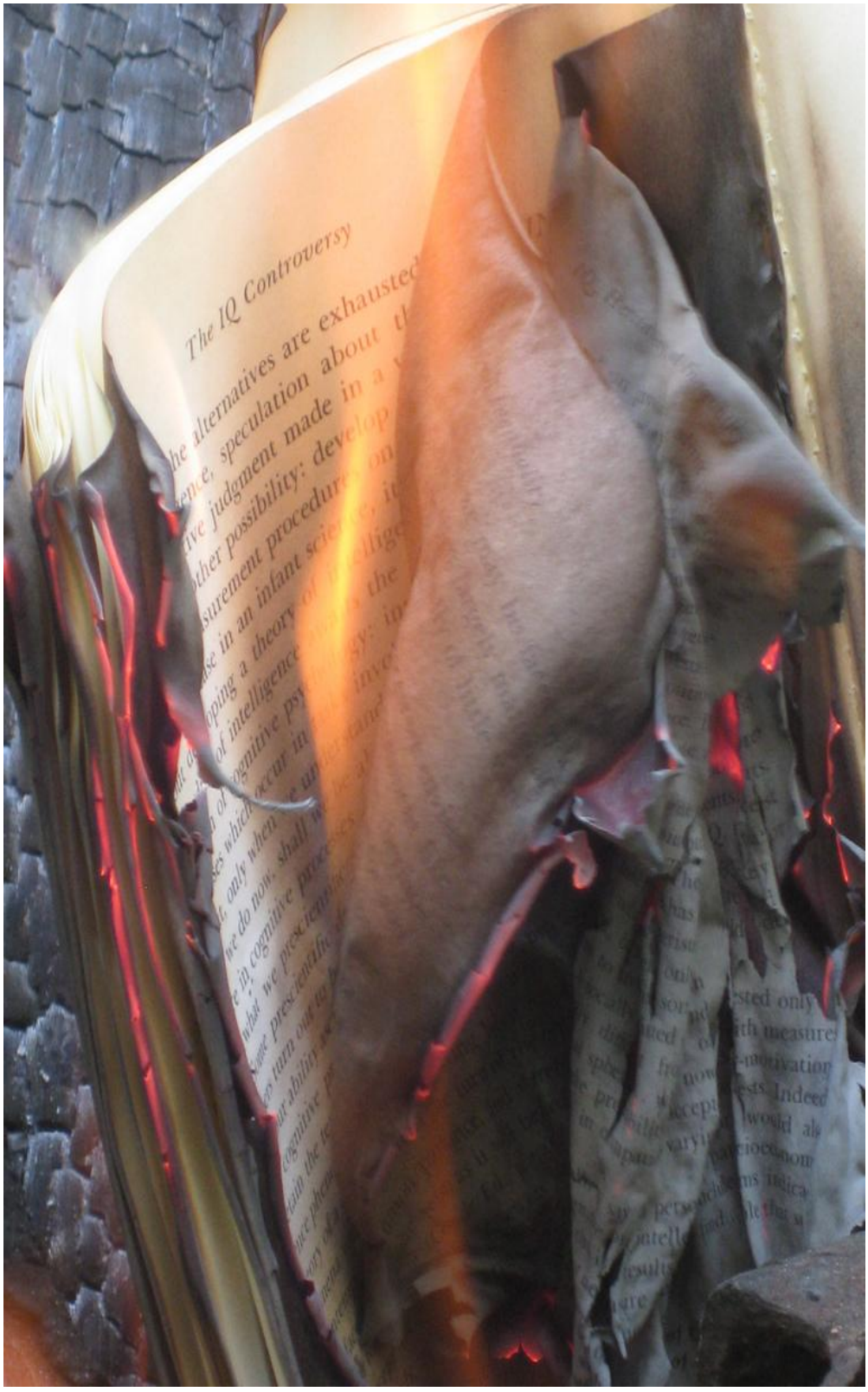
The IQ Controversy

University of Twente
Enschede - The Netherlands

Information Quality in Web 2.0

A Master Thesis by Gerrit J. MacLean

Supervised by

Roland Mueller, University of Twente, the Netherlands

Maya Daneva, University of Twente, the Netherlands

and

Markus Schaal, Bilkent University, Turkey

2008-2009

A story

In recent years, the author with authority has disappeared to the background of society. And in the dark depths and thick alleys of the internet a nameless and faceless horror arose: The anonymous author, known only by his address: 127.0.0.1. This horrifying monster claimed authority for itself and numerous people believed the monster, whatever it told them. And they told their friends that what the monster told them was true, since it was on the Internet. Some of them believed the monster, and others became a part of the monster themselves. And the monster which was called 'The anonymous author' or 'Web 2.0' grew and submerged the world slowly into darkness…

The years passed by and darkness prevailed in the hearts and minds of the citizens of the earth. But a spark of bright white light shone in the darkness… the spark of science. And another, a fluorescent blue light flashed in the dark, the light of criticism. There, in the dark alleys, where the light of science hardly ever protruded, were scientists and criticasters, waving their torches and flashing their razor sharp swords. They found the monster, which was a many eyed and many mouthed horror. With the monster at the tip of their swords they whispered: "From where is your authority? Defend yourself."

But the monster could not; it was no match for the blinding light and sharp swords of criticism and science. The triumphant look in their eyes told the monster that it was about to be defeated. But as the monster cowered back into the darkness, a knight in shining armour, bearing a shiny axe and a shield with an eye descended. With a thunderous crash he landed in front of the two with the torches. Eyes filled with fire and a voice like the distant echo of the thunder: "Do not bully the monster around. I will defend it."

On a pillar nearby sat a small figure with a faint mocking smile. "Fight all you like. But remember… I am the one who decides which side has won the day"

Contents

1 Overview

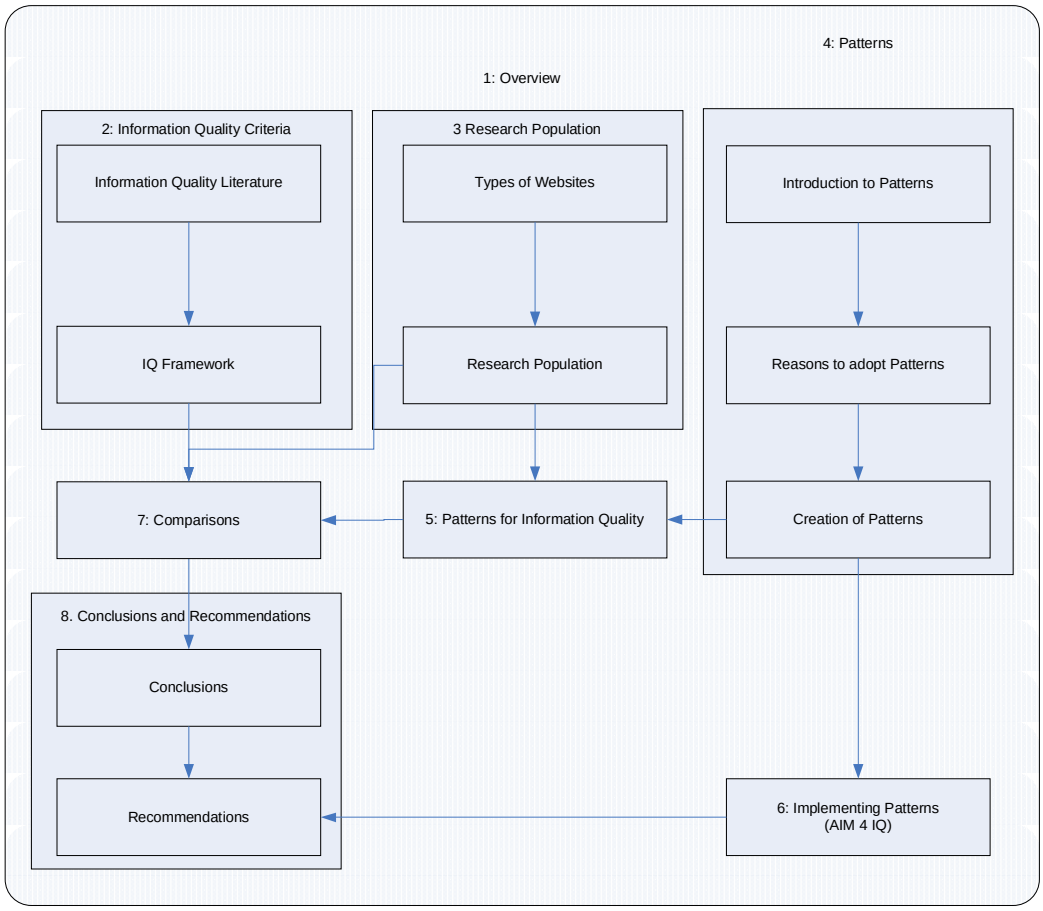

Figure 1: Outline of the Research

⇒ The numbered rectangles represent chapters.
⇒ The non-numbered rectangles represent sub-chapters.
⇒ The lines are logical links, representing that all chapters follow logically from their predecessors.

1.1 Introduction to the topic

In recent years, the development of websites has taken a big flight. In less than ten years, Google has become the world's most valuable brand (Millward Brown Optimor, 2009). According to web researcher Alexa.com, leading Web 2.0 sites have become established in the top of the most visited websites in the world.

People tend to be more on the internet and trust the information provided there. Especially websites which can bolster on a high reputation are trusted by people even at a scientific level. Under influence of the behaviour of other people, they will tend to consume those information objects which are often consumed by others, instead of those information objects which have a high quality (Salganik et al, 2006). Even reputable newspapers use the information provided by websites like Wikipedia. This resulted in a small scandal, as a newly appointed minister in Germany was treated to an additional first name by a prankster. As websites as spiegel.de and the newspaper

Bild copied this information, he had an additional name. This mistake was corrected, but raises questions about how and why we should trust such websites (Spiegel.de, 2009). In the Dutch juridical system, a court accepted a protest since Wikipedia was used as a source of information (Court Ruling, 2005). This procedure was adapted by the High Court, stating that Wikipedia has a clause that the information provided might not be accurate (Court Ruling, 2006). Yet, the information on Wikipedia is surprisingly high quality, as is the information on other websites with user generated content (Giles, 2005).

Quality of information is not as straightforward as it seems. Although information quality has been referred to as 'fit-for-use' (Wang & Strong, 1996; Knight & Burn, 2005) or 'It is information which must satisfy the needs of the user' (Parker et al, 2006), these definitions are not useful when it comes to improving and analyzing information quality. Therefore, we need a multi-dimensional framework which can encompass all the aspects of information needed for the indexing and categorization of the patterns in information quality, which will be analyzed in this research. Although there are some frameworks in the field which claim a solid scientific base, (e.g. Parker et al, 2006; Knight & Burn, 2005; Price & Shanks 2005) they all lack one aspect: completeness. Ironic, since completeness was one of the criteria of information quality they all mentioned. So a new framework was needed, aimed at completeness in a Web 2.0-environment. This resulted in a 42-criterion-rich framework.

At the moment, we can measure the quality of information on several of these criteria, especially the more technical criteria (Stvilia et al, 2008). For the less technical criteria this is still a subjective process, which is unavoidable, since many aspects of information quality are inherently subjective in nature. And this leads us to the conclusion: Whatever measurement instrument we develop; there will always be a subjective component in there. This is also the subject of current research (Price et al, 2008). Therefore, a new approach was needed, not targeting the quality of information as an end result, but as the result of a process.

These processes are not unique for every single website. Often they employ the same notions and working principles in their process of assuring and improving information quality. These notions and working principles are called 'patterns' which one can find at a lot of different websites. Collecting and indexing the patterns helps professionals communicate about what websites actually do to ensure information quality. In addition to this, it helps websites improving their processes concerned with information quality, by making the processes used at websites portable to other websites.

The websites from the sample, are not all equal. Websites differ in the information products they offer, from social networking to video-sharing and from encyclopaedias to trading goods. It is a broad range of information products that is offered by the research population. This does not only result in different patterns, but also in different strategic focuses of information quality on these websites.

This forms the content of the thesis: Analyzing the process that Web 2.0 sites use to assure the quality of the information provided. In Figure 1: Outline of the Research a graphical representation of the chapters and their logical links is given.

1.2 Research Goal

The goal of the research is to provide insight in the methods used for improving quality of information, when the creators and consumers of the information are numerous and unknown, by creating a pattern language for information quality improvement methods for websites with mainly user-generated content.

1.3 Relevance

This research is relevant for the further development of quality assurance among websites with a large share of their information generated by numerous and unknown users. As such it is relevant for the quality of information used for educational purposes, for business decisions and for media usage.

Recent articles even explicitly state the need for this research (Stvilia et al, 2008): "There is a need for empirical studies of existing IQ assurance models, with a goal to develop a knowledge base of conceptual models of IQ, taxonomies of quality problems and activities, metrics, trade-offs, strategies, policies, and references sources." Although such a broad goal is overstretching the time and scope limits of a master thesis, it indicates the relevance of this subject.

1.4 Project Boundaries

The research will focus on methods of improving quality of information.
The research will focus on methods on a business and process level.
The focus is on websites which have large quantities of user generated content.
The focus is on methods currently in practice and emerging methods. Legacy methods and methods for the future are out of the scope of this research.
The research is an explorative-theoretical research, with the goal of providing insight. It can be regarded as generating descriptive knowledge, although the description of these patterns can be used by websites afterwards to improve information quality.

1.5 Research Question.

Research Question: *What methods to ensure quality of information are employed by sites which have the bulk of their information generated by users of unknown expertise, of unknown intent and in an unknown context?*

There are several aspects to this research question, which have been redefined in the different sub-questions, to create simpler, more concrete questions. It is divided according to the various parts of the research that need to be explained or elaborated. The chapters follow roughly the sub-questions.

(Quality of Information Definition) (Chapter 2)

1. What is Quality of Information in this context?

2. What are the Characteristics of Information Quality in Web 2.0 environments?

   (Research Population) (Chapter 3)
3. Which characteristics are distinguishing for Web 2.0-sites?
4. What is a good sample of existing Web 2.0 sites for the research?

   (General Pattern Creation) (Chapter 4)
5. What elements of Patterns are relevant for documenting Information Quality Methods?

   (Specific Patterns Creation) (Chapter 5)
6. What are used patterns in Information Quality Assurance?

   (Implementation of Patterns) (Chapter 6)
7. How should these patterns be implemented at other websites?

1.6 Research context:

The context of this research is websites and the processes, procedures and communities behind these websites. Within these websites we will explore the methods used to ensure information quality.

1.7 Research Model

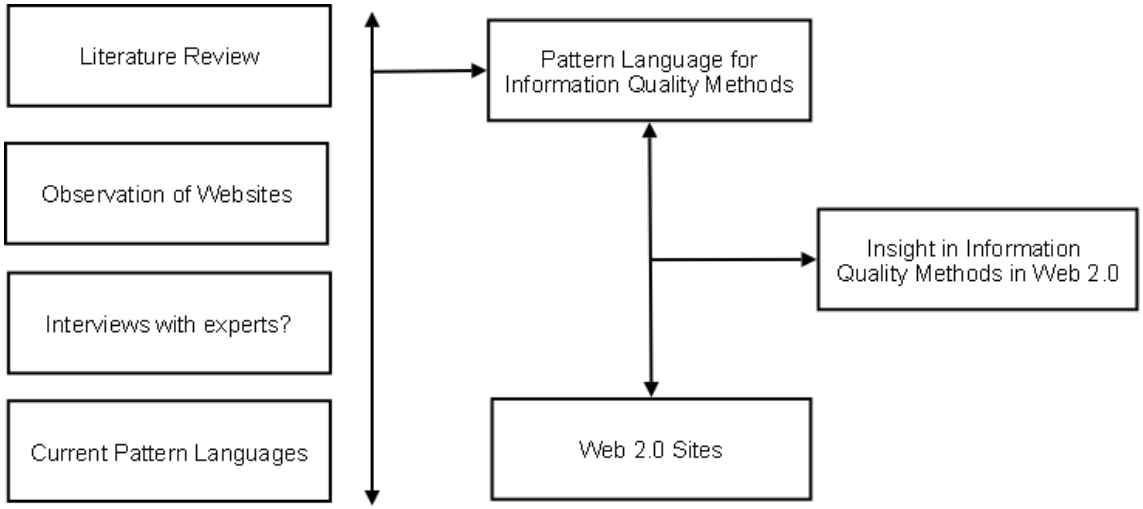

Figure 2: Research Model

The analysis of literature, the observation of the methods employed by websites, the currently used pattern languages and interviews with experts, will provide the necessary elements for a pattern language for Information Quality Methods, which will be used to analyze the Web 2.0 Sites, in order to provide insight in Information Quality Methods.

It may be noted that the literature review, observation of websites and possible interviews with experts will be employed as methods for the analysis of Web 2.0 sites as well. The process is more iterative than this research model suggests.

2 Information Quality Criteria

The research uses 42 criteria of information quality. A criterion is an aspect, which tells something about the characteristics of the information. These should not be confused with information quality dimensions, since the criteria presented here are clearly interdependent.

2.1 Justification

The 42 criteria of information quality were needed, because the research, collecting patterns and indexing websites, has need for a comprehensive set of information quality criteria. Completeness is in this case more important than other aspects, like mutual exclusivity, as was proposed by some authors (Eppler & Wittig, 2000). This does not mean that all other aspects were neglected in order to achieve a high level of completeness. Ignorance of the criterion conciseness would, for example, result in an impractically big framework. The choice for completeness as focal point of the framework, results obviously in a more complete framework. The trade-off is that some other criteria of information quality (e.g. conciseness, informativeness) will perform poorer than in other frameworks of information quality.

2.2 Research Methodology

The method used to find these aspects is twofold. First there is the systematic literature review, which yielded forty different criteria of information quality. Second, two more criteria of information quality were found, by observing websites which focused on an aspect of information quality, which was not mentioned in the literature.

2.2.1 Employed search strategy:

In order to cover the top 25 IS-journals, as well as most other sources of information, a selection of the employed *Search Engine*s had to be made. This resulted in four *Search Engine*s of choice: the scientific engines Scopus and Web of Knowledge, which were complemented with a manual search in the Communications of the AIS. These three engines are according to Schwartz & Russo (2004) sufficient to cover the top 25 journals in Information Systems. Google was added to the search, but only results from the first three pages were analyzed, as enrichment for the employed search strategy. In Table 1: Employed Search Terms, the employed Search Terms and the number of results found are displayed.

In addition to keyword search, backward search and forward search were employed to come to more relevant results. In addition to that Google Scholar was used for both keyword and backward searches however; this *Search Engine* is still beta and can therefore have no proven value as to reach completeness. Initially search results from 2002 and later were included for further analysis, but as the research progressed, we could raise that to 2004 and later; since two apparently independent literature reviews from past 2004 (Knight & Burn, 2005; Parker et al, 2006) were found.

2.2.2 Employed search terms:

Used Search terms	Conclusion	Scopus	Web of Science	CAIS	Google
Information Quality	Overload	153778	5356		139000000
"Information Quality"	Overload	756	400		1780000
"Information Quality", Framework		87	48	18	298000
"Information Quality", Taxonomy		5	3	2	20300
"Data Quality"	Overload	6725	3519	0	8500000
"Information Quality", Criteria		47	37	0	110000
"Information Quality", "Data Quality"		49	25	0	96500
"Data Quality", Framework		324	167	19	470000
"Data Quality", Taxonomy		24	10	2	222000

Table 1: Employed Search Terms

2.3 Results

For the basis of the framework, a literature study was conducted. Two prior literature reviews (Knight & Burn, 2005; Parker, Moleshe, De la Harpe and Wills, 2006) were used for coverage of information quality frameworks before 2004. Certainly, the body of knowledge stretches back even further, with definitions based on the seminal work by Wang and Strong (1996).

For the time span since 2004, an extensive literature survey revealed four additional frameworks until 2007. In total, 6 relevant sources are used: 2 literature reviews to cover all relevant literature from before 2004 and 4 more recent papers. These 6 papers are briefly described below.

First, Stvilia, Gasser, Twidale and Smith (2007) propose 22 attributes in the categories: Intrinsic, Relational/Contextual, and Reputational. Second, Su and Jin (2007) propose 15 attributes in the categories Syntactic IQ, Semantic IQ, Pragmatic IQ and Physical IQ. This paper does not only identify some new perspectives, but also identifies trade-offs between information quality criteria.

Rao and Osei-Bryson (2007) did a research in which they transferred the criteria for data quality towards criteria for knowledge quality. Since the concept 'information' is positioned between data and knowledge, this paper gives valuable insights in the criteria of information quality. They provide one new criterion, "Degree of context". Price and Shanks (2005) developed a framework based on semiotics. They provide a framework for shaping and categorizing the extensive list of criteria. This approach, which uses Syntaxes, Semantics, and Pragmatics, helps identifying to what category each criterion should be mapped. Yet, there are still some ambiguous criteria, which could fit in more than one of these categories. This is improved by the empirical refinement of the framework (Price & Shanks, 2005b).

The literature review by Knight and Burn (2005) is the older of the two literature reviews used. The paper is part of a research project aimed at developing an internet focused crawler that uses quality criteria. It evaluates 12 models, and combined them using the most often cited criteria. For the definitions of the individual quality criteria the paper adhered to Wang and Strong (1996), diverging to own definitions when the quality criterion was not represented in that model. The paper mentions the fact that there are no quality control procedures for the internet, and users have to make their own judgments about quality, which once again stresses the importance of this research.

The literature review as conducted by Parker et al. (2006) addresses 11 of the same papers as Knight and Burn (2005). They too point out that there are no quality control standards as how to publish information on the World Wide Web. It largely overlaps with the previous paper.

2.4 Grouping of Information Quality Criteria

Quality of information is characterized by numerous criteria, which are interdependent and not mutually exclusive. For the development of a comprehensive framework, different criteria have been adopted from many different branches of prior work. A grouping of these criteria is therefore necessary, for researchers to be able to maintain sound and consistent levels of abstraction and granularity when carrying out evaluation studies (like ours). Clearly, such a grouping involves linking and labelling the different criteria of information quality. For an improved understanding of the criteria in this framework, we have merged two ways of categorizing information quality criteria.

The first categorization scheme is a semiotic framework for data quality as proposed by Price and Shanks (2005) including their later modifications (Price and Shanks, 2007). They identify three different groups of information quality criteria based on the semiotic categories of syntax, semantics, and pragmatics.

2.4.1 Definitions of the groupings of Information Quality.

Definition 1: The **syntactic quality category** describes the degree to which stored data conforms to other information (e.g. rules or stored metadata).
This definition differs from Price and Shanks (2005) in that it avoids the data base centric explicit reference to metadata and replaces it by conformance to other information. As Price and Shanks (2005) describe, “the syntactic level consists of any relation between sign representations.” Therefore syntactic IQ criteria are concerned with the relationship between the information and other information (see Figure 2).
Definition 2: “The **semantic quality category** describes the degree to which stored data corresponds to represented external phenomena, i.e. the set of external phenomena relevant to the purposes for which the data is stored (i.e. use of the data).” (Price & Shanks, 2005). As Price and Shanks (2005) describe, “the semantic level consists of any relation between a sign representation and its referent.” Therefore semantic IQ criteria are concerned with the relationship between the information and the reality (see Figure 2).

Price and Shanks (2005) define the pragmatic quality category as follows: "The pragmatic quality category describes the degree to which stored data is suitable and worthwhile for a given use, where the given use is specified by describing two components: an activity (i.e. a task or set of tasks) and its context (i.e. location – either regional or national – and organizational sub-unit; typically created as a result of functional, product, and/or administrative sub-division)."
Price and Shanks (2005) describe the pragmatic level as "any relation between a sign representation and its interpretation". Therefore the pragmatic IQ criteria are concerned with the relationship between the information and the user (see Figure 2). We subdivide the pragmatic category further according to the aspect of how the quality information can be assessed. Naumann and Rolker (2004) distinguish between quality criteria that can be determined by the information content by the querying process or only by the user. The following definitions are based on the work of Naumann and Rolker (2004) and split the pragmatic quality category of Price and Shanks (2005) into three aspect-oriented categories:

Definition 3: The **user-pragmatic information quality category** describes the degree to which stored data is considered credible and trustworthy.
Definition 4: The **information-pragmatic information quality category** describes the degree to which the information is useful, applicable and understandable by the user for the task at hand.
Definition 5: The **process-pragmatic information quality category** describes the degree to which stored data can be found and accessed.

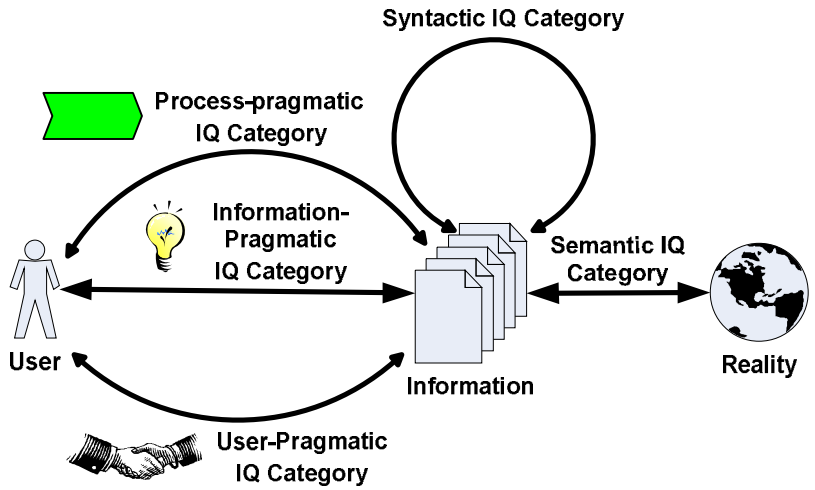

Figure 3: Visualization of Information Quality Categories

2.4.2 Visual representation of hierarchical grouping of information quality criteria.

The information quality criteria mentioned in Figure 4: Networked grouping of information quality criteria, are elaborated in the table in the next chapter. Not all links among information quality criteria are mentioned in the networked grouping, more links exist, but the most important as perceived by the researcher are displayed.

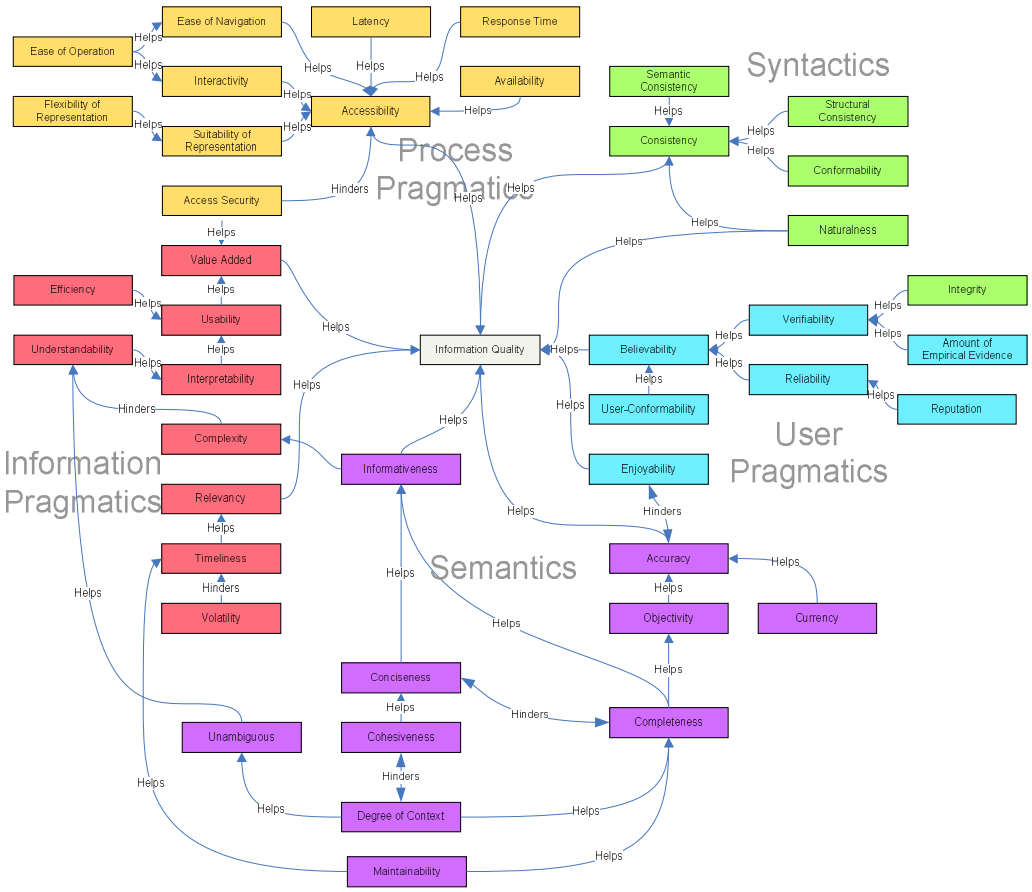

Figure 4: Networked grouping of information quality criteria

2.5 Definition of Information Quality

As our investigation progressed, two additional criteria of Information Quality, being User-conformability and Fun were discovered. Especially the latter is important, as websites thrive on only this aspect of information quality. (for example uncyclopaedia.com)

After the finding and rejecting of criteria as presented by the different authors, there was a need for synthesis among all different criteria. Some had to be adapted for clarity, but most for consistency across all criteria. Also; the criteria are in this section no longer grouped by author, but by semiotic groups. These rewritten definitions are the basis for analysis.

IQ Category	Name	Source	Subjective/ Objective	Definition
Syntactics	Consistency	Knight and Burn (2005)	Subjective/ Objective	The extent to which an information object is presented in the same format and compatible with other, similar information objects.
	Semantic Consistency	Stvilia (2007)	Subjective	The extent to which the same words and values are used to convey the same meanings and concepts, with respect to other, similar information objects.
	Structural Consistency	Stvilia (2007)	Objective	The extent to which similar attributes or elements of an information object are consistently represented using the same structure, format and accuracy as similar information objects.

	Conformability	Su and Jin (2007)	Subjective	The extent to which the data is free of contradictions and conformation breaks with respect to the current dominant culture.
	Integrity	Su and Jin (2007)	Subjective	The extent to which the scope of the metadata is adequate.
	Naturalness	Stvilia (2007)	Subjective	The extent to which the model or schema and content of an information object are expressed by conventional, typified terms and forms.
Semantics	Accuracy	Wang and Strong (1996)	Objective	The extent to which the information object represent the external phenomena correct and free of error.
	Completeness	Wang and Strong (1996)	Objective	The extent to which information incorporates all key factual information of the external phenomena it represents and is free of significant omissions.
	Conciseness	Wang and Strong (1996)	Objective	The extent to which information compactly represents the external phenomena.
	Objectivity	Wang and Strong (1996)	Subjective	The extent to which information is unbiased, unprejudiced and impartial with regard to the external phenomena it represents.
	Cohesiveness	Stvilia (2007)	Subjective	The extent to which the information object is focused on one external phenomenon
	Informativeness	Stvilia (2007)	Subjective/ Objective	The amount of information contained in an information object divided by the length of the information object.
	Maintainability	Su and Jin, (2007)	Objective	The extent to which information can be organized and updated to comply with the external phenomena on an ongoing basis.
	Degree of Context	Rao and Osei-Bryson (2007)	Subjective	The extent to which context is provided for in information object.
	Unambiguous	Price and Shanks (2005)	Subjective	The extent to which the information as it is represented, maps only one possible external phenomenon.
	Currency	Stvilia (2007)	Objective	The age of an information object.
User Pragmatics	Believability	Wang and Strong (1996)	Subjective	The extent to which information is regarded as true and credibly mapping the real world object by the information consumer.
	Verifiability	Stvilia (2007)	Objective	The extent to which the correctness of information is verifiable or provable by the information consumer.
	Amount of Empirical Evidence	Wang and Strong (1996)	Objective	The extent to which the quantity or volume of available data or metadata is appropriate to support the conclusions and claims made.
	Reliability	Knight and Burn (2005)	Subjective	The extent to which the provider of the information is regarded as reliable by the information consumer.
	Reputation	Wang and Strong (1996)	Subjective	The extent to which provider of the information is regarded as reliable by society.
	User conformability	Observation	Subjective	The extent to which the information is free of contradictions and conformation breaks with respect to the user.
	Enjoyability	Observation	Subjective	The extent to which the consuming of the information object is regarded as enjoyable.
Information Pragmatics	Value-added	Wang and Strong (1996)	Subjective	The extent to which the information object is beneficial, provides advantages from its use for the task at hand.
	Usability	Knight and Burn (2005)	Subjective	The extent to which information is clear and easily used for the task at hand.
	Relevancy	Wang and Strong (1996)	Subjective	The extent to which information is about the subject for the task at hand.
	Timeliness	Wang and Strong (1996)	Subjective/ Objective	The extent to which the information is sufficiently up-to-date for the task at hand.
	Efficiency	Knight and Burn (2005)	Subjective	The extent to which the information object is able to quickly meet the information needs for the task at hand.
	Interpretability	Wang and Strong (1996)	Subjective	The extent to which the information object can be interpreted by the information consumer to tackle the situation at hand. (Non specific-ness)
	Understandability	Wang and Strong (1996)	Subjective	The extent to which the information object is represented in language, signs and expressions familiar to the information consumer.
	Complexity	Stvilia (2007)	Subjective	The extent of cognitive complexity according to some index or indices.

	Volatility	Stvilia (2007)	Objective	The amount of time the information remains valid in the context of a particular activity.
Process Pragmatics	Access Security	Wang and Strong (1996)	Subjective	The extent to which access to information is restricted appropriately to maintain its security.
	Accessibility	Wang and Strong (1996)	Subjective	The extent to which information is available, or easily and quickly retrievable.
	Latency	Naumann and Rolker (2000)	Objective	The amount of time until first information reaches a user after a request.
	Response Time	Naumann and Rolker (2000)	Objective	The amount of time until complete information reaches a user after a request.
	Ease of Operation	Wang and Strong (1996)	Subjective	The extent to which the information is easy to manipulate, aggregate and combine with other information.
	Availability	Knight and Burn (2005)	Objective	The relative amount of time which information is available to the information consumer. (Up time)
	Ease of Navigation	Knight and Burn (2005)	Subjective	The extent to which data are easily found and linked to.
	Interactivity	Su and Jin, (2007)	Subjective	The extent to which the information retrieval and creation process can be adapted by the information consumer.
	Suitability of Representation	Price and Shanks (2005)	Subjective	The extent to which the presentation of the information is suitable for your needs.
	Flexibility of Representation	Price & Shanks (2005)	Subjective	The extent to which data can easily be manipulated and the data presentation customized as needed.

Table 2: Definitions of Information Quality Criteria

2.6 Information quality criteria among frameworks

Information Quality Criteria/Patterns	Wang & Strong 1996	Naumann & Rolker 2000	Knight & Burn 2005	Parker et al. 2006	Roa & Osei-Bryson 2007	Su& Jin 2007	Stvilia et al. 2007	Stvilia et al. 2008	Price & Shanks 2005	MacLean 2009	Total
Consistency			1	1	1			1		1	5
Semantic Consistency							1			1	2
Structural Consistency	1	1					1			1	4
Conformability				1		1			1	1	4
Integrity						1				1	2
Naturalness							1	1		1	3
Accuracy	1	1	1	1	1	1	1	1	1	1	10
Completeness	1	1	1	1	1	1	1	1	1	1	10
Conciseness	1	1	1			1				1	5
Objectivity	1	1	1	1						1	5
Cohesiveness							1	1		1	3
Informativeness							1	1		1	3
Maintainability						1				1	2
Degree of Context					1					1	2
Unambigous					1				1	1	3
Currency	1				1		1			1	4
Believability	1	1	1	1						1	5
Verifiability	1	1		1			1	1		1	6
Amount of Empirical Evidence	1	1	1							1	4
Reliability		1	1				1		1	1	5
Reputation	1	1	1	1			1			1	6
User conformability										1	1

Fun										1	1
Value-added	1	1	1	1		1			1	1	7
Usability			1		1					1	3
Relevancy	1	1	1	1		1	1	1	1	1	9
Timeliness		1	1			1			1	1	5
Efficiency			1	1						1	3
Interpretability	1	1								1	3
Understandability	1	1	1	1		1			1	1	7
Complexity							1	1		1	3
Volatility					1		1	1		1	4
Access Security	1	1	1	1		1	1		1	1	8
Accessibility	1		1	1		1	1	1	1	1	8
Latency		1				1				1	3
Response Time					1	1				1	3
Ease of Operation	1	1		1	1					1	5
Availability		1	1							1	3
Ease of Navigation			1							1	2
Interactivity						1				1	2
Suitability of Representation				1					1	1	3
Flexibility of Representation	1								1	1	3
	18	19	19	16	10	15	16	11	13	42	

Table 3: Occurence of Information Quality Criteria among selected frameworks

2.7 The edge between Quality Criteria and Absence of Criteria.

There is one criterion in the frameworks encountered which was on the edge of the criteria, and even might be classified as a pattern. This was Awareness of bias (Shanks & Corbitt, 1999).Therefore, the question was raised: "Is it truly a criterion of information quality?" Creating awareness of bias not a goal in itself, since objectivity is the desirable criterion, but awareness of bias might be a good alternative if objectivity is unachievable.

On the other hand, shouldn't we treat this as a pattern, that the 'making someone aware' of the bias, is a method to enhance the perceived objectivity? The same can be said about for example currency. Is it desirable that the user of the information is made aware of the 'time-dimension' of the data? (e.g. statistical analysis based on 2006 figures, because 2008 figures aren't available) It is desirable, but isn't it more desirable that the data is up-to-date?

The answer is in this case that we have adapted the *Declaration of Failure* as a pattern. This pattern has the advantage that it incorporates the awareness of bias, the awareness of currency or any other lacking information quality criterion, in a process to improve information quality. Hence, we regard 'Awareness of Bias' not as a criterion of information quality.

2.8 Appreciation of Information Quality Criteria

Information criteria come in five different semiotic groups, but are also valued differently. Therefore, a context dependent grouping should be added to the framework, consisting of five new categories. According to the most recent developments in the Kano-model, in each context information quality criteria can be regarded as either (Zultner & Mazur, 2006):

1. Expected (“atari mae”)
2. Desired (“ichi gen teki”)
3. Exciting (“mi ryoku teki”)
4. Indifferent (“mu kan shin”)
5. Reverse (“gyaku”)

These factors are graphically represented in Figure 5: Appreciation of Information Quality Factors. In the middle there is an area where the appreciation of the different factors is neutral, this is where the lines for expected and exciting factors end.

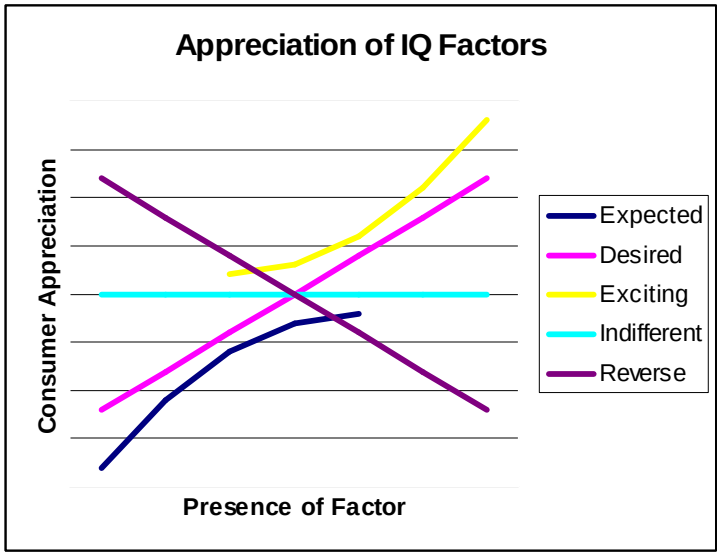

Figure 5: Appreciation of Information Quality Factors

2.8.1 Expected factors

Expected factors are those who need a minimum level, before there is any information quality at all. One can get an impression of this by simply asking himself: What will happen if this criterion reaches zero (or infinite)? Is the information than no longer valuable at all? In that case, the information quality criterion is an expected factor. Criteria which need to have a minimum-level are the most common among the Process Pragmatic group, like Response Time, Latency and Ease of Operation. If one of these criteria horribly underperforms, the information becomes worthless.

2.8.2 Desired factors

Desired factors do not need a minimal level, but still can affect the quality of information. Typical: Flexibility of Representation, Objectivity and Verifiability. When one of these criteria is not represented, it makes the information not completely worthless, but it should be mentioned that the information is lacking in one of these criteria.

2.8.3 Exciting factors

Exciting factors are those criteria which can create a real wow feeling if they are strongly present. When it is not present one will hardly miss it, but when it is present, you get a good feeling out of it. The most obvious exciting factor is Enjoyability,

especially when not expected. Users can go and search for information objects which excel at this exciting factor, in which it becomes a straight factor.

2.8.4 Indifferent factors

Indifferent factors are those criteria about which the information consumers do not care. “Mu kan shin” literally translates as “Not the gateway to the heart”, and represents those factors about which customers are completely indifferent. There is a group of information quality criteria which can become indifferent factors. For example, relevancy, when customers are surfing the internet for their leisure.

2.8.5 Reverse factors

At times, information quality factors may become reverse factors, when the presence of such an information quality criterion is undesirable. An example is Enjoyability, which may become a reverse factor when information consumers are looking for e.g. news or stock data.

2.9 Shortcomings and remarks

Price & Shanks (2005) point out some important shortcomings to Wang & Strong, (1996) which apply for this approach as well. The first is the interdependencies which link several criteria of Information Quality. That is something that applies to this taxonomy, for example in the area of Timeliness, Volatility, and Currency. These criteria of information quality are strongly interdependent, which would be a shortcoming. We acknowledge this fact and have tried to visualize and model the most important of these interdependencies in Figure 4: Networked grouping of information quality criteria. However, this is not a problem in our research context, since the purpose of this taxonomy is classification of patterns which help information quality. The fact that some patterns target several information quality criteria is not a problem.

Another shortcoming noted by Price & Shanks (2005) is that some criteria are not generic. Thus different patterns may only occur in certain contexts, where the information quality criterion is applicable. This is not as such a shortcoming, but it is a reason why there are some portability problems with the application of patterns across different types of websites. It restricts the portability of the different patterns, and the applicability of the taxonomy. Whether or not a pattern can be implemented by a website is a matter of strategic focus and the nature of the information objects of the website.

Another shortcoming or design choice of this approach is the vast number of criteria; which imposes a problem on the requirement that an information quality framework should be concise. The framework proposed has 42 criteria, 40 from literature, 2 from observation, which all have a (slightly) different definition. This is about twice as many as the most comprehensive articles (Naumann & Rolker, 2004; Dedeke, 2000). The answer to this is fairly simple: this taxonomy is about creating a collection of labels, with which we can classify the patterns used for assurance of information quality. These methods can bear different labels, if they address different criteria of information quality. The problem of overlap, which forces the frameworks concerned

with measuring information quality to great specificity, is not an issue. When a pattern is discovered, which would affect a certain criterion, but that criterion of information quality is not in the framework, the relevance of the pattern might be misunderstood. Another argument for making this taxonomy this comprehensive, is the discovery of possible gaps in quality assurance.

Eppler & Wittig, (2000) have defined four goals which an information quality framework should achieve a) it should be systematic and concise, b) a scheme should help to analyze and solve information quality problems, c) it should provide a basis for information quality measurement and proactive management and d) that it should provide the research community with a conceptual map.

Those requirements are partially fulfilled by our model of information quality criteria. The five categories provide a systematic scheme for the concise information quality criteria. We provide a strong base for information quality measurement and we provide the research community with a conceptual map. We do provide a solution for information quality problems and with the patterns in chapter 5 we support proactive information quality management.

The fourth requirement, that it should provide the research community with a conceptual map, can be considered as achieved as well. It does indeed provide a conceptual map, although it might be suitable for only a limited number of purposes. We have used this framework as a building block for the sorting of information quality patterns.

2.10 Conclusions

From the findings in 2.3 and the followed methodology, we can conclude that the information quality criteria together form a more comprehensive and complete framework than the frameworks previously published. The fact that during the research only two additional criteria of information quality were discovered underlines this fact. According to principles of the Grounded Theory (Dick, 2000) this indicates saturation in the theory. This results in a usable diversification of the framework, which can be used in the further analysis of patterns and websites.

3 Research Population

In this chapter the research population will be analyzed and indexed according to the characteristics of the websites. Researchers have characterized Web 2.0 as a 'marketing buzzword' (O'Reilly, 2009) or as a set of principles and practices (O'Reilly, 2009). For our research we found it more feasible to distinguish websites from Web 2.0 sites by the observation of 4 types of behaviour.

3.1 Types of websites

In the Web 2.0-environment, we found four main categories. This classification in categories of behaviour was reached in an emergent way, as the websites studied displayed similar behaviour among four groups. We have used this division since patterns were found to be within categories transferable and applicable to other websites, but across categories this was harder to do. This division is based on the dominant behaviour that websites display, although most websites in the research population display several types of behaviour. All websites were found to display one dominant type of behaviour. The four types of behaviour are briefly explained below.

3.1.1 Collaborative Content Creation

Collaborative Content Creation behaviour is the creation of information objects, cooperatively, by two or more persons with possibly different knowledge and skills and supplying these information objects to users. The Collaborative Content Creation environments are mainly the wiki-style websites. The most striking example is Wikipedia. In these websites, multiple users can collaborate to create a single information object. In practice that are mainly text-based articles, since the technology and practices to collaboratively create and enhance pictures, videos, songs or graphs are still immature. However, there are websites in which the users collaboratively create graphs, or funny pictures.

3.1.2 Media Provision

Media Provision behaviour is the provision of information objects (like photos, videos and articles), which are created by a single user, to other users. The Media Provision websites are the upload-and-share kind of sites, like video and photo sharing sites. The most visited website of this kind is Youtube.com. The media provided on these websites cannot be collaboratively created, although sites may have incorporated Metadata Generation, to create ratings and tags.

3.1.3 Metadata Generation

Metadata Generation behaviour is the collecting of tags, ratings and opinions about information objects or real-world objects and providing this information to other users. The Metadata Generation websites are often recommendation and tagging sites, like Digg.com and Del.icio.us. The key to these sites is that they do not necessarily have to offer the media themselves, but collect information about media which in turn can help users to consume some information objects. Often Metadata Generation is combined with Social Networking and/or Media Provision.

3.1.4 Social Networking

Social Networking behaviour is the communication among users and the information about users and their respective relations. Social Networking started out as an additional feature in Media Provision sites, but at the moment, social networking sites can flourish without other prima focuses. The most used websites in this field are Myspace and Facebook, although Myspace also has strong hints of Media Provision. The key is that Social Networking is not about information objects, but about links to other users of the website.

3.1.5 Validation of the classification of websites.

For the validation of the classification made, the behaviour of websites in our research population was described by 8 independent reviewers, next to the researcher. The reviewers were light and casual users of the websites, all students. The questionnaire in Appendix D. was used to distinguish websites between their dominant and secondary behaviour. Reviewers were asked to only fill in websites that they were familiar with, since asking these users to become familiar with all websites in the research population would place a heavy time burden on these volunteers. All questionnaires were returned, with at between 13 and 48 items rated, resulting in 282 items rated by 8 volunteers. This resulted in a mean linearly weighted agreement between the raters and the researcher of $\kappa(wl)= 0,532$ and a mean quadratic weighted agreement between the raters and the researcher of $\kappa(wq)= 0,607$ (Vanbelle & Albert, 2009). This indicates a moderate to substantial agreement, according to Landis & Koch (1977). This gives confidence that the description of Web 2.0 as a set of behaviours, and the classification of websites according to those behaviours is feasible.

After the questionnaire, the reviewers were consulted in informal conversation, to determine whether the reviewers understood the behaviour described correctly. We make the note that although some websites do display hybrid behaviour, no class 'hybrid' was added. This is because even the most hybrid websites were classified as a single type as dominant behaviour, by all but one respondent. The specific results are to be found in Appendix E.

3.2 Justification of the research population

The websites were chosen because these are the top Web 2.0 sites in the Netherlands (Alexa, 2009). The Web 2.0 sites which generate the most traffic are shown in the table below. This table also underlines the impact of Web 2.0, as 16 of the top 50 sites in the Netherlands are Web 2.0. The data are from February the 3rd, 2009, from Alexa.com. Observation shows that these data change only little in the course of the research.
We make the note that for ethical and practical reasons, sites in the erotic video business were not included in this research.

Web 2.0 Rank		Overall Rank	Category
1	Hyves.nl	2	Social Networking
2	Youtube.com	4	Media Provision (Video)

3	Marktplaats	7	Media Provision (Trade)
4	Wikipedia	9	Collaborative Content Creation (General)
5	Blogger.com	15	Media Provision (Blogging)
6	Facebook	18	Social Networking
7	Flickr	25	Media Provision (Photo)
8	Dumpert.nl	31	Media Provision (Video/Photo)
9	IMDB	32	Media Provision (Movies)
10	Myspace	35	Social Networking
12	Photobucket	38	Media Provision (Photo)
13	Web-log.nl	39	Media Provision (Blogging)
15	Ebay	43	Media Provision (Trade)
16	Wordpress.com	49	Media Provision (Blogging)

Table 4: Web 2.0 sites in the top 50 most visited websites

⇒ Source: Alexa, the web information company (Alexa, 2009)

According to the Grounded Theory methodologists research subjects may be identified on an ongoing basis, adding to the research population as theory emerges (Dick, 2000). Therefore, the first research subject is Wikipedia. The choice for Wikipedia as first research subject is caused by the fact that this website has a open and clear structure. It is transparent which makes it easier to research. An additional advantage is that there is already a good body of knowledge on the subject of Wikipedia. The scientific *Search Engine* Scopus returned 434 scientific articles involved with Wikipedia as of 27 October 2008. This is satisfactory as other sites gained about the same number of hits. That is 193 for Youtube.com, 815 for Yahoo! (which is older), 414 for Ebay, 119 for Facebook and 163 for Myspace. Also, that it is top 10 site in The Netherlands, as well as most Western countries and even world-wide, makes it a solid choice as first research subject.

The second research subject is another global leader as well, Youtube.com. We included it because the focus of Youtube.com is clearly different from the focus of Wikipedia, but it is also one of the international top 10 websites. Next to these websites, more websites were observed, all of them in the top 50, but also other websites were added to the research, because of their interesting nature, for example the site icanhascheezburger.com which is a website which creates information objects, which are different in nature than other website, or offline media. Or for the sake of completeness, for example vimeo.com, which provided another sample for a video website. Other websites, which provided insights, but are not included in this research further are simple.wikipedia.org, conservapedia.com, speecha.com, and hotmail.com, as well as many other websites.

Web 2.0 Rank		Overall Rank	Category
1	Hyves.nl	2	Social Networking
2	Youtube.com	4	Media Provision (Video)
3	Marktplaats	7	Media Provision (Trade)
4	Wikipedia	9	Collaborative Content Creation (General)
5	Blogger.com	15	Media Provision (Blogging)
6	Facebook	18	Social Networking
7	Flickr	25	Media Provision (Photo)
8	Dumpert.nl	31	Media Provision (Video/Photo/Audio)
9	IMDB	32	Media Provision (Movies)
10	Myspace	35	Social Networking
12	Photobucket	38	Media Provision (Photo)
13	Web-log.nl	39	Media Provision (Blogging)
15	Ebay	43	Media Provision (Trade)
16	Wordpress.com	49	Media Provision (Blogging)
unknown	Zoover	unknown	Metadata Generation
unknown	Boardgamegeek.com	unknown	Metadata Generation
unknown	Last.fm	unknown	Metadata Generation
unknown	Digg.com	unknown	Metadata Generation
unknown	Uncyclopedia.info	unknown	Collaborative Content Creation (Fun)
unknown	Wikia.com	unknown	Collaborative Content Creation (Special Interest Groups)
unknown	Wikihow.com	unknown	Collaborative Content Creation (Solutions)
unknown	Vimeo.com	unknown	Media Provision (HD Video)
unknown	Icanhascheezburger.com	unknown	Media Provision (Lolcats)

Table 5: Research Population

⇒ Source: Alexa, the web information company (Alexa, 2009)

⇒ Added interesting research subjects by the researcher on 3 February, 2009.

3.3 Description of the websites in the research population

3.3.1 Hyves.nl

Hyves.nl is the number 2 website of the Netherlands, a typical Social Networking website. Due to network effects, this website is the website of choice among Dutch users of social network sites. The website displays Media Provision and Metadata Generation behaviour as well.

3.3.2 Youtube.com

Youtube.com is the leading video website in the Netherlands, ranked fourth in traffic, and ranked third worldwide. It shows videos to users which are uploaded by other users, and provides the ability to give feedback. It is a Media Provision site, but displays significant Metadata Generation behaviour and Social Networking behaviour as well.

3.3.3 Marktplaats.nl

Ebay.com's subsidiary Marktplaats.nl is one of the two trading websites within the research population. The website is ranked 7th in traffic in the The Netherlands. It is consequently classified as Media Provision, but shows small traces of Metadata Generation. It displays the least signs of Web 2.0 behaviour of all the websites in the research population.

3.3.4 Wikipedia.org

Wikipedia.org is the only non-commercial website in the top 10. It is ranked 9th in traffic in the Netherlands, but 7th globally. The goal of this website is to provide the world with an encyclopaedia, and to achieve this, users are encouraged to add and revise content on the website. It is a typical example of a Collaborative Content Creation website, but displays signs of Media Provision as well. Metadata Generation and Social Networking are minor behaviours of the website.

3.3.5 Blogger.com

Google-held blogger.com is the leading pure weblog-site in the Netherlands, ranked 15th in the Netherlands and 9th worldwide. Social Networking site Hyves.nl.nl has blogging functionality as well, therefore it is unclear whether this website is the leading website for blogging. It is a Media Provision website, with Metadata Generation and Social Networking as minor behaviours of the website.

3.3.6 Facebook.com

The leading Social Networking website of the world (8th overall ranking) is ranked 18th in the Netherlands, where Hyves.nl is the leading social networking website. It is one of the fastest growing websites in the world (Alexa, 2009), and benefits from network effects. Beside Social Networking Behaviour, it displays Media Provision and Metadata Generation behaviour.

3.3.7 Flickr.com

Yahoo subsidiary Flickr.com is a photo sharing website, with recently added functionality to upload and share videos as well. The traffic rank of Flickr is 39 world wide, and 25 in the Netherlands. Flickr is displaying dominant Media Provision behaviour, but has strong Metadata Generation functionality in addition to that. There are also some Social Networking traces to be found on this website.

3.3.8 Dumpert.nl

Dumpert.nl, the media sharing subsidiary of the popular Dutch weblog Geenstijl.nl has surpassed Geenstijl.nl in popularity, becoming the nr 31 most visited website of the Netherlands. The goal of dumpert.nl is providing photos, videos and audio files to the Dutch audience. The website shows signs of Metadata Generation behaviour.

3.3.9 IMDB.com

The Internet Movie DataBase (IMDB.com) is currently ranked 32nd in the Netherlands and 29th global. The goal of the website is to summaries, reviews, cast lists and theatre schedules to the general public. The website also has an extensive rating and recommendation scheme, which makes the focus on Metadata Generation, but Media Provision is prominently present as well. There are also some traces of Collaborative

Content Creation behaviour on the website, such as the option to update plots and upload posters.

3.3.10 Myspace.com

The Social Networking site myspace.com has a focus more on music than other social networking site. This focus on a specific target group has brought myspace to the nr. 6 position in the world, outranking Facebook, but in the Netherlands, the website is ranked 35th. The website is a social networking website, but with strong hints of Media Provision behaviour and some Metadata Generation behaviour.

3.3.11 Photobucket.com

Photobucket.com is at the moment the leading photo sharing website of the world, placed 26 among the most visited websites. In the Netherlands photobucket.com is 38th, well behind competitor flickr.com. It is a Media Provision website with hints of Metadata Generation and Social Networking as well.

3.3.12 Web-log.nl

Web-log.nl is a website for blogging, which is currently ranked 39th in the Netherlands. It is mainly a Media Provision website, with some hints of Metadata Generation and Social Networking.

3.3.13 Ebay.com

Ebay.com is ranked 43rd in the Netherlands, and ranked 18th worldwide. It is a trading website, and displays some Media Provision behaviour and smaller signs of Metadata Generation behaviour. It displays, along with competitor Marktplaats.nl the least Web 2.0 behaviour of the research population. Note that Marktplaats and Ebay are owned by the same company.

3.3.14 Wordpress.com

Ranked 49th in the Netherlands, it is the last website of the top 50 in the research population. It is ranked 36th in the worldwide traffic ratings.Wordpress.com is another blogging website, which displays mainly Media Provision behaviour, although traces of Metadata Generation and Social Networking are found on the website.

3.3.15 Zoover.com

Zoover.com is an internationally oriented website for holiday ratings and reviews. The interesting rating system makes this website a typical Metadata Generation website. The traffic rank is below 100,000. The Dutch counterpart, zoover.nl is ranked 125th among websites in the Netherlands. The main competitor is Tripadvisor.com. It was added to the research population because of the interesting multi-criteria *Rating Engine*.

3.3.16 Boardgamegeek.com

Boardgamegeek.com is a website which collects ratings and opinions about board games, and recommends games to users. As such it is classified as a Metadata Generation website, although it displays strong hints of Social Networking as well. It is ranked 7,925th world wide. It was added to the research population because it has a

strong focus on a special interest group, and has an advanced *Recommendation Engine*.

3.3.17 Last.fm

Last.fm is a music oriented website, which displays the most hybrid behaviour of all websites in the research population. It is classified as a Metadata Generation websites because of extensive rating, tagging and recommendation functionality on the websites. However, users might use it solely as a radio station, using the website in a Media Provision way, or users might use Last.fm as a tool to compare their musical tastes to their friends, using it in a Social Networking way. Finally, the website displays Collaborative Content Creation behaviour in the biographies of the artists. It is ranked 321st in traffic among websites worldwide. It was added to the research population because the information objects provided (songs) are subject to multiple consumptions of the same information object by the same user.

3.3.18 Digg.com

Metadata Generation website Digg.com is a recommendation website, built around the idea that users can indicate which websites they like, and thus provide *Impersonal Recommendations* to other users, and obtain *Personalized Recommendations* if they have a general taste. It is ranked 243rd among websites world wide.

3.3.19 Uncyclopedia.info

Uncyclopedia, "the content-free encyclopaedia" is a parody on Wikipedia. However, it turns out to be rather successful and is ranked 18,280th among all websites. It is a Collaborative Content Creation website which aims at bringing fun to the information consumers and at creating funny information objects. It has traces of Social Networking and Metadata Generation behaviour as well. While conducting the research, this website was transferred to the more general Wikia-platform. It was added to the research population because it was one of the few Web 2.0-sites which has a focus on enjoyability.

3.3.20 Wikia.com

Wikia.com is a wiki-based special interest group website, aimed at both online and offline communities. Its business model is an advertisement supported encyclopaedia. It is ranked 296th among all websites. The company behind wikia.com is owned by Wikipedia founder Jimmy Wales. It is a Collaborative Content Creation websites, with traces of Media Provision, Metadata Generation and some social Networking. It was added to the research population because it provides an additional sample of Collaborative Content Creation websites, and this with a specific focus on special interest groups. For example, avid fans of the Pokemon franchise, can find and add detailed information on pokemon.wikia.com, which would be removed from Wikipedia.

3.3.21 Wikihow.com

Where Wikipedia rejects the more applied side of information, Wikihow.com is a how-to manual, based on wiki-principles. It was added to the research population for added comparability and new insights in Collaborative Content Creation behaviour.

Beside this, it also displays signs of Media Provision, Metadata Generation and traces of Social Networking. It has a traffic rank of 859.

3.3.22 Vimeo.com

The video sharing website Vimeo.com is a Media Provision website, which displays strong Social Networking and Metadata Generation behaviour as well. It is one of the more hybrid websites in the research population, and was added to the research population to provide an additional example of a video sharing website. Vimeo.com has a traffic rank of 530.

3.3.23 Icanhascheezburger.com

Icanhascheeburger.com is a website in the research population, since the information object that this website provides is an internet phenomenon. The internet phenomenon is called 'lolcats' and consists of a photo of a cat with a text in the font 'Impact'. Often this text is poorly spelled to parody the poor spelling of internet language. The website offers pictures of cats where other people can add their text to make it an enjoyable total. It is classified a Media Provision website, although it has strong hints of Collaborative Content Creation and Metadata Generation. It was added to the research population for an increased understanding of websites which focus on enjoyable information objects. It is ranked 4545th among the websites of the world.

4 Creating Patterns

In this chapter the format of the core of the research will be explained, being the created pattern language. A pattern is a form of documenting knowledge, in order to communicate phenomena among experts. (Gamma et al, 1995) It is, like a book or a scientific article, a form of written communication. Patterns are not written as single entities, but are part of a collection of patterns: A pattern language.

4.1 Introduction to Patterns

Patterns do not stem from the field of knowledge management. The theory of patterns was developed in the seventies by Christopher Alexander. He published the first patterns in books called: "A Pattern Language" (Alexander, 1977) and "The timeless way of building" (Alexander, 1979). Although he found response in the field of architecture, it was not until the discovery of patterns by the software engineering field that patterns really took off (Gamma et al, 1995). But where Alexander was concerned with living buildings [sic] and good places to live in, Software Architecture was more interested in a practical solution for communicating standard solutions in their field. Today, a shift is taking place towards the more intangible side of Information Systems, as is shown by Till Schuemmer (2005), where he analyzes patterns for social networking. This shift creates room for more focus on living and well-being instead of a practical way to communicate solutions.

4.2 Reasons to adopt patterns

Adopting patterns as a means of communication knowledge holds several advantages. The first is that patterns are strongly problem-solution oriented, and are considered an intuitive way to communicate standard solutions developed by experts. The second advantage is that patterns are easier to read than articles or books with a certain problem in mind. This is because patterns have standard forms, which include sections as 'Problem' and 'Solution'. If a certain pattern is not applicable to the problem at hand, the pattern can be skipped. The third is that patterns offer section which can be skipped if deemed irrelevant. On top of this, there are optional elements which can be included if the minimal form of the pattern is considered to shallow to cover the depth of the solution.

4.3 Creation of new patterns

The method used in this master project to collect patterns is a structured observation of websites, complemented with literature review. This process is iterative. Since a pattern which is mentioned in literature in adjacent fields of research, doesn't necessarily have to be in the field of information quality in Web 2.0. Therefore, viewing and reflecting on patterns from literature and searching for occurrences of these patterns on the websites has to be an iterative process. While performing the analysis, regular reflection by talking to users and specialist was conducted. To formalize this process, Grounded Theory was used as an approach to this.

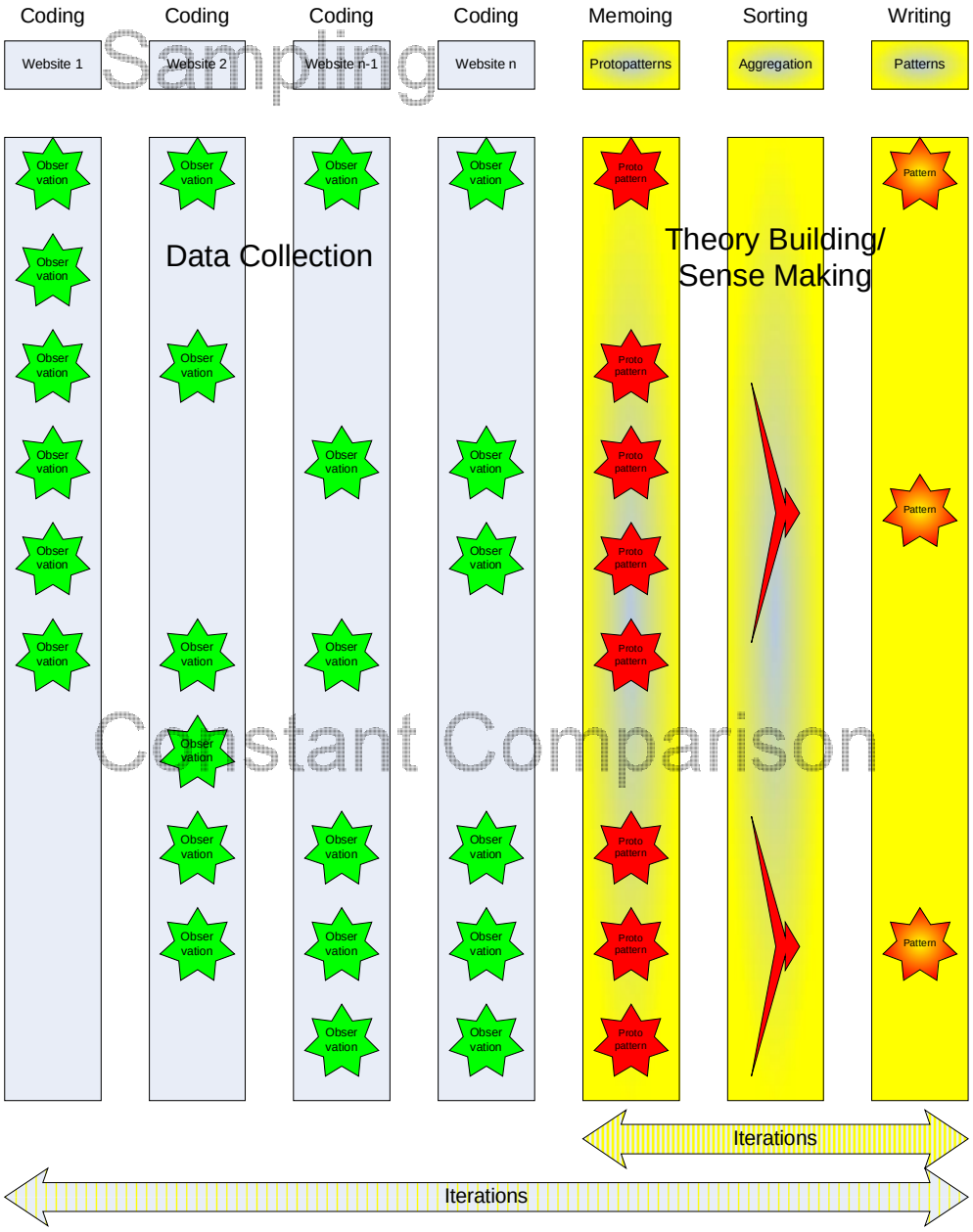

Figure 6: Theory creation process

4.3.1 Grounded Theory

The approach for the research was an iterative, reflective and creative process of discovery. For the definitive rooting of the research in a solid base of scientifically accepted practice, we have adopted the grounded theory approach. This is different from the majority of research that is conducted in the field of Industrial Engineering and Management, where the focus is often on the more application-related side of the theory. The basic Grounded Theory was invented by Glaser and Strauss in 1967 (Glaser and Strauss, 1967), emerging from the practices of psychology. Instead of testing a hypothesis, thus verifying a theory, this approach is focused on creating a

theory, which is hidden in the (qualitative) data. The main assumption is that if there are recurring words or patterns in an analysis, this might provide the basis for a theory hidden in the data. This theory hidden in the data, forms the basis of a possibly good theory. This is referred to as an 'emergent approach' [sic], referring to the fact that the theory follows the data instead of the data verifying the theory. The steps described here are an adoption of Bob Dick's: Grounded Theory; a thumbnail sketch (Dick, 2000). First described is the standard way, in chapter 4.4 is described how we adapted the Grounded Theory for research in Web 2.0 sites.

4.3.2 Constant Comparison

The heart of the process is constant comparison. More than an active process which can be started and finished, it is a mindset, a reflection which has to be part of the nature of the researcher, in order to make sense of the observations. That reflection on observations has to be conducted permanently, in line with the other observations made.

4.3.3 Data Collection

Interviews are typically the main source of qualitative data in the Grounded Theory approach (Dick, 2000). However, this is not the only method; focus groups are also known within this approach (Dick, 2000). Informal conversation is also an accepted practice in data collection. No practices of inquiry are openly rejected in this approach, since all possible observations may uncover a theory that is hidden in the data.

4.3.4 Note taking

The first analysis a researcher conducts is note-taking. This means observing an interview, making small notes about facts that appear in this observation. In this process, one can identify categories of phenomena and the subsequent properties of these categories. The fact that the interviewer/researcher takes notes, instead of observing full transcripts of conducted interviews is more a practical guideline than a strong research ground. Parallel to Note taking, the researcher should be memoing important facts.

4.3.5 Coding

As the research continues, facts will be clustered to form patterns. To a researcher, this means to observe small blocks of information one at a time. While observing these notes, a researcher should ask himself: "What is going on here?", "What is it I am actually seeing?", "And how does this relate to the data collection earlier?" This will converge into one or several categories. With a category in mind, a researcher should be able to identify the different elements that tell something about that category. When new observations do no longer add anything to the knowledge about the category, the category is likely to be saturated.

4.3.6 Sampling

The sample taken in a Grounded Theory approach is always determined by the situation at hand. It is a form of ad hoc sampling However, there are no clear guidelines as how to select a sample. There is a goal that the research population should be as diverse as possible, as to identify a more comprehensive set of theories

and phenomena within the field of research. Dick (2000) mentions the fact that a sample, like a theory, should be emergent. This means that new research samples can be identified on an ongoing basis, as to test emergent theories. This verification step is called 'theoretical sampling' (Dick, 2000).

4.3.7 Memoing

In the process of memoing, a researcher writes memos about the possible patterns and phenomena that occur in the research. These memos are an emergent hypothesis about categories or the relationship among them. While coding uncovers the aspects of the theory that is hidden in the data, memoing uncovers the relationships among these aspects.

4.3.8 Sorting

The sorting is a process, through which the researcher integrates the pieces of the emergent theory and explicates implicit links between categories and between codes. It is what "puts the fractured data back together" (Glaser and Strauss, 1967), the process includes creative investigation on how each concept fits in and how it will carry forward in the cumulative build-up of the theory. It is an activity that needs to be done by which the researcher can create a structure for the emergent theory.

4.3.9 Writing

Writing up the final report is the last phase, in which the researcher needs to put words to the emergent theory that has been formed. A coherent story should have been formed by the time that the writing process actually starts. If the theory is not coherent, it might be advisable to go through another iteration of the process.

4.3.10 The role of literature

The role of literature in the Grounded Theory is emergent. This means that the need for literature can be determined as it becomes relevant. This does leave openings for criticism as to the relevance of the research in the light of the existing body of knowledge. But it also prevents from becoming biased and thus influencing the emerging theory. There is a second role for literature in Grounded Theory: it is a source of data as well, which can be used to form and enrich the emergent theory.

4.4 Implementation of the Grounded Theory for researching Web 2.0 sites

4.4.1 Data collection

The Grounded Theory had to be adapted on several points for this research. This is caused by the fact that the Grounded Theory was mainly developed as an interview analysis technique, whereas we needed a technique to analyze websites and to form a coherent theory from the observation of these Web 2.0 sites.
The Grounded Theory leaves room for different sources of data. Dick (2000) advises the researcher to conduct the first interviews as unobtrusive as possible. The same applies for the data that can be collected from observing websites. The interviewing of people was substituted by the observation of websites in this research. The observation of websites is done through use of the websites, mimicking the behaviour

of a regular user. This includes consuming the information objects, adding new information objects, altering information objects (only applicable in collaborative content creation settings) and registering as a user. These activities provide insight in the processes and procedures of a website.

4.4.1.1 Observatory Research

The part observatory research was conducted at all websites of the research population. This includes viewing information objects and the processes that were in place. Looking at algorithms used for sorting and recommending information objects was not possible, since these algorithms are typically intellectual property, which the websites will not share. In order to collect data from websites and to counter researcher-blindness, print-outs of the websites and their respective interfaces were made, after which we could index and categorize all elements on the website. A simple website like Wikipedia or Youtube.com has already numerous elements in only their main site. Typical over 50 elements which can be identified as having a contribution to information quality or social networking, and several objects aimed generating revenue.

4.4.1.2 Participatory Research

This variant of research was conducted at Wikipedia. The task here was to join, to add some useful things to Wikipedia, and to become a member, accepted and respected by the majority of the active members. As such, the researcher became a member of Wikimedia The Netherlands, the organization behind Wikipedia. This gave him the opportunity to visit the Dutch Wikimedia conferences, attend lectures and most of all; talk with specialists on the topic of Web 2.0 and information quality. In this context he spent over 80 hours as a 'Recent Changes Patrol', someone who monitors recent changes. A patroller (a 'wikignome') takes corrective action if needed and passes a warning to malicious editors. Also he welcomes new editors with constructive contributions and corrects spelling and grammar errors.

It did occur that vandalism is often done by anonymous users, on typical high school subjects (geology, biology, history). However, the verification that these subjects are indeed the prime target of vandalism was out of the scope of this research. Another point of interest is that policies, methods and processes used to improve information quality are publicly available. When analyzing these methods they appeared very good with respect to the criteria objectivity, verifiability and accuracy. This is an advantage over other encyclopaedias, and would be the proof needed to conclude that Wikipedia was indeed outperforming other encyclopaedias which lack transparency in these processes. But problems were encountered in the enforcement and adherence to these rules. Heavy contributors and established users started displaying aristocratical behaviour. Yet, if they are committed to the cause and believe in free information for everyone, they create high quality information objects.

4.4.1.3 Informal conversations with users

Conversations were conducted with users, mostly friends and acquaintances, in different stages of the research. All users were familiar with both Wikipedia and Youtube.com, as well as other websites in the research sample, but usage differed. Some were registered users, but most were anonymous users and only consumed the

information objects. It appeared that users could quite easily grasp the concept of different criteria of information quality. But users were found unable to adequately track back an information quality criterion to a real life problem. E.g. if users were asked to give an example of an Accuracy problem, they would often come up with a Completeness problem.

This type of research was also conducted after questionnaires, as described in 3.1.5, if respondents would provide unexpected answers to seemingly unambiguous questions. In this way, the informal conversation provided insight in the interpretation of the questions asked, or the usage of the website. Sometimes phrasing and wording of the subsequent iteration of the questionnaire could be improved after informal conversation. It also provided the insight that users who mainly use a single functionality of a hybrid website (e.g. last.fm or imdb.com) will describe the behaviour displayed by that functionality as dominant, and other behaviour as present.

4.4.1.4 Informal conversations with specialists

Specialists were consulted through face-to-face communication, presentations and email. This includes the advice offered by the mentors at the university. Informal conversations were often the source for new insights and classification. Although these conversations were intended mainly at gathering new data and providing discovering new occurrences of proto-patterns, it proved valuable as well to categorize and structure the emergent theory. As such these conversations took place in both the Data Collection and the Theory Building/Sense Making part of the research.

At several occasions, highly experienced users from the Wikipedia community were contacted during the Dutch 2008 Wikimedia Conference and the informal Wikimedia meetings. These informal conversations provided insight in most of all the social systems underlying Wikipedia. These conversations had their value mostly in the Data Collection part of the research.

4.4.2 Note Taking

In this research note taking was found to be overlapping with memoing. This was caused by the fact that the there were no interviews conducted, but websites observed. Where notes are used to capture the information in interviews in a short time span, websites can be visited multiple times. Therefore, in this research the function of notes could be replaced by renewed observation of websites, and direct memoing. This process created a log, in which all observations were recorded.

4.4.3 Coding

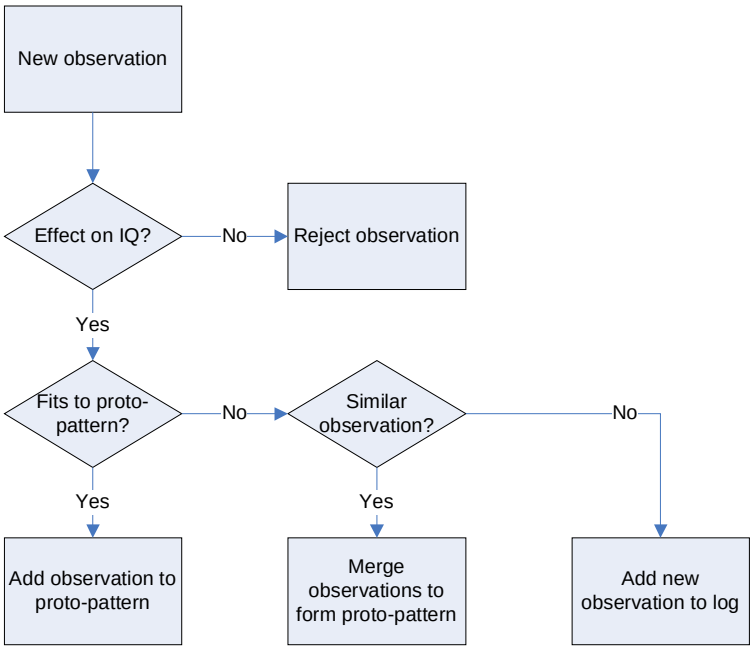

Figure 7: Coding Process

⇒ The rectangles represent processes.
⇒ The diamonds represent decisions.

Since Note Taking was replaced by renewed observation and direct memoing, Coding became more important. Coding took place after a new observation was made on a website. If an observation was similar to an existing proto-pattern, the observation could be added to the existing proto-pattern. Else, coding was conducted by creating small proto-patterns, consisting of an observation and the presumed effect it has on a particular criterion of information quality. Or, if a similar observation was present, but not yet a pattern, the two observations could be merged and be enhanced to form a proto-pattern.

4.4.4 Sampling

The first sample taken was Wikipedia. Due to its highly transparent nature, high traffic and strong focus on user-generated content, it was an ideal first research subject. The second website to be observed was Youtube.com. Being the Web 2.0 site with the highest traffic at the time of research, this site has a strong focus on user generated content as well. The next sample was the social network site Hyves.nl, which is one of the top sites in the Netherlands. This was because social networking was found to be one of the main issues alongside information quality. Further samples were added in an iterative way, referring back and forth a way which cannot be conducted with regular interviews. This additional freedom allowed for more variation. The research sample included all websites which could be indexed as Web 2.0 from the top 50 websites as measured by traffic for the Netherlands, as described in Chapter 3. For additional comparability similar sites were added in order to create more comparable observations. These are summarized in Table 4, as found in Chapter 3. These are also the top-websites for Web 2.0 in the Netherlands.

4.4.5 Resampling for single observations

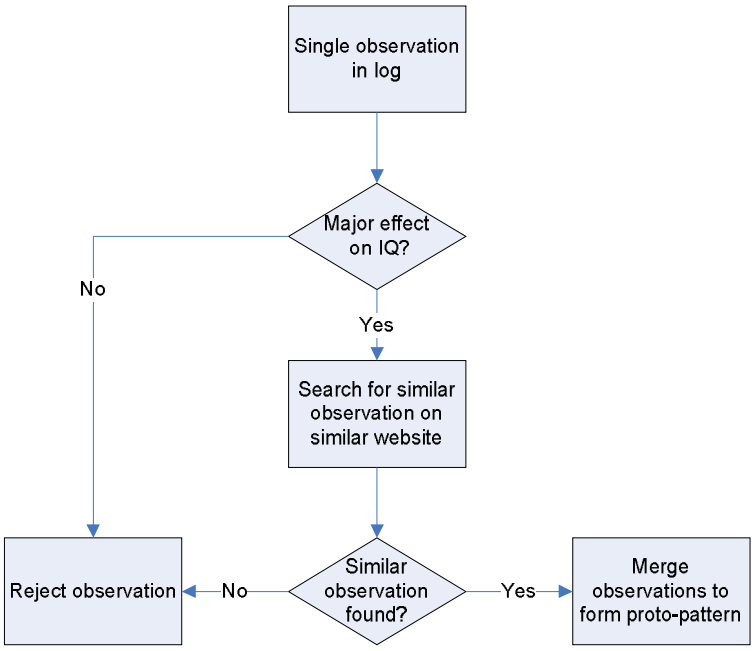

Figure 8: Sampling for single observations Process

⇒ The rectangles represent processes.

⇒ The diamonds represent decisions.

Besides the first stage of sampling, in which the Web 2.0 sites of the top 50 websites were researched and analyzed, there was another type of ad hoc sampling. When an observation was not found on more than one website, attempts were made to find this particular behaviour on other websites, resulting in more websites added which employed the same methods of information quality improvement. This websites were often quite similar to the websites on which the first observation occurred. Websites as wikihow.com were added to the research sample to create additional proto-patterns from observations which had no similar observation in the research sample up to that moment.

4.4.6 Memoing

Instead of using cards, as Dick (2000) proposed, in this research a digital log could be used to write down thoughts and notions about the possible relations between observations and categories. This Direct Memoing is feasible because websites can be accessed multiple times, and are very patient in comparison with humans. This log also took the form of a proto-pattern language, which was a collection of observations aggregated in proto-patterns of behaviour of websites to improve information quality.

4.4.7 Writing

The writing was more gradual than proposed in the original Grounded Theory. This was caused by the fact that the researcher always had a computer at his disposal, and could switch from research to writing and back more easily than when interviews were conducted. This is presumably an improvement to the Grounded Theory, but is only applicable when the research subject is not human, and thus not offended by quick shifting between research and writing.

4.4.8 The role of literature

Several types of scientific literature were accessed and incorporated in the research. As in the Grounded Theory approach is described, literature was accessed as it became relevant to the research. The literature sources can be divided in three categories:

1. Research-topic related: Articles concerning information quality and the web. e.g. (Knight & Burn, 2005)
2. Report-form related: Articles concerning pattern languages. e.g. (Meszaros & Doble, 1996) and Pattern languages concerning other subjects. e.g. (Alexander et al., 1977)
3. Methodology related: Articles and books concerning research methodology. e.g. (Allan, 2003)

Most notable was Till Schuemmer's (2005) work on social networking. As it emerged, it was clear that social networking and information quality methods are bordering each other. For the construction of a framework on information quality, a thorough literature review was conducted, whereas for the construction of the pattern language, literature was accessed in a more iterative fashion. Literature served as both source for possible patterns in the Data Collection part, as well as a guide to structuring in the Coding and Sorting phases of the research. Literature concerning the Grounded Theory was accessed to improve the understanding of the theory.

4.5 Standard format of patterns

Patterns in literature have been published in different forms, shapes and sizes. This is caused by the different areas of application in which patterns are used. For the transfer of knowledge in the field of information quality in Web 2.0 environments, adaptations to the format of Till Schuemmer (2005) were made in an early stage. As the research progressed it became clear that simple adapting another format would not work all the way, so a new standard format was created. For this the guidelines of the Hillside group were used. (Meszaros & Doble, 1996) These guidelines insist on a division between mandatory elements and optional elements when needed. Therefore, additional elements may appear in the patterns. The mandatory elements are briefly described below. The first 5 elements of a pattern form brief description which should give a sufficient impression whether the pattern is applicable in a specific situation.

4.5.1 Name

The first mandatory element is the name of the pattern. Since it is present in a pattern language, it is often represented as a noun, and as evocative as possible, in accordance with Meszaros & Doble. (1996) It is always accompanied by a picture or a graph, which should create additional insight in the pattern.

4.5.2 Alternative names

The name of the pattern is chosen with care, but not the only name possible. This section describes the possible alternative names of this pattern, found at other authors or made up to allow more diverse communication.

4.5.3 Intent

The intent of the pattern describes what result an implementation of this pattern should achieve. In this research all patterns directly or indirectly aim at improving information quality. It contains a short description of the way in which this pattern improves information quality.

4.5.4 Context

The context in which this pattern is applicable. This includes the implemented patterns and the nature of the site. This nature can be Collaborative Content Creation, Media Provision, Metadata Generation or Social Networking. It is more or less comparable to the applicability section in 'Design Patterns'

4.5.5 Addressed Criteria

A list of the information quality criteria addressed by this pattern.

4.5.6 Problem

The problem to which the pattern is supposed to give a solution.

4.5.7 Solution

The solution-section describes the general solution to the problem.

4.5.8 Rationale

The reason that this pattern will improve information quality, or resolve the problem at hand. This section also includes assumptions made to the functioning of the pattern.

4.5.9 Known Uses

This section includes the observations made which lead to this pattern. It includes a website on which the pattern was found; and than a description of how this website has implemented this specific pattern.

4.5.10 Related Patterns

This section includes other patterns which relate to this pattern. This can be refinements or more general patterns. It describes how patterns relate to each other, as to make it a coherent pattern language. (Meszaros and Doble, 1996)

5 Patterns for Information Quality in Web 2.0

In this chapter the patterns for Information Quality in Web 2.0 are elaborated, describing the problems these patterns address and the solution that is given by these patterns. These patterns form a pattern language, which means that a pattern is seldom a single solution, but has relations to other patterns. The picture below displays the patterns and their respective relations.

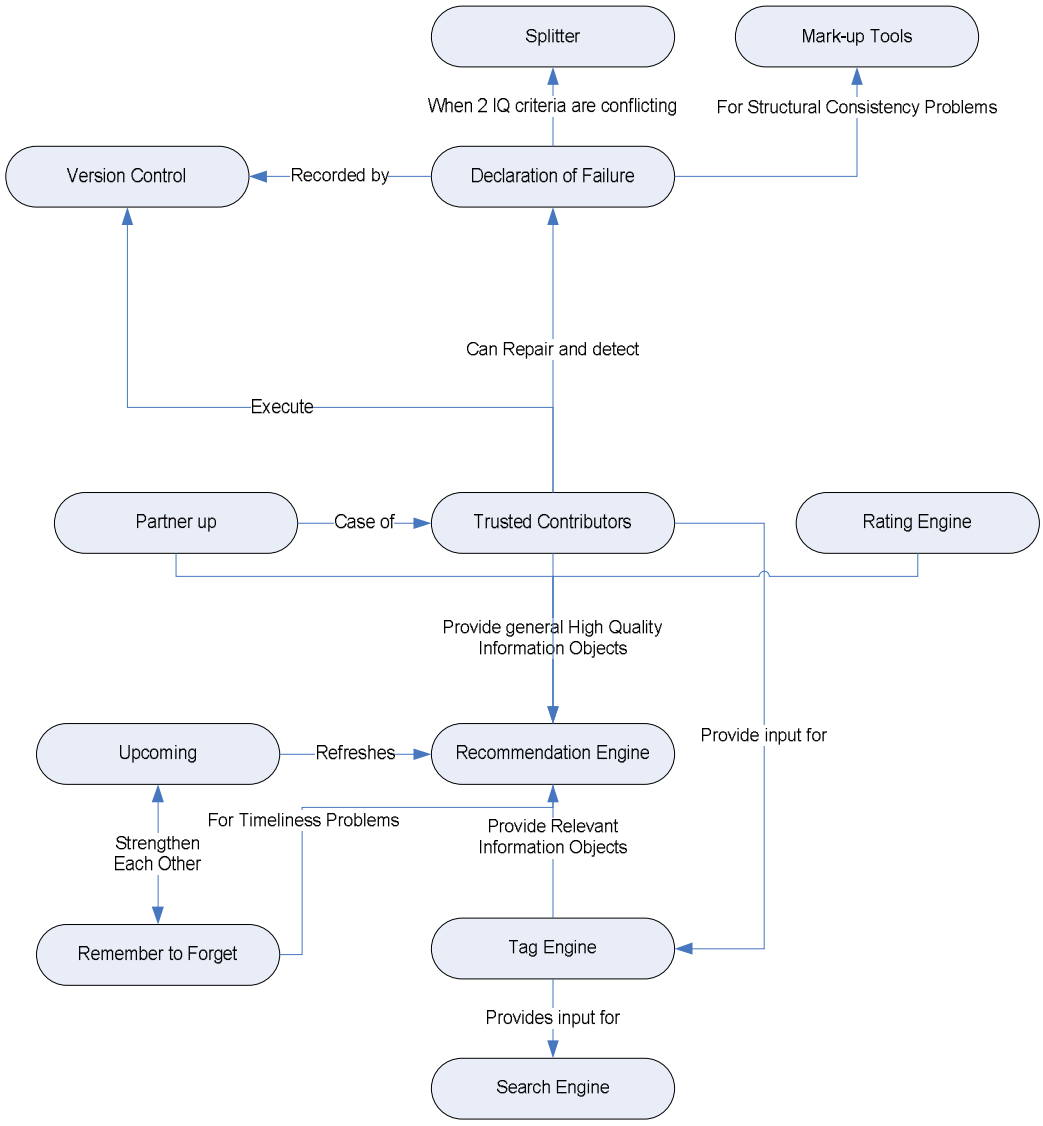

Figure 9: Relations among patterns

⇒ Rounded rectangles represent patterns.
⇒ The arrows and between the patterns represent relations between patterns.
⇒ The text accompanying the arrows represents the cases in which there is a relation between the patterns.

Patterns are abstract solutions to abstract problems. But more specific implementations may occur. Figure 10: Implementations and Refinements of Patterns depicts the patterns with their respective Implementations and Refinements.

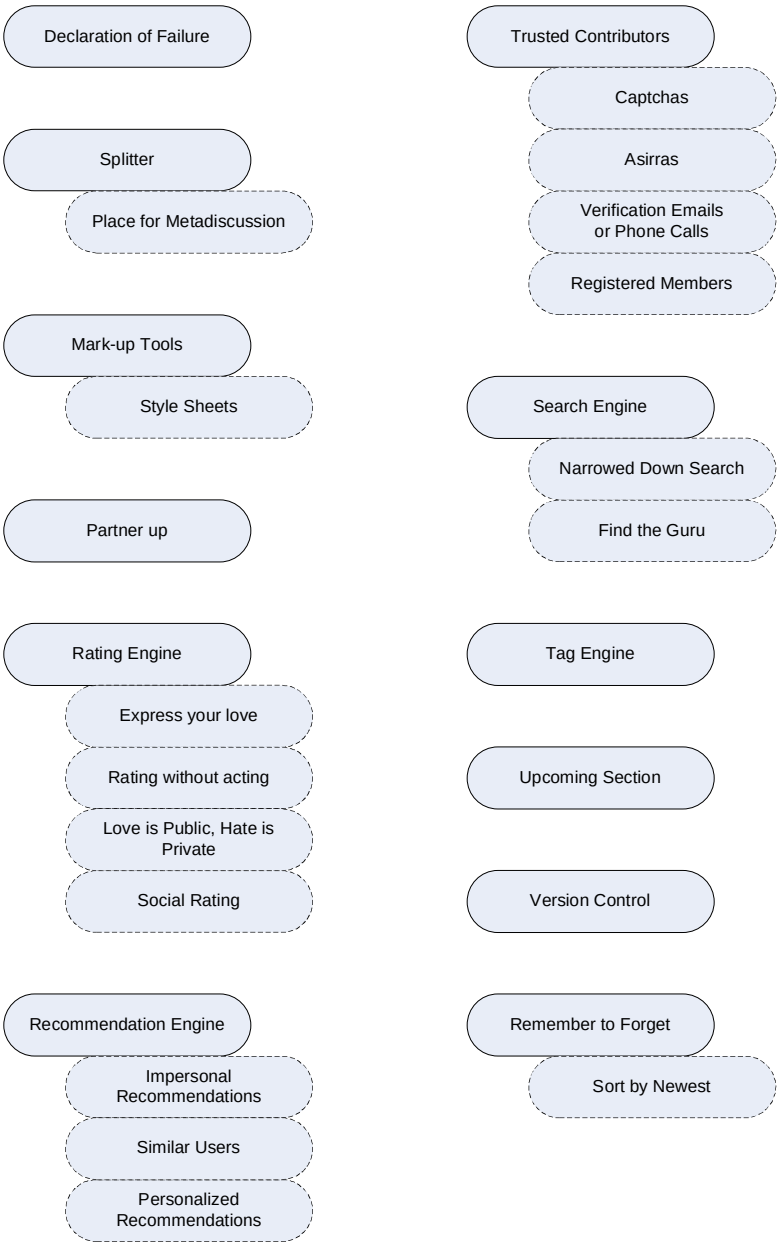

Figure 10: Implementations and Refinements of Patterns

⇒ Rounded rectangles represent patterns.

⇒ Rounded rectangles with dotted lines represent the Implementation and Refinements of their respective parent patterns.

5.1 Declaration of Failure

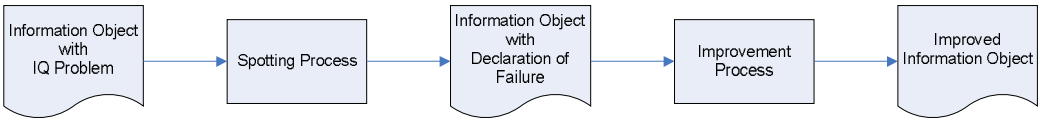

Figure 11: Process of *Declaration of Failure*

⇒ This graph depicts the general process that is used in a *Declaration of Failure* process.
⇒ The rectangles represent processes
⇒ The rectangles with a gulf underside represent information objects
⇒ The arrows represent the process flow.

5.1.1 Alternative names

Detection-Solution Process

5.1.2 Intent

The intent of this pattern is first to declare that the information repository underperforms on a certain criterion of information quality. Depending on the implementation of this pattern, the information repository can either request an improvement on this criterion, or redirect users to other information objects that perform better on that aspect.

5.1.3 Context

This pattern is particularly useful in Collaborative Content Creation-environments, and can be used almost any aspect of information quality. It can also be used to encourage Metadata Generation.
When a strategic focus on certain information quality criteria has been chosen, one can encounter information objects in the information repository which are insufficient on one or some information quality criteria. It might be advisable to delete these information objects which underperform on too many criteria, but preserve those information objects which underperform on only one or some criteria.

5.1.4 Addressed Criteria

This pattern is widely applicable, on all the aspects of information of Information Quality, except Relevancy and the Process-pragmatic criteria. Which aspects are indeed addressed is fully depending on the implementation of the pattern.

♠ ♣ ♥ ♦

5.1.5 Problem

There are many possible ways to create an information object which refers to a certain real world object. But these many ways do vary a lot in their performance on the different aspects of information quality. If there is a problem with an aspect of information quality in an information object, this is the most generally applicable solution.

5.1.6 Solution

First there has to be a way to spot and identify the information quality problem. The spotting can usually be done by either the community supporting the website (volunteers) or by the visitors of the information object (users). Sometimes the improvement can be done by the spotter himself, as is typically the case with spelling errors. But when the spotter cannot improve the information object, they need to have tools to mark the information object as failing on an aspect of information quality. This is a *Declaration of Failure*. There should be different types of *Declaration of Failure*, of course for the different information quality criterions, but also ranging from small parts of information objects to complete (groups of) information objects.

When a (part of an) information object is marked with a *Declaration of Failure* there are several ways to improve the overall user experience.
First, if the information object is not performing well on any characteristic of information quality, the easiest way is to simply delete the information object. Than the *Declaration of Failure* should be a redirect to other recommendations.
Second, if the information object is performing fairly well on the important aspects, but is clearly lacking on one of the aspects within the strategic focus of the information repository, one can ask for an improvement in the information object itself. In that way a volunteer or a user is stimulated to improve the information object. The other possibility is that the information object performs poor on one criterion, but inside the repository is already another information object that performs better on that criterion. Then the information repository can offer a redirect to the other information object.
Third, if the information object is performing fairly well on the important aspects, but is underperforming on one of the aspects outside the strategic focus of the information repository, one can ask for an improvement in the metadata (for example, a discussion page.) An alternative is to redirect them to a different information repository which has a focus on that specific criterion. If that other information performs badly on that aspect as well, the redirect can become a request for improvement in itself.

5.1.7 Rationale

This pattern improves information quality by creating higher quality information objects. The reason that this pattern will improve information quality is by the assumption that the general public is better in improving all aspects of information quality than a single user. The other assumption is that users who are passing by are sometimes willing to contribute to an information object. It also acknowledges the fact that someone who spots the information quality problem is not always able to solve the problem. Based on this, a *Declaration of Failure* is at the same time a request to all users. Another good aspect of this approach is that it, so far, does not attract any increased vandal activity.

5.1.8 Known Uses

5.1.8.1 Wikipedia

Wikipedia is the clearest example of the usage of this pattern. Which is not that strange, as it is also the most visited website in Collaborative Content Creation. On

http://en.wikipedia.org/wiki/Wikipedia:Template are numerous examples of Templates used as a *Declaration of Failure*.
When a [Citation Needed]-tag is added, than is that a typical *Declaration of Failure*, on the criterion of Verifiability. Someone spotted a statement that needs backing by reference; the [Citation Needed]-tag is the tool that is provided. This tag urges the users and volunteers to add a possible reference to the statement made. Other implementations in Wikipedia include the [Clarify]-tag, which is a *Declaration of Failure* on the criterion: Understandability and the [Sic]-tag, which is a *Declaration of Failure* on the criterion: Semantic Consistency. (Spelling errors are a semantic consistency problem essentially.) Also the notifications that a certain article doesn't live up to the Wikipedia standards of neutrality are an implementation of this pattern. The last mention should go to the disambiguation pages; which are in essence a *Declaration of Failure: Unambiguous*, and in that a redirect to the less ambiguous pages which are in the information repository itself.

5.1.8.2 Uncyclopedia.org

At Uncyclopedia they approach things a bit different; they try to make fun of everything. Knowing that the entire information repository performs horribly on the aspect of Accuracy, they offer redirects to Wikipedia. Of course in the typical uncyclopedia-style, they sometimes have added a box to an article with the text: "For those without comedic tastes, the so-called experts at Wikipedia think they have an article about this subject." In that way, they provide a redirect to an article at an information repository, which performs better on the aspect of Accuracy.

5.1.8.3 Wikihow

The application counterpart of Wikipedia, Wikihow has its own strategic focus. They use much of the same set of *Declaration of Failure* as Wikipedia. On the site http://www.wikihow.com/wikiHow:Templates/Article, Wikihow has numerous options presented of this pattern. It is a toolbox that anyone can use to signal the problems with information quality that could occur. It includes Declarations of Failure for: Accuracy (Accuracy); Attention (General IQ problems); Citation Needed (Verifiability); Clarity (Understandability, Unambiguous); Cleanup (Conciseness); Copyedit (Consistency, Semantic Consistency); Format (Structural Consistency); In Need of Pictures (Suitability of Representation); Introduction (Degree of Context); Personal References (Interpretability) Split (Informativeness) and Stub (Completeness). This acknowledges the fact that improvement on all these aspects of information quality is still manual labor; it cannot be done by machines or bots. Typical is that Wikihow.com does not include any references to Wikipedia.org, while the information repositories might both benefit from these links.
Wikihow has another interesting implementation of this pattern. At the bottom of each page there is a question asking: "Was this article Accurate?" accompanied by two buttons: Yes and No. In this way, they are actually pushing the *Declaration of Failure* on the aspect of Accuracy. At the same time, it is an implementation of the *Rating Engine*-pattern.

5.1.8.4 Youtube.com

Youtube.com also has an implementation of this pattern. When the metadata for a certain video is absent, they encourage users to add a rating to the video. In a certain

way, also the absence of any written reactions can be provoking to users, and they can add their reaction. This strange phenomenon is illustrated by numerous sites, where there is a culture of typing a reaction with only the content 'First' as being the first reaction on an information object.

5.1.8.5 Last.fm

Last.fm has a limited, but nonetheless interesting implementation of this pattern. The pattern occurs when a song doesn't link to a video (Youtube.com based). Then they actively request for videos to be added.

5.1.9 Related Patterns

5.1.9.1 *Splitter*

Splitter can be an implementation of this pattern, when it comes to one information object that can be split into two information objects on more or less the same subject. In that case, the problem was in conciseness or cohesiveness.

5.1.9.2 *Mark-up Tools*

These tools can be the solution to problems in the field of Structural Consistency.

5.1.9.3 *Recommendation Engine*

It can be a wise decision to leave information objects which have a *Declaration of Failure* present out of the *Recommendation Engine*.

5.1.9.4 *Version Control*

Declarations of Failure and the possible removal of problems can be recorded by a *Version Control* system.

5.2 Splitter

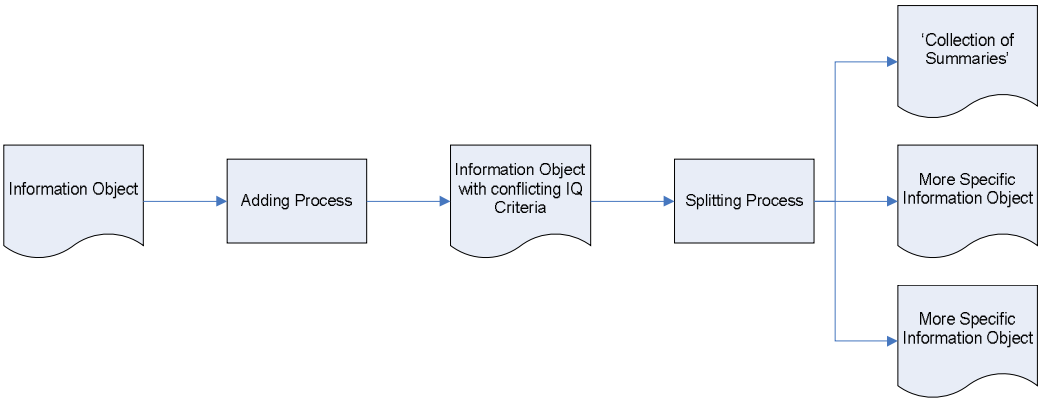

Figure 12: Process of *Splitter*

⇒ This graph depicts the general process that is used in a '*Splitter*' process.
⇒ The rectangles represent processes
⇒ The rectangles with a gulf underside represent information objects
⇒ The arrows represent the process flow.

5.2.1 Alternative names

Division Approach, Multiple Versions

5.2.2 Intent

Improving information quality, by splitting information objects which desire to score well on conflicting criteria of information quality.

5.2.3 Context

In a Collaborative Content Creation environment, users create and add information to an information object. But they might have a different opinion on what a good information object is. As such they will edit in different ways. This is not a problem, but enrichment for the information repository. However; there can be information quality criteria which conflict; a gain on one, results in a loss on the other.

5.2.4 Addressed Criteria

This pattern addresses pairs of criteria, which are depending on the implementation of the pattern. Most commonly it is Conciseness - Completeness. Other pairs include Cohesiveness-Cohesiveness, Timeliness – Accuracy and Fun – Objectivity.

♠ ♣ ♥ ♦

5.2.5 Problem

When users are encouraged to add their specific knowledge to an information object, the object will grow. Since knowledge is not divided in independent blocks, but the

knowledge is a body, with links and joints. Information objects are subject to meta-discussion as well. However, this great body of knowledge is too big to be practical.

5.2.6 Solution

Split the information object as it reaches a certain size. This can be a split in two, three, or even more. The division can be either in different topics, or it can be that there is an overview information object, with small summaries of the more specific information objects. An alternative approach is that an information object is split according to the different types of media that are present in the information object. This solution does not only hold for information objects, but for entire information repositories as well. By splitting information repositories, the focus on an information quality criterion or on a single media-type, can be enhanced.

5.2.7 Rationale

This pattern improves information quality by creating higher quality information objects. The rationale behind the *Splitter* pattern is fairly simple. It is not possible to have all criteria of information quality satisfied at the same time. There are certain criteria which are damaged when others are improved. The most typical example is conciseness and completeness, which have to be balanced to get to an informative information object.
Therefore, information objects (and even information repositories) have to be split into two or more objects, if one of the information quality criteria gets violated because of improvements on another criterion.
It is important that information objects are only split if both information quality criteria are important to the information repository. Otherwise, it might be better for information quality to split the entire information repository, so that the different information repositories have different focuses, and they do no longer conflict inside the information repository.

5.2.8 Known uses

5.2.8.1 Wikipedia

Wikipedia has numerous implementations of this pattern, ranging from templates to propose a split, to discussion pages, to the Wiktionary.
Splitting pages is most often proposed when the entry on a single information object grows too large to be useful. Than some part is proposed to be split into a separate information object, which is than left in the overall information object with a small summary, with a link to the more complete information object. This is typical the case with great historical events and countries, as is depicted in the picture below.

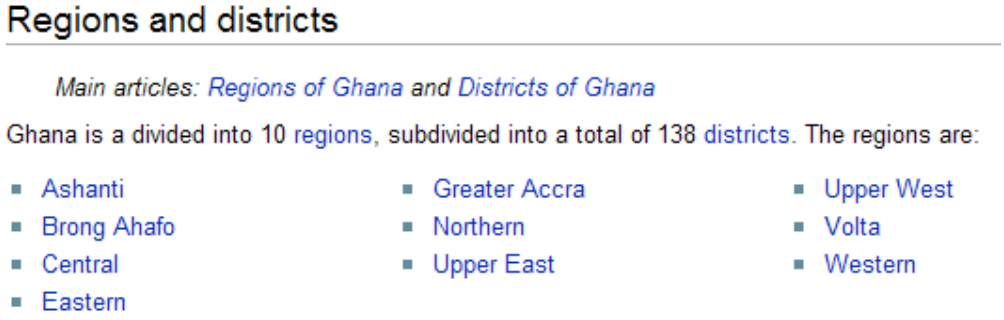

Figure 13: Example of *Splitter*

However, there are more implementations of this pattern. For the 'dictionary definitions' of words there is no room in Wikipedia. However, the entire repository is split; and there is the 'Wiktionary', in which there is a place for such definitions. Of course, the different languages in which the online encyclopaedia is available are an implementation of the *Splitter*-pattern as well.

5.2.8.2 Wikia.com

Wikia.com is a typical site which gets information out, and creates an own repository for it. Typical these are collections of information objects which are rejected by Wikipedia, where these were deemed 'not worthy of a place in an encyclopaedia', as often happens with fictional figures in television series. But, behind these television series there is often quite a fan community, with ample power to build an encyclopaedia. Wikia.com grants a place for these communities to build an encyclopaedia on that particular subject.

5.2.8.3 Youtube.com variants

The adult variants of Youtube.com (which are not included in the research further) are in fact implementations of this pattern. Those websites are in itself not a true split from Youtube.com, but have emerged on the fact that Youtube.com simply rejected the content. Thus there was a Split, but Youtube.com did not provide a place for the rejected content. So these other sites saw the gap in the market and emerged on that business.

5.2.9 Implementations and Refinements

5.2.9.1 *Place for Meta-Discussion*

An often found phenomenon is the *Place for Meta-Discussion*. It is a place where disputes can be solved and user can ask for elaborations on an information object. In Media Provision environments, all reactions are meta-discussion, but in Collaborative Content Creation environments, active division between the main information object and the meta-discussion is advisable. When a website has *Place for Meta-Discussion* is not regarded as an active implementation of the *Splitter* pattern in the research.

5.2.10 Related Patterns

5.2.10.1 *Declaration of Failure*

Splitter can be a solution to the *Declaration of Failure*, especially when conciseness is the information quality criterion which is violated.

5.3 Mark-up Tools

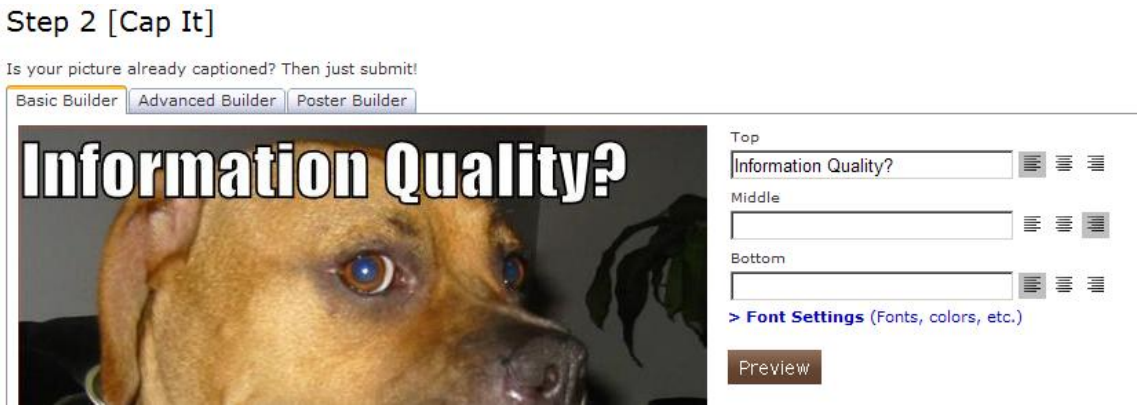

Figure 14: Example *Mark-up Tools*

5.3.1 Alternative names

Style Sheets

5.3.2 Intent

Creating a more cohesive and standardized lay-out of information objects, not by forcing contributors to make something look in a certain way, but by providing the most easy and user-friendly *Mark-up Tools.*

5.3.3 Context

This pattern is only applicable in environments in which users have to invest manual labour to contribute an information object to the information repository. In Collaborative Content Creation environments and in Media Provision environments, users are free to create and add content. When the users are diverse, it is rather hard to make a uniform total out of it. Users have different styles, different quality measures and different possibilities with their computer programs. This results often in highly diverse spectrum of information quality and information style.

5.3.4 Resulting Context

A more uniform lay-out of the information objects, but with enough freedom to enable the diversity needed for an information repository. This more uniform lay-out should improve the focus of the information consumer on the information itself, since the lay-out is no longer a disturbing factor in the information repository.

5.3.5 Addressed Criteria

Structural Consistency

♠ ♣ ♥ ♦

5.3.6 Problem

When a lot of users are free to create and add content, it is difficult to make a uniform total out of it. Users have different styles, different quality measures, different possibilities with their computer programs. This results in a lot of different styles used, thus damaging the structural consistency as is often a trademark of well-built sites. Also, it indirectly damages understandability and usability of the information objects.

5.3.7 Solution

Offer *Mark-up Tools* to the end-users. But make sure that these *Mark-up Tools* are pretty rigid, in order to ensure the Structural Consistency. If too much functionality and flexibility is provided within the mark-up tool, it is at risk to have little effect on Structural Consistency after all. *Mark-up Tools* can also be provided in the form of *Style Sheets*, which users can use afterwards, as they apply these *Style Sheets* to certain information objects.

5.3.8 Rationale

This pattern improves information quality by creating higher quality information objects. Users who are willing to contribute to the information repository are usually well intended users. Thus they are willing to invest a little effort in adding higher quality information objects to the repository. So if easy-to-use tools are offered to this group of contributors to enhance the Structural Consistency, they will be likely to use these tools. If the users aren't willing or able to use the *Mark-up Tools*, then some *Trusted Contributors* might be willing to apply these tools to contributed information objects.

5.3.9 Danger spots

Mark-up Tools might draw a lot of resources from a server, thus damaging Accessibility of all other information objects.
A too good mark-up tool might get abused for purposes it was not intended for.
A too rigid mark-up tool can damage the information as it is provided by the contributors, since they might need more freedom than the mark-up tool allows.

5.3.10 Known uses

5.3.10.1 Icanhascheezburger.com

Icanhascheezburger.com and their sister sites, (ihasahotdag.com, graphjam.com, and even failblog) offer standardized *Mark-up Tools* for their captions. It is simple, a contributor can add text on 9 places (top, middle, bottom – left/centre/right-aligned) and that is about it. The font is already standard in trademark font, the colour is white, and the photographs to capture are already on the server. Since most contributors use these *Mark-up Tools*, the Structural Consistency among the information objects they provide is high. This is not only the fact within the different sister sites, but even among the different sister sites.

5.3.10.2 Youtube.com

Youtube.com has recently added the possibility to add small grey blocks with text to the movies submitted. These small grey blocks are foremost an additional feature, a

service to the contributors. But the fact that the blocks are only grey blocks, and the font, and font size is the same throughout all movies, makes these blocks recognizable as added comments, using the Youtube.com-interface.

5.3.10.3 Wikipedia

Wikipedia has *Style Sheets* for the different standard categories, like geographical places, birds, bands and chemical elements. These *Style Sheets* impose a coherent lay-out on the information object, with the additional benefit that pretty dry information can be displayed in a concise way. Without taxonomy boxes for example, there would have to be lengthy sentences explaining which class, order and family a certain bird would fall in to.

5.3.10.4 Flickr.com

This photo and video sharing community has several *Mark-up Tools*, of which the most striking is the collaboration with Picnik. Because of this, there is the possibility to rotate, crop and resize the photos. Furthermore a user can adjust the exposure, colours and sharpness of the picture. There is also a functionality which reduces red eyes and an autofix-functionality which attempts to make the overall photo quality better. For the more experienced user, there are advanced functions, which include several premium functions for which the user has to pay. Main competitor Photobucket.com has similar functionality, but provided by Fotoflexer.com. Another mark-up tool that Flickr provides has to do with comments that are given on a photo. It allows users to draw simple rectangles in an additional layer on the photo, and to attach comments to it. It is a tool for highlighting details of a photograph.

5.3.10.5 Vimeo.com

This video sharing site uses an approach to *Mark-up Tools* which doesn't only allows editing the information object, but also enables protection and sharing options. In that way, it is a mark-up tool for meta information as well. In the uploading process they have incorporated the tag-engine as well as a tool to give credit to the other users of Vimeo, who have cooperated to create the video.

5.3.10.6 Artisteer/Wordpress.com

The Wordpress-platform allows for much diversity, by giving users the freedom to upload their own themes. The Artisteer company has created a tool to create these themes without a need for programming skills. In this way, this company has created a Mark-Up Tool to create *Style Sheets.* It is interesting that Artisteer claims that it is not endorsed by Wordpress, or any of the other web log platforms that Artisteer provides *Mark-up Tools* for.

5.3.11 Implementations and Refinements

5.3.11.1 *Style Sheets*

When *Mark-up Tools* are used in a text-only environment, the name *Style Sheets* is more appropriate, since it is not really What You See Is What You Get, but the plain text is marked up afterwards.

5.3.12 Related Patterns

5.3.12.1 *Declaration of Failure*

In Collaborative Content Creation settings, when *Mark-up Tools* can be applied after an information object has been created, it can be used in a *Declaration of Failure: Structural Consistency.*

5.4 Partner Up

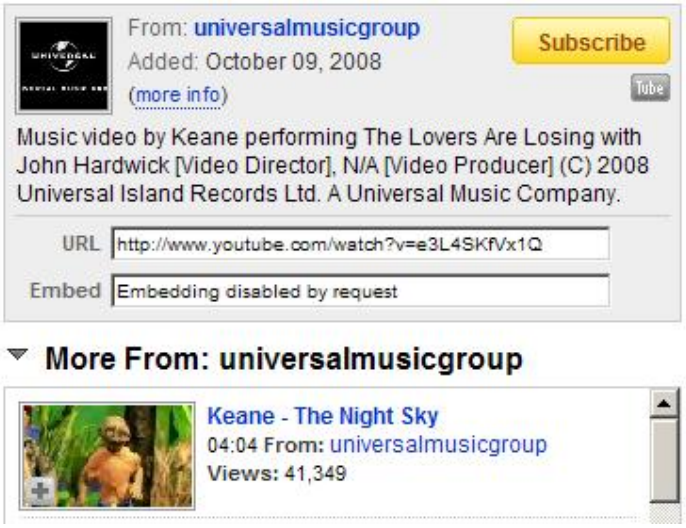

Figure 15: Example *Partner Up*

5.4.1 Alternative names

Professional Contributors, Contributor for Hire

5.4.2 Intent

Assuring information quality by partnering up with well-known content providers, so that the content they deliver is of high quality.

5.4.3 Context

In a Media Provision website or Meta Data Generation website, all users offer information objects to the general public. When a lot of users can deliver content, overall quality will be lower (this is caused by the problems as described in Andrew Keen's 'Cult of the Amateur'). This might drive away the information consumers, and have a negative effect on the success information repository.

5.4.4 Addressed Criteria

In general, all aspects of information quality can be addressed, if an expert would do better in creating the information object than an amateur. Reputation and Reliability are the most prominent among these, as they are considered to be connected to the author.

♠ ♣ ♥ ♦

5.4.5 Problem

When a lot of users can deliver content, quality will be lower. This might drive away the real good creators of content, thus reducing information quality in the information

repository. This in turn might lower the traffic drawn to the website, and as such reduce the attractiveness.

5.4.6 Solution

Partner Up with the professional content creators. Offer them an advantage (e.g. promotion) in exchange for the content they can deliver to the knowledge repository. Some professional content creators may be satisfied with a brief mention, that they are credited for the information object. Others will not work for less than a financial compensation.

5.4.7 Rationale

This pattern improves information quality by creating higher quality information objects. The reason why this pattern works is because there are specialists in the field which can create high quality information objects, and there are amateurs, which cannot create such high quality information objects. These specialists are often found working in the field, creating information objects with a clear commercial intent. They can be individuals, employees, freelancers or scientists. They can also be companies or collectives which can create high quality information objects together. When such a specialist creates an information object, the quality is expected to be (much) higher than when an amateur creates an information object.

5.4.8 Known uses

5.4.8.1 Youtube.com

Youtube.com has an extensive partner program. They *Partner Up* with e.g. Warner Music, and all other major labels. The cooperation between the companies was under heavy debate during the writing of this report, yet it is still an observation of a known use. Nonetheless, Youtube.com still has this form of cooperation with the other major music labels. It is a flower blooming on the dung pile of the pirate industry. Instead of constant battling over abused content, Warner Music now delivers content. This is a win-win-situation. On the one hand, Youtube.com always gets the first, high quality content on their website, without violating any copyright. On the other hand, Warner Music has a platform for promoting their new music. The partners have found a balance in this, by letting Warner promote new music videos whenever another video of them is watched. This is observable by the fact that, if a user watches a video; the "more videos from this provider" is visible, if one views a video from Warner, whereas it is hidden when the same user watches a video from an unknown content provider.

5.4.8.2 Google Knol

Google Knol beta is essentially a *Partner Up* program with the providers of content. Instead of letting anonymous users create and edit content, the creation of content is done by well-known specialists in the field. While any user can become a 'page owner' of a specific subject even if they aren't that notable, Google tries to verify the most active contributors as real experts in the field.

5.4.8.3 Last.fm

This music website has a most interesting *Partner Up* program, which allows artists, without regard for their skill or their level of acceptance among the general public, register themselves and claim royalties for every time a song of them is played. If an artist doesn't claim any royalties, that is acceptable is well to Last.fm But besides a *Partner Up* program, this is also a part of larger business model.

5.4.8.4 News station and celebrity blogs

Many blogging sites have special links to the blogs of news station, like Wordpress has Partnered Up with CNN. The mutual advantage is that news stations provide high quality content, and in return their entries are displayed on a prominent place at the blog. Popular people, like soccer players or politicians may as well be the partner for a website. The mutual advantage is that this popular person draws traffic to a website, and in return his blog is displayed in the front page of the blogging website.

5.4.9 Related Patterns

5.4.9.1 *Recommendation Engine*

When the quality of information objects is certain because the information provider is known, these information objects can be recommended to information consumers. This is more applicable to *Impersonal Recommendations* than to *Personalized Recommendations*.

5.5 Rating Engine

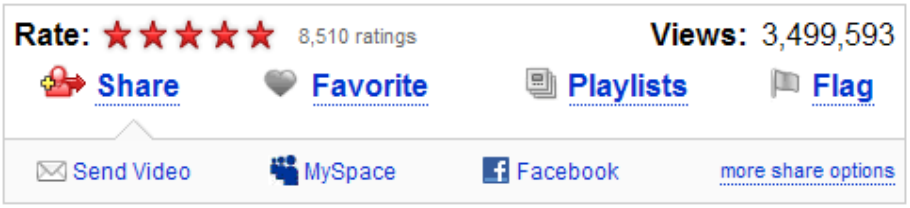

Figure 16: Example *Rating Engine*

5.5.1 Alternative names

Rating, Grades

5.5.2 Intent

Collecting subjective opinions as meta-data about information objects. This is preferably done in a standardized way, in order to make aggregation and refactoring of these data possible.

5.5.3 Context

In a Media Provision or Metadata Generation context: When a lot of users contribute, the information quality of the submitted information objects may vary. It is hard to automatically asses the information quality of an information object, especially when more subjective criteria (e.g. Understandability, Conformability and Fun) are involved. However, users who view the information object, can fairly well assess the overall quality of the information object.

5.5.4 Addressed Criteria

Depending on the implementation of the pattern, but usually the more subjective criteria are targeted. Especially information objects with high scores on Fun and other Wow-factors can easily attract high ratings. On the downside, information objects with poor scores on the hygiene factors will attract low ratings.

♠ ♣ ♥ ♦

5.5.5 Problem

When everybody is free to contribute, and all contributions become independent information objects, it is hard to distinguish which information objects are good and which are poor. This knowledge could be valuable, especially when combined with a *Recommendation Engine*.

5.5.6 Solution

Let information consumers decide for themselves, by giving them the ability to rate the information object, or even the information provider. Users are often more than willing to let their opinion be heard, and are even willing to invest a little time in that. Therefore, make the rating process pretty easy, but resistant to vandalism and abuse.

5.5.7 Rationale

This pattern improves information quality by creating better meta-information. *Rating Engine*s work because users are willing to provide positive feedback to information object they enjoy consuming and negative feedback to information objects that they did not enjoy. This gives the impression of influence (or power) over the information object in the form that they can influence future consumption of the information object by the general public. This implies an implementation of the *Recommendation Engine*.
Another way in which users are using ratings is to ensure that they will not consume the same information object twice, or that they will consume the information object more often. This rationale is applicable in environments with information objects that are intended for multiple consumptions by the same information consumer (e.g. music).

5.5.8 Known uses

5.5.8.1 Digg.com

Digg.com is built around an *Express your Love* rating system. In Digg.com the *Rating Engine* is linked to a *Recommendation Engine*, which collects and aggregates the ratings and other the metadata, and combines this in either a categorized recommendation, or a personalized recommendation. The aggregation for the *Impersonal Recommendations* is fairly simple; a sum of the positive diggs (sic) and the negative buries, the higher the total, the higher it appears on the list. To prevent clogging of the recommendations, the site has implemented a *Remember to Forget*, and to ensure new feed in the *Recommendation Engine* an *Upcoming Section*.

5.5.8.2 Icanhascheezburger.com

This Fun-oriented group of websites is largely dependent on ratings. Since it is (at the moment) impossible to automatically judge if the combination of a photo and caption is funny, the ratings of the user are the only source to make a good judgment. As Digg.com, they have implemented a simple rating system, with thumbs up and thumbs down. But here, the total count is not aggregated when displayed, but both the thumbs up and thumbs down are displayed. The sites display only the highly appreciated photos which are distilled by the *Upcoming Section*.

5.5.8.3 Youtube.com

Youtube.com uses a 5-star rating system. The advantage of a 5-star rating system is that it allows for a diversification of the ratings given to an information object. The usage of stars instead of grades removes the bias on grades as imposed by the school system. The diversification enables more sensitivity in the grade giving, which would result in probably more accurate grades. However, due to the highly subjective nature of ratings, it is impossible to measure this. In Youtube.com, the *Rating Engine* is coupled to an advanced *Recommendation Engine*, which allows for personal and *Impersonal Recommendations*.

5.5.8.4 Boardgamegeek.com

This website uses a personalized rating system, linked to a rich and personalized *Recommendation Engine*. This site has a two-way rating system, one giving a rating

to the overall fun of a game, and another which is named 'weight'. Board games which take a lot of time and strategic planning are considered heavy, but fast and luck based games are considered light. This implementation is acknowledging that board games have different qualities, which cannot all be captured in a single metric. Even the scales (resp. 1-10 and 1-6) are different for the different metrics. This is an implementation of the *Splitter* in the field of the *Rating Engine*. There is a simpler *Express your Love* rating system in place for the evaluation of user contributions to the forum.

5.5.8.5 Zoover.com

The holiday review site Zoover is perhaps even more interesting when it comes to diversified rating systems. When rating a holiday accommodation, there are several options to diversify the rating given. The options include 'General impression', 'Service', 'Location', 'Price/Quality', 'Rooms', 'Food' and 'Child Friendly'. Unique to this rating system is that it is possible to leave one or several of the fields blank. It is unclear what is done with partial ratings, since there is an aggregate grade displayed for accommodations in the main page, but equal weights to unequal factors seem illogical. For the destinations, there is another rating system, again with the optional elements, which is different for summer and winter destinations. Differences include après-ski or nightlife and culture or novice ski area. This difference enables different ratings for summer and winter destinations.

5.5.8.6 Last.fm

Last.fm uses an *Express your Love* rating system. Songs which are appreciated by the users can be rated by a single click. Songs which are not appreciated can be blocked from a playlist by a button labeled 'Ban this track'; which removes the track from the random playlist of the user.

5.5.8.7 Wikihow

Wikihow has another interesting implementation of this pattern. At the bottom of each page there is a question asking: "Was this article Accurate?" accompanied by two buttons: Yes and No. In this way, they are actually pushing the *Declaration of Failure* on the information quality criterion of Accuracy. At the same time, it is an implementation of the *Rating Engine*-pattern.

5.5.9 Implementations and refinements

5.5.9.1 *Express your Love*

The *Express your Love* implementation of this pattern is a specific implementation without diversification. It allows users to either express their love for an information object, or to loathe an information object, as that the user really don't like the (information quality of the) information object. There is nothing between those choices, as is the case with more complex implementations of the *Rating Engine*. This should not be confused with abuse report buttons, which are an implementation of the tagging of information objects.

5.5.9.2 *Rating without Acting*

This implicit implementation assumes that user behaviour gives an indication about the quality of information object, even though these users do not give feedback in the form of ratings. In the case of video or music, (completely) consumed information object may indicate that they information object is perceived as enjoyable by the information consumer. Thus the number of views gives an indication of the quality of the information object, as is prominently displayed on Youtube.com. This implementation of the pattern looks like Till Schuemmer's pattern *Activity Counter (Schuemmer, 2005)*.

5.5.9.3 *Love is Public, Hate is Private*

To prevent repercussions among users who rate each other's information objects unfavourably it is wise to make unfavourable ratings private. The reason to keep favourable rating public is to have a reference which enhances the Reputation of a contributor.

5.5.9.4 *Social Rating*

A specific, new implementation of the *Rating Engine* is *Social Rating*. It is visible when Media Provision websites have implemented an icon which links to social Metadata Generation websites, such as Digg.com, del.icio.us, reddit.com or stumbleupon.com. This is actually an externalization of the Rating pattern, and it serves as a place for attracting new information consumers as well.

5.5.10 Related Patterns

5.5.10.1 *Recommendation Engine*

A *Rating Engine* is nearly always accompanied by a *Recommendation Engine*. This is the way communicate the found quality characteristics of information objects to the information consumers.

5.6 Recommendation Engine

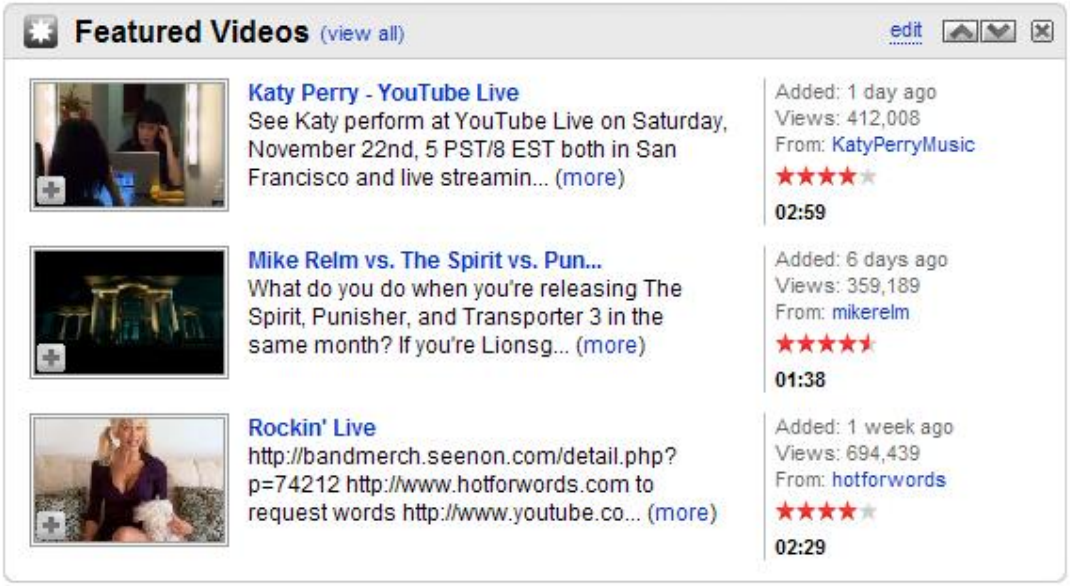

Figure 17: Example *Recommendation Engine*

5.6.1 Alternative names

Hall of Fame (Schuemmer, 2005)

5.6.2 Intent

The intent of this pattern is creating Recommendations which lead a user to information objects which they might enjoy.

5.6.3 Context

This pattern is usable in a Media Provision setting, usually with an implementation of a *Rating Engine* or *Tag Engine*. Information consumers are diverse, but may have some characteristics that they share. And these characteristics often result in similar preferences. Users tend to favour complete articles over stubs, high quality videos over low quality, and good-recorded songs over poorly recorded songs. However, a website should get these information objects to the attention of the information consumers.

5.6.4 Addressed Criteria

Relevancy, Ease of Navigation, (Secondary, all criteria except Process Pragmatics.)

♠ ♣ ♥ ♦

5.6.5 Problem

When a user is drawn to a website, the website wants to give a favourable impression and keep making a good impression every time he returns. Besides a good lay out and good interaction design, the website wants to offer the user the information objects which the user is likely to enjoy consuming.

5.6.6 Solution

The solution is twofold: it is consists *Impersonal Recommendations*, which are supposed to apply to all users; and *Personalized Recommendations*, which are specific to a consumer. *Impersonal Recommendations* are information objects of which the quality is high. The website can come to know which information objects are of high quality, because of a good implementation of the *Rating Engine* or through more objective measures, such as views of an information objects, or the resolution of a video or picture. *Personalized Recommendations* are harder to implement, because an implementation of the *Trusted Contributors*-pattern is required. Then the website can link consumed information objects to a specific user and can recommend that user new information objects based on the behaviour of other, similar users. There is also a third way; which is recommendations associated on the information object. This is a good way to implement this pattern when most of the users are unregistered.

5.6.7 Rationale

This pattern improves information quality by retrieving higher quality information objects. The most basic logic behind a *Recommendation Engine* is that the preferences and the associated behaviour of users are not random. There are lines and patterns in that behaviour among users, users can by classified and other users with similar tastes may prove a source of information for those users. The other part is that users like to be offered good information objects, and they are likely to consume the information objects offered to them.

5.6.8 Known uses

5.6.8.1 Digg.com

Digg.com is the most prominent user of the *Recommendation Engine*; as its business model is built almost solely around the recommendation of information objects to the visitors of the site. The first is the implementation of the *Recommendation Engine* with regard to new or unregistered users. They have *Impersonal Recommendations*, ordered in a list with the latest addition to the general recommendation list on top. With each recommendation there is a number of 'diggs' which is an indication of the relative popularity of the information object recommended. There is also another way of ranking the information objects, which is slightly less prominently displayed, but still on the front page. There is a collection of recommendation ordered by relative popularity. To refresh the recommendations offered, Digg.com forgets recommendations after a period of time. This is an implementation of the *Remember to Forget* pattern.

5.6.8.2 Youtube.com

This video website uses all three ways of the *Recommendation Engine*; but in different contexts. First, Youtube.com tries to recommend general high quality videos to all users. When a user visits the site, it gets recommendations for featured videos, which are selected by the editors of the website. By these *Impersonal Recommendations* users can get a favourable (first) impression of the videos on Youtube.com. Second, this site uses a Personalized *Recommendation Engine* for when a registered user visits the site. Based on previously watched videos, Youtube.com recommends other videos which the user will like. Third, Youtube.com has an

implementation of the *Recommendation Engine* based on the information object consumed. While watching a video, on the right hand side is a field with related videos. These recommendations are also displayed after a video has finished.

5.6.8.3 Wikipedia.

Wikipedia uses this pattern in the 'Featured Article'-section, which is a high quality article, which is not relevant for the information consumer, but a high quality information object in general. The Featured Article-status also has an additional benefit; it motivates users to enhance an information object, so it can be displayed in this section. The disadvantage of putting an article in the display as a general recommendation is that it attracts vandalism.

5.6.8.4 Last.fm

Last.fm lets users listen to their own radio station, of which each new song is actually a personal recommendation for a song. In that way, they first try to generate songs that a user might like, by associating the artists with each other. To achieve this, they link artists with similar tags to the artist who station is listened. Another way of associating is looking for other users with more or less the same preferences in music and offering the other music that these users have on their computers as a recommendation. Additionally; when an unregistered visitor comes to the site; there are hot artists on display; based on relative increase in plays. If possible, these artists are associated with the country the user visits the site from; this feature acknowledges the high sensitivity of music to cultural and temporal differences. Also there is an entire section named 'Charts' in which there are *Impersonal Recommendations* for all users. There is an additional diversification possible based on genre if the user has a certain preference.

5.6.9 Implementations and refinements

5.6.9.1 *Impersonal Recommendations*

This is the simplest implementation of the *Recommendation Engine*, which has as the main advantage that it does not require any personal information from *Trusted Contributors* or *Registered Members*. Regardless of the users, the General Recommendation will offer a general high quality information object, although it is impossible to offer high Relevancy. Despite being irrelevant, users will see that the information object recommended is of high quality. Therefore this specific implementation will enhance perceived quality of the website for new visitors and unregistered users. Wikipedia uses this specific implementation in their 'Featured Article'-section.
A major positive side-effect of this pattern is that *Trusted Contributors* will invest more effort in contributing high quality information objects, since a mention in the *Impersonal Recommendations* is motivating, much in the way that Till Schuemmers *Hall of Fame (Schuemmer, 2005))* is motivating for users, but not on a user level, but on an information object level.

5.6.9.2 *Similar Users*

This is a specific implementation of the *Recommendation Engine*, which does not recommend information objects or external phenomena, but recommends other users

with similar interests, tastes or skills as other users. *Similar Users* has specific advantages if information objects are highly heterogeneous, such as with music. Digg.com has this specific implementation of this pattern, but uses it alongside *Impersonal Recommendations*. It has much of the characteristics of Till Schuemmer's pattern *Birds of a Feather (Schuemmer, 2005)*.

5.6.9.3 *Personalized Recommendations*

The counterpart of *Impersonal Recommendations*, this pattern acknowledges the fact that people are different in taste and style. It improves recommendations on the information quality criterion of Relevancy, although it might damage other information quality criteria.

5.6.10 Related Patterns

The *Recommendation Engine* does need some information to work. Of course they could gather information by plain analysis of characteristics of the information objects, in all implementations of this pattern observed; another pattern was linked to this pattern.

5.6.10.1 *Rating Engine*

This pattern is often accompanied by a *Rating Engine*, so that the ratings gathered about an information object are used to scale the information object higher or lower when recommending it to a known or unknown user. Only wiki-based sites tend to neglect this for their recommendations.

5.6.10.2 *Trusted Contributors*

Trusted Contributors is essential to create *Personalized Recommendations*. From all the sites visited, only Wikipedia does not link the *Trusted Contributors* pattern to the *Recommendation Engine*.

5.6.10.3 *Tag Engine*

The *Tag Engine* is important to associate information objects with each other which are not automatically searchable or classifiable.

5.6.10.4 *Partner Up*

When a *Partner Up* pattern is implemented, this is an opportunity for the improvement of the *Recommendation Engine*. When the provider of content is known and trusted, he can be relied upon for tagging and providing high quality information objects. These objects can be incorporated into the *Recommendation Engine* without further analysis.

5.7 Trusted Contributors

Figure 18: Example *Trusted Contributors*

5.7.1 Alternative names

Sign up, Registering Members

5.7.2 Intent

Registering Members, so the website knows who can be trusted and who should be watched carefully. In addition to this, the registering of members enables a personalized environment, with personalized settings and recommendations.

5.7.3 Context

In a Social Networking Context, to get the right information and the right links to the right users, the website has to verify the nodes (which are persons) in the network. In a Collaborative Content Creation, Media Provision or Metadata Generation environment, when a lot of users can add and edit information objects, it is unknown whether these users are trustworthy. If the website can make members register, they become more known, and it easier to track their activities. If they perform well, they can become *Trusted Contributors*.

5.7.4 Addressed Criteria

Because of the first intent, most information quality criteria which are prone to vandalism can be indirectly improved by this pattern. The second intent improves the Ease of Operation and the Relevancy of the consumed information objects.

♠ ♣ ♥ ♦

5.7.5 Problem

When a lot of users can add and edit information objects, it is unknown whether these users are trustworthy. They can be either good users, or malicious spambots. Usually the website has no more information than an IP-address and the information of that IP-address in its current session.

5.7.6 Solution

Let users register themselves for additional benefits and increased ease of use. It helps fighting vandalism, since their activities are linked to a username and an email

address. To ensure that only well-intended users register as members, there should be a barrier to bots registering. This can be achieved by implementing *Captchas* or *Asirras* in the registering process. Additional benefits for registered users can be an incentive to make anonymous users register. As users are longer registered, they can become more trustworthy and (automatically) get increased benefits.

5.7.7 Rationale

This pattern improves information quality by creating higher quality information objects. The rationale underlying this pattern has different aspects and insights to it. The first part is that users who have given away some of their identity; for example an email-address or a real name, will restrain themselves from vandalism and other malicious use of the site.
The rationale behind using *Captchas* or *Asirras* is that only malicious users will try to use bots for registering user accounts
Trusted Contributors will also provide valuable insight in the consumption patterns of information consumers, since *Registered Members* provide the opportunity to link sessions or individually consumed information objects together to a more overall consumption pattern.
Furthermore, users will not likely register, unless there are some additional benefits to registering. The rationale behind the custom interface or *Personalized Recommendations* is that it is the reason why users register in the first place, and at same time it enhances the experience of the user.
There is also an effect that when users have invested effort to sign up, they are more likely to return to the website.

5.7.8 Known uses

Nearly every website has the ability to register. Some interesting implementations of the *Trusted Contributors* pattern are mentioned here.

5.7.8.1 Facebook

Facebook, as Social Networking site, cannot live without an implementation of the *Trusted Contributors* pattern. Facebook ask for a name, an email address and a password. On top of that, they ask for gender and date of birth. Verification of that a user is in fact human is not done by a captcha, but by assuming that if someone has an email address, he is verified by someone else to be human. However, for everyone that is added as a friend, another captcha has to be filed. The only way a user can remove this step is by verifying himself through a phone number. In that case the user receives a text message in which he finds a verification number. After typing this verification number, he can invite friends without having to type *Captchas*.

5.7.8.2 Hyves.nl

Hyves.nl, as Social Networking site, cannot live without an implementation of the *Trusted Contributors* pattern. The site asks for a name, email address, username and password. The verification of that the applier is in fact human is done by a captcha.

5.7.8.3 Wikipedia

Wikipedia has the possibility to register, but provides the least additional benefits. Because of a philosophy that everybody should be able to contribute, and users are

well-intended until proven otherwise, also anonymous users can contribute to the website. However, *Registered Members* build up a reputation, which brings the trust needed to give additional rights. Once a user has over a hundred edits in the German Wikipedia, he gets the ability to flag revisions as trustworthy. When a user gets over 400 edits in the Dutch Wikipedia he gets the ability to change titles as additional bonus. For greater responsibilities, there are administrators (which are elected) and system operators.

5.7.8.4 Google Knol

Google goes to lengths to verify the trustworthiness of an author, especially when it comes to high profile articles about difficult subjects. If that is the case, they try to verify that the author in the case is indeed the person who he claims to be. Thus they collect references contact users to verify the claim.

5.7.9 Implementations and refinements

5.7.9.1 *Captchas*

Captchas (Completely Automated Public Turing test to tell Computers and Humans Apart) are deformed words, letters or numbers that cannot be read automatically by text recognition, but are readable to humans. By typing in the correct letters or numbers, access is granted and the user can continue the process of signing up.

5.7.9.2 *Asirras*

Asirra (Animal Species Image Recognition for Restricting Access) is a new variant of *Captchas*; which is developed by Microsoft Research. Instead of recognizing words or letters, it is based on the human ability to tell cats from dogs. This is harder to hack by means of computers and is easier for humans to conduct. It is still beta, but it is a promising technology.

5.7.9.3 *Verification emails or phone calls*

Numerous websites use verification emails to an address to make sure that the address given is indeed an existing address for user. Other websites even go as far as making phone calls to register trusted users.

5.7.9.4 *Registered Members*

Not every user who registers is a *Trusted Contributor*. As long as users do not contribute or try to influence other users' opinions, it is not necessary to verify that users are indeed well intended and human.

5.7.10 Related Patterns

5.7.10.1 *Recommendation Engine*

This pattern enables the *Personalized Recommendations* which are a specific implementation of the *Recommendation Engine*.

5.7.10.2 *Partner Up*

Partnering Up is also a way of contacting a *Trusted Contributor*. The main difference is that a *Trusted Contributor* can be a nobody outside the information repository, while a Partner is trusted and well known to the outside world.

5.7.10.3 *Tag Engine*

The *Tag Engine* can be fed by input from *Trusted Contributors.* A way to improve to achieve this is by letting users create a tag cloud of tags that they often use, so they only have to click to add a tag.

5.8 Search Engine

Figure 19: Example *Search Engine*

5.8.1 Alternative names

Search tool

5.8.2 Intent

This pattern is aimed at creating an easy and rather intuitive way to find information objects in the information repository.

5.8.3 Context

As the information repository grows, the number of information objects increases and users are getting a harder time to find the information object. However, if the website has something that could aid in the search of the information consumer, he would probably use it. The *Search Engine* pattern is found at nearly all major sites.

5.8.4 Addressed Criteria

Relevancy, Ease of Operation, Ease of Navigation, Accessibility

♠ ♣ ♥ ♦

5.8.5 Problem

As the information repository grows, the number of information objects increases and users are getting a harder time to find the information object. This is problematic, as websites often want their information repository to grow, but they have to keep in mind that the information objects still need to be findable.

5.8.6 Solution

Implement a *Search Engine*. There are numerous ways to do this, and *Search Engine*s have been quite extensively developed. Problems might occur when information objects that can not be automatically indexed (typical anything not text-based.) In that case the information repository has to add tags and headers to power the *Search Engine*.

5.8.7 Rationale

This pattern improves information quality by retrieving higher quality information objects. The *Search Engine* helps information quality through helping users find what they are looking for. This greatly increases Relevancy for the users when they have used a search phrase in a *Search Engine.* Also, the current implementations are easy to use, and improve navigation and the accessibility of the information objects.

5.8.8 Known uses

5.8.8.1 Youtube.com

Youtube.com has implemented a *Search Engine* to search through their site. They have as soon as a user starts typing also suggestions for search terms. That means that if a user is only halfway done with typing a popular search term, Youtube.com will suggest that search term in a drop down menu under the search bar. They do provide the freedom to continue typing in case the user is searching for something more specific.

5.8.8.2 Last.fm

Last.fm has a *Search Engine* which is more intended at categorical searches. A user can specify one of the 6 categories on which he wants to search: Music, Event, Tag, User, Group or Label. This narrows down the search to a limited number of instances.

5.8.8.3 Wikipedia

Wikipedia has a main site on which there is an implementation of the *Search Engine*, which can be categorized by the language in which the search should be conducted. In the individual language dependent sites, there is an implementation which redirects to the article with the matching title, or otherwise to articles in which the searched term is present. The special thing happens when a searched term is not present, instead of redirecting to more or less matching terms, there is the ability to create that information object.

5.8.9 Implementations and refinements

5.8.9.1 *Narrowed Down Search*

When users have several categories from which they can distinguish before they enter a search term, this is considered a *Narrowed Down Search*. It is possible to do a *Narrowed Down Search* by letting the user choose a category in a drop down menu. An alternative is to let user add tags if they search for a specific item. For example: adding the tag [video], will only return video results. An advantage is that adding tags allows for more different categories than would be practical in a drop down menu, and it enables searches in multiple categories. Another advantage is that different division can be made. A disadvantage is that using tags performs less at Ease of Operation than a dropdown menu.

5.8.9.2 Find the Guru (Schuemmer, 2005)

The *Find the Guru* pattern as proposed by Schuemmer (2005) is in this context an implementation of the Search Engine, with a specific focus on the Trusted Contributors who are experts on their field. It has relations to the Recommendation Engine and the Trusted Contributors.

5.8.10 Related Patterns

5.8.10.1 *Tag Engine*

The *Tag Engine* creates the tags needed to be able to use a *Search Engine* in any non-text based environment.

5.8.10.2 *Recommendation Engine*

In a way, a *Search Engine* is an implementation of the *Recommendation Engine*, but associated on a certain keyword. Like recommendations, the results of a search can be selected and displayed according to Timeliness, popularity or any other quantifiable measure.

5.9 Tag Engine

Figure 20: Example *Tag Engine*

5.9.1 Alternative names

Tag Cloud, Categorization, Labels

5.9.2 Intent

Creating labels and tags for entries submitted. Thus being able to group, index and link them.

5.9.3 Context

In a Media Provision or Metadata Generation context, when users upload or create content, it is hard to automatically identify what they have uploaded. Therefore, a website can enable users to tag the objects they create.

5.9.4 Addressed Criteria

Ease of Navigation, Relevancy

♠ ♣ ♥ ♦

5.9.5 Problem

When users upload or create content, it is pretty hard to identify what exactly they have uploaded. This is especially though when the information objects are videos, pictures or anything else that is not automatically indexed.

5.9.6 Solution

Contributors have the knowledge about the information objects that they have uploaded, since they thought it was worth the trouble of uploading or submitting it. Therefore, let them add tags, which can be used in the *Search Engine*.

5.9.7 Rationale

This pattern improves information quality by retrieving higher quality information objects. Tags are assumed to be logically related to the object they are derived from. If there is a picture with a cat, the majority will recognize it as a cat. So, human effort can create a tag which relates logically to the picture represented.

5.9.8 Known uses

5.9.8.1 Del.icio.us

This website uses tagging as it main functionality. What is more, it has the functionality of a collaborative *Tag Engine*, in which there is not only a Metadata Generating part (in this case, tagging of websites); but also a Social Networking side to the *Tag Engine*. That is, other users can view the tags a user has assigned to a website, and use these tags for their own benefit.

5.9.8.2 Youtube.com

Youtube.com has the functionality to add tags when a user uploads a video. These tags are used to associate the video to other videos and as feed for the *Search Engine*. There is no room for adding tags when the video is consumed.

5.9.8.3 Last.fm

Last.fm uses this pattern in two different ways; a user can either tag artists that they are similar to other artists, or one can tag that they belong to a certain genre. Of course, a user can do both. This is used as information for the *Recommendation Engine*.

5.9.8.4 Google Image Labeller

This is a *Tag Engine* with a special twist: It encourages users to participate in the tagging process, using a game form. In that way, users are encouraged to label the pictures, in an accurate way, without the need to pay the users. The tags created in this way are in turn used by the Google image search. It is also one of the few Google services for which registering is not required.

5.9.9 Related Patterns

5.9.9.1 *Search Engine*

Typically in Media Provision environments, which are not quite as easy to search through as text based collaborative content creation environments, the *Search Engine* uses the tags (and perhaps the title) to search through the information objects.

5.9.9.2 *Recommendation Engine*

In Media Provision environments, recommendations can be made by association on the information object. To do so, these items should be tagged, otherwise it might prove pretty hard to place a recommendation for a new information object.

5.10 Upcoming Section

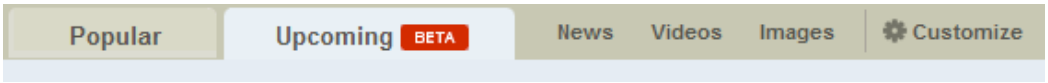

Figure 21: Example *Upcoming Section*

5.10.1 Alternative names

New entries

5.10.2 Intent

Ensuring new entries in the *Recommendation Engine*, by adding a section in which new ideas have the time to grow, to become popular and to make it to the 'big' *Recommendation Engine*. At the same time it is a filter which throws out the poor new entries.

5.10.3 Context

In a Media Provision or Metadata Generation context, when the website has implemented a *Recommendation Engine*; but the recommendations stay the same over time. This while an intent of the website is to provide new recommendations to the users over time.

5.10.4 Addressed Criteria

Timeliness, Relevancy

♠ ♣ ♥ ♦

5.10.5 Problem

When users rate an information object favourably, it should appear high in the *Recommendation Engine*. However, once an information object is consumed, it is unlikely that the user will appreciate another recommendation for that information object. Once the user has exhausted the recommendations, he will leave and not return to visit. On the other hand, new entries have a hard time getting popular, since the *Recommendation Engine* will not show the newly added information objects.

5.10.6 Solution

Add an *Upcoming Section*, in which new entries have the time to grow and become popular. In the mean time, poor entries will not get that popular, so they will not come to the 'big' *Recommendation Engine*. They have to be visible to the users though, so that users who are willing to rate and assess the information objects can do so.

5.10.7 Rationale

This pattern improves information quality by retrieving higher quality information objects. The reason why an *Upcoming Section* sorts effect is because light users will hardly visit an *Upcoming Section*, as they are saturated by the *Impersonal Recommendations*. Heavy users, which are more experienced, will visit the *Upcoming*

Section, where they will be confronted with relatively poorer information objects. This leads them to rate the objects they consider high quality to promote them to the 'big' *Recommendation Engine.*

5.10.8 Known uses

5.10.8.1 Digg

Digg.com uses an *Upcoming Section* to filter the poor new entries from the good ones. At the same time they have a pretty hard and elaborate process a user has to go through before an entry can even be submitted. This ensures that at least one user has gone through the effort of submitting a new information object, so he must rate it favourably. The *Upcoming Section* at Digg is split among the categories that are available in the 'big' *Recommendation Engine.*

5.10.8.2 Icanhascheezburger.com

The Icanhascheezburger.com group of websites has implemented a two way approach of the *Upcoming Section.* First there is an *Upcoming Section* in which new information objects are displayed according to their relative age. But, creating and submitting is made as easy as possible, an approach contradicting the approach of Digg.com. The second approach is the Vote section, in which the *Upcoming* or new information objects are displayed, with the explicit request for a rating. Then users can rate the pictures with captions by clicking 'win'; fail; or skip. The actual average rating is displayed when the user sees the next information object. In this way, the site collects numerous ratings, which help them to display the best on the front page. Because the next information object the user sees is selected at random from the available information objects, it is impossible to give on information object an artificial boost.

5.10.9 Related Patterns

5.10.9.1 *Recommendation Engine*

This pattern is only applicable with some kind of a *Recommendation Engine.* It ensures the feeding of new recommendations to the engine.

5.10.9.2 *Remember to Forget*

While this pattern is aimed at the feeding of new information objects in the *Recommendation Engine*, it might become problematic when the number of visitors of a website is going down. In that case, the new information objects in the *Upcoming Section* might be unable to provide new input. But, if the old ratings and information objects are forgotten over time, this problem can be solved.

5.11 Version Control

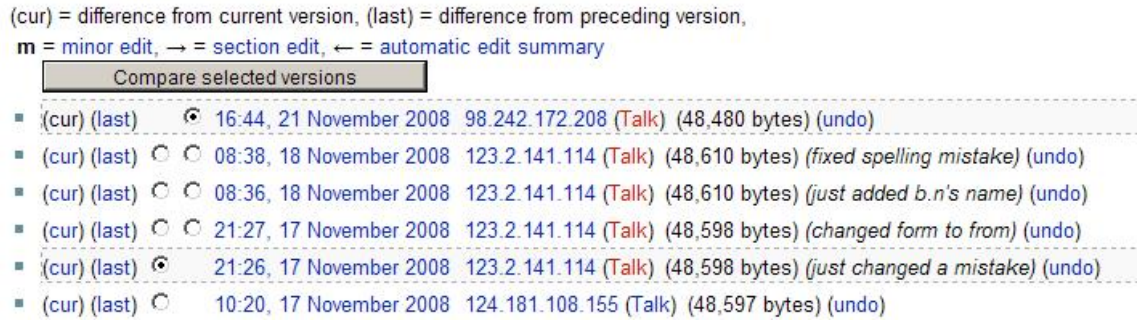

Figure 22: Example *Version Control*

5.11.1 Alternative names

Selective Rollback

5.11.2 Intent

Controlling versions of an information object; as to enable rolling back to previous version and controlled versions on the way.

5.11.3 Context

In a Collaborative Content Creation environment, when an information object changes often, a website wants to know which versions are good, which are bad, what is the best version, what is the most recent version, what is pure vandalism and what the differences between the versions are. Also a website needs a means to counter vandalism. A Collaborative Content Creation website can do this by implementing a *Version Control* system in the information repository.

5.11.4 Addressed Criteria

Volatility, Verifiability

♠ ♣ ♥ ♦

5.11.5 Problem

When an information object changes often, a website wants to know which versions are good, which are bad, what is the best version, what is the most recent version, what is pure vandalism and what the differences between the versions are. Moreover, when the information quality of an information object is damaged, the website wants to be able to undo it.

5.11.6 Solution

Use a form of *Version Control*, which allows users to search for the best information object that is available. And at the same time use this *Version Control* to roll back any malicious edits made to an information object. This pattern can also be used to assess which articles need a revision, or to protect articles from apparent vandalism.

5.11.7 Rationale

This pattern improves information quality by creating (more accurate: restoring) higher quality information objects. The Rationale behind this pattern is that it is easier to roll back vandalism and make in that way a high quality encyclopaedia, instead of protecting and reviewing all edits before they get through.

5.11.8 Known uses

5.11.8.1 Wikipedia

Wikipedia uses *Version Control* to roll back to previous versions, and to identify vandals and malicious users. Furthermore, they use a kind of *Version Control* to identify which articles should be improved a lot before they are encyclopaedic relevant. For this, they use a time-based (typically 2 weeks) approach to give the editor the opportunity to improve the article.

5.11.8.2 Wikihow.com

Wikihow uses *Version Control* in much the same manner as Wikipedia, but are less subject to vandalism because of the lower traffic and the lower profile of wikihow.com

5.11.8.3 Ebay.com

Ebay.com is one of the few websites outside Collaborative Content Creation which have implemented '*Version Control*'. The way this website has implemented this pattern is by letting users view the revision history of the objects that are for sale. If the description of an item has changed, after the header 'Description' the link '(revised)' is added. Clicking this link will open a summary of revisions, but it is not extensive, nor is their a possibility to roll back to previous versions.

5.11.9 Related Patterns

5.11.9.1 *Trusted Contributors*

Trusted Contributors is important for a *Version Control* implementation, because trusted users do not require as much attention as unregistered users who only have an IP-address to verify the claim.

5.12 Remember to Forget

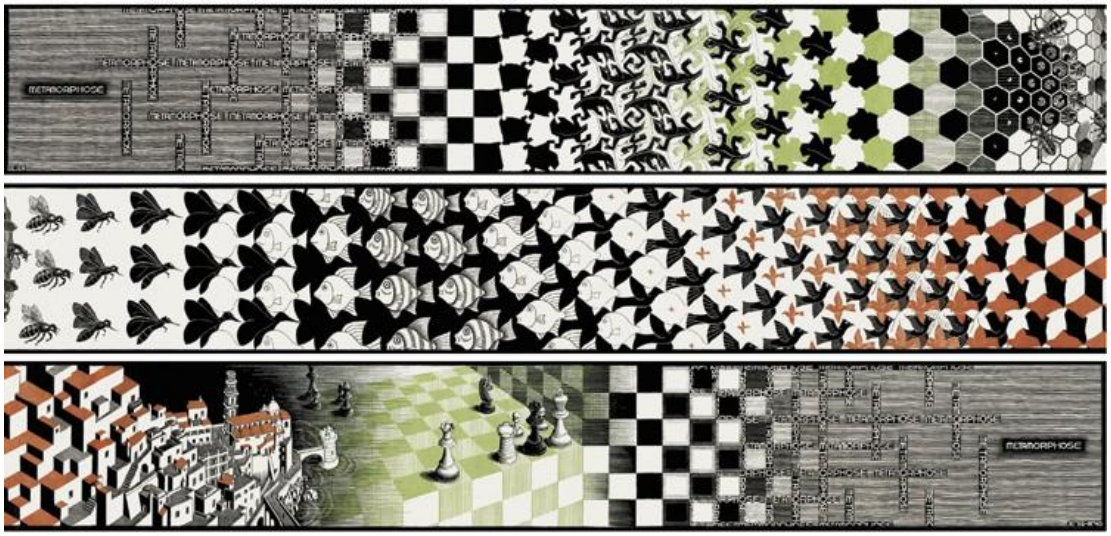

Figure 23: M.C. Escher Metamorphosis, (picture is unrelated)

This pattern was presented as a thumb sketch by Till Schuemmer in 2005 (Schuemmer, 2005)

5.12.1 Alternative names

Only Recent Data

5.12.2 Intent

The intent of this pattern is keeping the information in the repository timely and up-to-date. This improves information quality, as well as makes it attractive for information consumers to revisit the website.

5.12.3 Context

In any context, when a *Recommendation Engine* is implemented, the information might get outdated. This applies for friends, text, pictures and even historical data. When this historical data is used to provide recommendations or another purpose, this pattern assures that these recommendations are timely.

5.12.4 Addressed Criteria

Currency, Timeliness, Relevancy

♠ ♣ ♥ ♦

5.12.5 Problem

When recommendations are given to users, they will likely consume the recommended information objects. But, users are not likely willing to consume the same information object multiple times (with the scarce exception of songs, which are

often consumed multiple times). Therefore, the quality of recommendations will be falling if the same user visits the website more often.

5.12.6 Solution

Remember to Forget the information gathered and the recommendations provided. By forgetting old ratings, comments and tastes, this pattern incorporates the fact that tastes change. The culture and appreciation of information objects might change over the years. This is especially important in Metadata Generation contexts.

5.12.7 Rationale

This pattern improves information quality by retrieving higher quality information objects, because the information objects are more timely. Therefore the chance that users have previously consumed the information object decreases, which has a positive impact on experienced relevance.

5.12.8 Known uses

5.12.8.1 Digg.com

Digg.com thrives on this aspect, by forgetting the websites submitted within a day. Therefore websites can become popular within this day, but the next day, when the users return, the popular website of yesterday has been replaced by a new one. Digg has implemented *Remember to Forget* by the date a website is first posted on the site, not by the ratings given.

5.12.8.2 Icanhascheezburger.com

The *Upcoming Section* of the Icanhascheezburger group of websites has implemented a kind of *Remember to Forget.* When new pictures are submitted, these pictures are scanned through the *Upcoming Section* for a period of time, if after that period the picture still hasn't made it through, it is forgotten and removed from the *Upcoming Section*.

5.12.8.3 Last.fm

The Recommendation Engine of Last.fm has a specific implementation of the Remember to Forget Pattern. This is, that the Recommendation Engine displays artists and songs who are 'popular this week', with that forgetting the artists that were popular before.

5.12.9 Implementations and Refinements

5.12.9.1 Sort by newest

A way to partially omit this pattern is by sorting the recommendations by the newest first. In this way the Icanhascheezburger websites are organized, by incorporating all information objects which are of high quality, but the newest first, the same effect is sorted. A drawback may be that the offering of all information objects may clutter the repository, while the added value is small.

5.12.10 Related Patterns

5.12.10.1 *Recommendation Engine*

In Media Provision environments, recommendations can be made by association on the information object. To do so, these items should be tagged, otherwise it might prove pretty hard to place a recommendation for a new information object.

5.12.10.2 *Rating Engine*

It is possible to forget ratings given to an information object over time. It is quite a safe way in Collaborative Content Creation contexts, since in those contexts the information objects are dynamic, and therefore the rating given to a previous version of an information object may no longer apply to the information object.

5.12.10.3 *Upcoming Section*

The *Upcoming Section* is important, as *Remember to Forget* actually removes information from the repository, but the *Upcoming Section* provides new input for the *Recommendation Engine*.

6 Implementing patterns

As shown in Chapter 5, there are several patterns which can help websites with improving information quality. Since patterns represent abstract and high level solutions to general problems, they have to be adapted to the specific situation and problem at hand. The process to guide websites through such an adaptation is named Assessment and Improvement Method for Information Quality. (AIM 4 IQ). For this process, the input is needed for the strategic focus of the information repository, which should be an information quality criterion, as well as the type of website. First we present the general methodology, which we elaborate in (fictional) scenarios. This means the websites are real, but the behaviour as described is partly fictional. Not all websites are in the research sample, the first two websites are Web 2.0 sites, which were not included in the research sample, but were insightful in the broadness of possible applications of Web 2.0. The next two websites are included in the research sample, but the implementation of patterns is fictional. The last example is purely fictional, as a hypothetical example of a Web 2.0 site which displays highly hybrid behaviour. These scenarios are presented to give an impression of the possible implementations of the patterns.

6.1 The AIM 4 IQ process

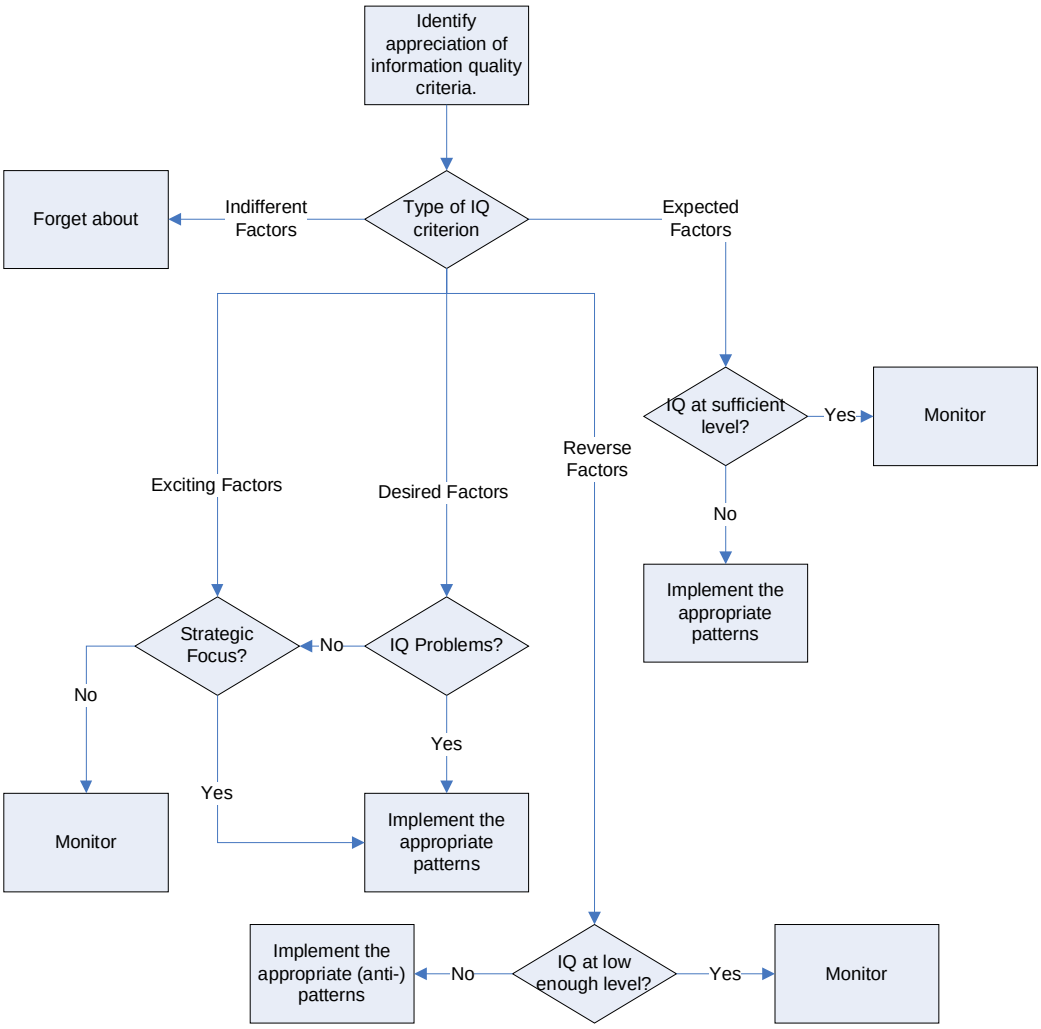

Figure 24: AIM4IQ Process

⇒ The rectangles represent processes.
⇒ The diamonds represent decisions.

The tool created to analyze the process websites can use to ensure information quality is a straightforward 4 step process.

1. Identify which quality criteria of the taxonomy apply to the type of information objects in the information repository. Identify which criteria are Expected, Desired, Exciting, Reverse and Indifferent factors.
2. Choose a strategic focus or strategic foci, for the information repository from among the Desired and/or Exciting information quality criteria.
3. Identify information quality problems with regard to information quality criteria.

4. Apply patterns improving information quality on these criteria and countering problems. Make sure that the applied patterns do indeed apply to the context in which the information repository is.

6.1.1 Determining which information quality criteria apply

When an information repository handles certain information objects, different information quality criteria apply. Articles (written words) do not require the same approach to information quality as videos, photos or songs. These information objects all have different applicable criteria. These are not only mentioned in the framework, but other criteria, derived from the criteria in the framework as well. For example, videos have little need for the accuracy criterion, since it is very hard to make a video of something that isn't there.

6.1.2 Choosing a Strategic focus

Choosing a strategic focus for your information repository is, strongly related to the type of website that the repository wants to be, and the target audience for the website. Any of the Desired of Exciting factors can be a strategic focus. If the information repository offers several types of information objects, the strategic foci of the information objects have to match, although they can be different. If for example the information repository focuses on high quality news photographs, with a concise description of what is happening in the picture, the information quality criteria for the photographs and the articles accompanying the photographs are different. The photographs have a need for completeness, the articles for conciseness. The strategic focus of the entire repository is supported by the different information quality criteria of the different information objects.

6.1.3 Identifying information quality problems

This is usually not an action, but an emergent problem, it is that information objects fail to meet expectations. Problems emerge when the level of quality is insufficient for the task at hand, or the information objects are outperformed by a competing website, which offers better quality information objects. In this respect, an information quality problem can also be interpreted as an opportunity to gain competitive advantage over the competition.

6.1.4 Applying patterns

Patterns are general solutions, which are widely applicable to counter problems. But the patterns should be tailored to meet the specific needs of a website. In the patterns presented in this thesis, there are specific implementations, as well as examples of occurrences of the pattern in practice.

6.2 Scenarios

6.2.1 Scenario: Skiline.cc

Skiline ski tracking is a web company which allows skiers to add, rate, track and time their ski routes. It is a typical Metadata Generation website, with a strong focus on a specific target group. With GPS-functionality, users can add ski routes and runways they have travelled and even add new routes. The added value for a user is that he or

she can find the distance travelled, highest speed achieved and longest run without stops. Users can also compare themselves to other users. However, users can only map to other users they know, and therefore find little added value. To increase added value, Skiline implements a *Recommendation Engine*. The type they implement is a *Similar Users*; which maps users of similar skills in types of runways, speed and frequency to the user of the service. By looking at the other routes these users have taken, Skiline recommends similar tracks in one of the ski regions which the user has chosen. The way they collect the enjoyability of routes and runways is by viewing the number of times it is done by a user, assuming that users will take runways more often if they find them enjoyable. This is an implementation of the *Rating Engine*, typically the type *Rating without Acting*.
After several years, it appears that the quality of recommendations is falling. Several of the recommended runways are closed, and users get poor recommendations if their skill has increased. To counter this effect, Skiline implements *Remember to Forget* on the ski routes, ski runways and the users. When recommending new routes or runways to users, only the last thirty routes taken are considered. This results in better recommendations. The more experienced users provide feedback for runways, commenting that it is illogical that skiing and snowboarding are treated in the same manner. Where skiers do have little problems with stretches of flat runways, snowboarders have a need for a continuous gentle slope. To improve recommendations Skiline splits the user groups and the recommendations using a partial *Splitter* pattern. Users can still get *Impersonal Recommendations* but *Registered Members* who have provided the information can get specific ski or snowboard recommendations.

6.2.2 Scenario: Yoyogames

Yoyogames has the innovative software of Gamemaker, which allows users to create, upload and share small games in the community. It is a typical Media Provision site, which offers software to create games and has a *Recommendation Engine* to offer games to their users. The software offered to create games is an implementation of the *Mark-up Tools*, which helps users to create games easier and quicker than conventional programming methods. To make sure that users are motivated to submit high quality games, the website has implemented a *Recommendation Engine*. Since popular games tend to be popular for quite a long time, users started to experience troubles with identifying new games. To counter this, the website introduced an *Upcoming Section*, in which the new games could become popular before they would be displayed and recommended to the general public. The *Upcoming Section* is called 'What's hot' and the other *Impersonal Recommendations* are called: 'Gamemaker Legends.' After some haggle and fierce discussions among the users of the website, it becomes apparent that blood and violence are not allowed on the website, since young users also visit the site. However, there is a significant group who enjoys blood and violence. Therefore they decide to split the information repository, into a child friendly website (the aforementioned yoyogames.com) and an unrelated underground website for games containing 18+ content. This is an implementation of the *Splitter*-pattern. After several years, it appears that the 18+ website is more popular than the original website. It draws away the high quality contributions, and the overall quality is falling. To counter this, Yoyogames implements a *Partner Up* program, where

professional developers and talented contributors can upload games, and in return get a share of the profits.

6.2.3 Scenario: Wikihow.com

Wikihow 'The How-to Manual That You Can Edit' [sic] is a typical collaborative content creation website in which users can add guidelines and informative articles. The focus of the website is to provide information objects which score better on Usability and Interpretability. This focus is mainly enforced by *Trusted Contributors* who adhere to the policies.
Users did start to contribute information objects, but the quality of was highly variable, some articles being complete manuals on how to revise an engine, whereas other guidelines are sketchy and incomplete. To counter this, Wikihow implemented a *Declaration of Failure*, a number of templates, which could be added to articles by *Trusted Contributors* to indicate different problems. Under the page 'Request for article improvements' a number of possible problems could be indicated. When an article was improved, the request could be removed. One of the comments was a lack of consistency among the information objects. Although *Trusted Contributors* could implement the policies to make these improvements, the better solution was to implement *Style Sheets*, which fixed the lay-out of classes of information objects. However, users started to ask for improvements on articles which would better fit an encyclopaedia, as they were focussed more on the underlying principles and the explanation why things worked as they did, instead of how-to. To counter this, Wikihow implemented a *Splitter*-pattern on an article level, with a focus on redirecting users who asked for such an improvement on an information object to Wikipedia, which is more focused on providing encyclopaedic articles.
As the number of visitors to the website increased, so did the number of malicious editors, which tried to push their specific products or services, and of course plain vandals, which did nothing but empty pages and put obscene language at places. These vandals were countered by an extensive *Version Control* system, which was operated by *Trusted Contributors*. Because of the swift rollback that these users exercise, the relative amount of time that an article remained in vandalized state was reduced. To keep users coming back to the website, wikihow.com has a section at the main page called: Spotlight Article. To assess which articles should be in the spotlight, only articles without any *Declaration of Failure* are considered to be suitable for a place in the Spotlight. However, as the number of articles and the respective quality grows, the number of articles without any *Declaration of Failure* grows as well. Although any of these articles would be considered good enough, it is desirable to display the best. In order to gather information about what articles are the best, Wikihow.com implements a *Rating Engine*, which allows users to rate articles on an A-F scale, in the same way as the American School system. This project is quickly abandoned as it becomes clear the ratings from outside the United States are more random distributed, creating only more ambiguity. This is caused by European contributors who do not understand that an A stands for 'very good' and an F for 'very poor.' For a more natural rating system, Wikihow.com implements stars instead of grades in a 5-star scale. This results in the additional clarity that was needed for a good assessment of the quality of articles.

6.2.4 Scenario: Hyves.nl

Hyves.nl, the foremost Social Networking website of the Netherlands, started small, as any social networking site would. The members of this site are all verified that they are *Trusted Contributors* by a *Captcha*, and a *Verification Email*. Without these verifications, it is impossible to send messages to other users. Without these verifications, it is possible to add other users as friends, but impossible to send private messages. Only public messages are allowed, although all users can receive private messages. When users sign up, they have a number of required fields, a limited number of suggested additional information and a limited number of free fields that can be filled in order to create a profile. Based on the required fields and the suggested additional information, comparable information objects are created.
When users have established a network, it is possible to recommend new connections that the users might want to establish. This is an implementation of the *Recommendation Engine*, more specific a personalized recommendation. When users are looking for someone specific, there is a *Search Engine*. The additional functionality of the *Search Engine* lets users search for users who have specific skills, hobbies, knowledge or tastes. This is an implementation of the *Find the Guru (Schuemmer, 2005)*. Nice additional functionality is that it allows users to view what are the links between you and the intended guru. This could help users to get an introduction to the person. After the emergence of mobile internet, the need for a version which would not be as resource heavy became apparent. Thus Hyves.nl introduced a light version: mobile.hyves.nl.

6.2.5 Scenario: Spellexchange.wiz

Spellexchange.wiz is hypothetical hybrid website which allows users to upload, share, create and improve magic spells. The hypothetic users are considered wizards, and the spells as described in the case are the hypothetic information objects.
Spells are not easily shared, a lot of factors should be considered. When users sign-up, the first thing that is verified is whether they are true wizards, not bots or zombies crafted by devious warlocks. To verify this, the website has implemented *Asirras* and *Verification emails* to make sure that the contributors are indeed genuine wizards. When users are new, they are not allowed to alter other user's spells until they have at least contributed one working spell that is verified by a senior *Trusted Contributor*. If a user reviews a spell (by attempting to cast it) he can provide reviews on that spell. To measure overall successfulness of a spell, a *Rating Engine* is introduced, which aggregates ratings of all spells. To counter other types of failure, like incomplete lists of ingredients, exploding heads and other minor inconveniences, a *Declaration of Failure* is implemented. Types of this *Declaration of Failure* are completeness (important for ingredients) accuracy, (important for difficult Latin spells), understandability (for what effect the spell should sort) and complexity (mainly a problem for beginners). To keep track of changes, a *Version Control* is implemented, in which the more easy spells and the recently *Registered Members* are closely monitored. There are also some words which block the automatic updating of the spell, as they are known to be in the standard incitation for e.g. minor demon summoning, the spontaneous vomiting of the caster or giving the room a hideous pink-purple wall decoration. When a spell is checked by a *Trusted Contributor*, it gets an additional green *Style Sheet*, which ascertains the user that the spell does indeed

sort the claimed effect. If the spell sorts the effect, but the effect is dangerous, a green-red style sheet is applied to the spell. Non-working spells get a yellow-black question mark lay-out, which is a *Declaration of Failure* at the same time. To provide more structural consistency, a standardized mark-up is enforced by making information about ingredients, tools and incitations available in their own fixed spot on every page. This is another implementation of the *Style Sheets*. As the information repository grows, the number of spells online becomes so large that there has to be a *Search Engine*. This *Search Engine* enables *Narrowed Down Search*, so that users can search on either problems, the effect a spell has, or on the components used.

7 Evaluation

In this chapter the three research subjects will be juxtaposed to come to a conclusion. This results in several graphical representations, compromised of classes of websites/patterns, websites/IQ-criteria and patterns/IQ-criteria.

7.1 Justification

Juxtaposing these aspects is important because after the observation of the patterns, there was a need for a clear overview of the usage of these patterns. Therefore, each of the patterns was assessed in the light of 23 different websites, which were diverse and which were the most used Web 2.0 sites in the Netherlands; complemented with some other sites which were not among the top 50 websites concerning traffic, but were insightful in the creation of patterns. Then these sites were assessed on their strategic focus concerning information quality. This is important because it helps on the one hand to discern errors in our research made in the classification of patterns and the impact on information quality criteria. On the other hand it helps to discern errors that the respective websites have made concerning the implementation of their methods compared to the strategic focus they have. To ascertain this, additional research is needed, measuring the state of the information quality in the different criteria. This is unfortunately beyond the scope of our research.

7.2 Results

There are several graphs and one matrix constructed as a result.

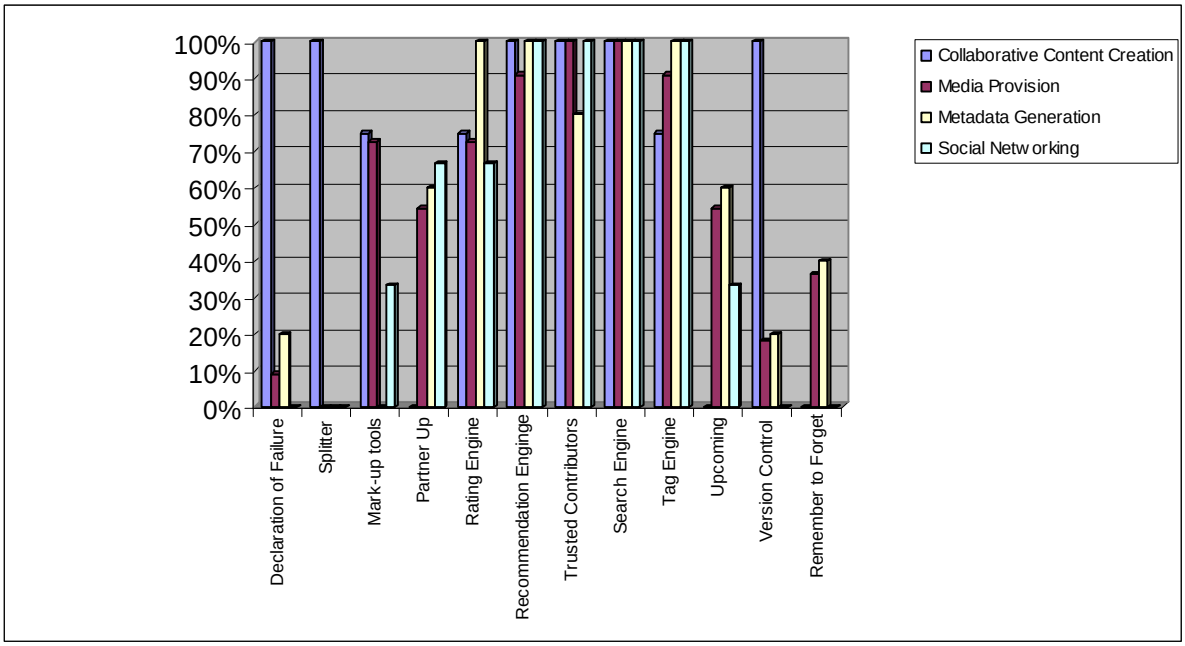

Figure 25: Usage of Patterns in Web 2.0 sites

In Figure 25: Usage of Patterns in Web 2.0 sites we see that some patterns are omnipresent, such as the Search Engine. Others, such as Splitter are only found in Collaborative Content Creation sites. Patterns may be found useful only in a specific

context, which can be found in the 'Context'-section of each pattern. The fact that each pattern occurs relatively often is caused by the high level abstraction of each pattern. Would they be more specific, the usage of a specific pattern would be smaller.

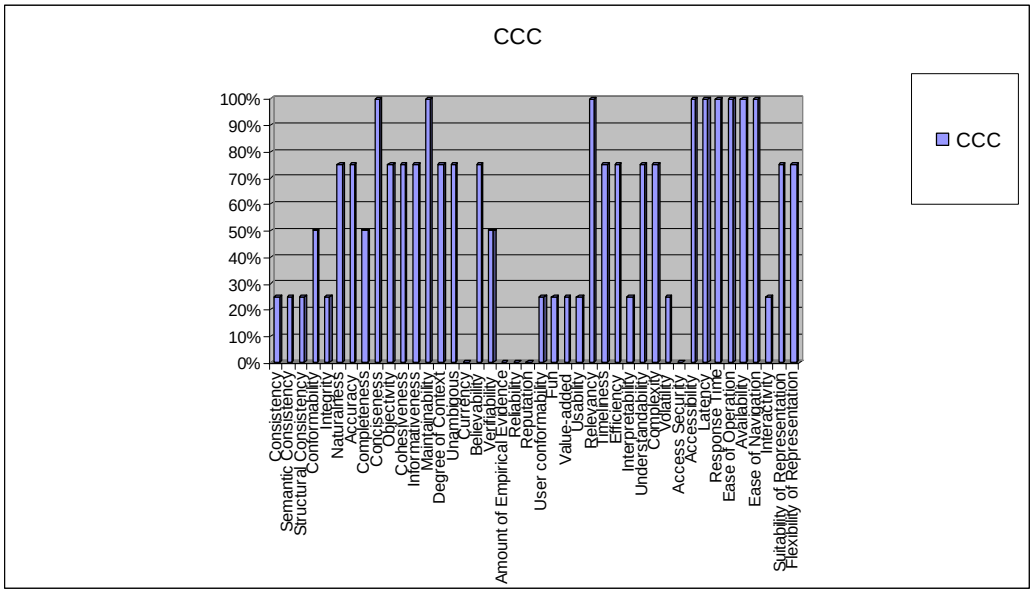

Figure 26: Occurrence of Information Quality Criteria in Collaborative Content Creation websites

In Collaborative Content Creation websites we see that the Process-Pragmatic Category is important. This is caused by the fact that these factors are all Expected Factors, which are essential for the user experience. Relevancy is important as well, much for the same reason. The fact that Conciseness is indeed a focus for Collaborative Content Creation is a bit puzzling. Since Collaborative Content Creation websites often have little trouble generating enormous amounts of information, it is important that this information is chopped up in smaller sizes, as to facilitate easy consumption of the information objects. The high occurrence of maintainability among this type of websites is expected, since Collaborative Content Creation websites have the technology to facilitate the maintenance of information objects on an ongoing basis. We also spot relative high Timeliness, as opposed to a low Currency. This is because information objects can be updated on an ongoing basis, but that is only necessary when the real world object the information object refers to changes. Therefore, Timeliness is far more important than Currency in a Collaborative Content Creation context. Also a low interest for Access Security is observed. This is caused by the wiki-philosophy of 'Assume Good Faith', which actually resents the usage of access restriction, since more people contributing will result in higher quality information objects.

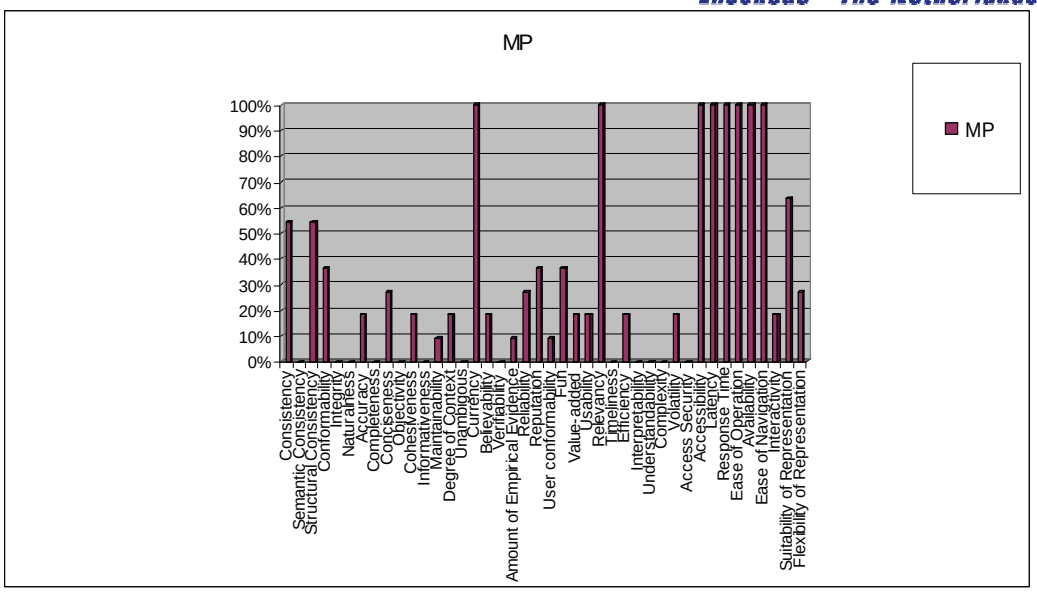

Figure 27: Occurrence of Information Quality Criteria in Media Provision websites

In Figure 27: Occurrence of Information Quality Criteria in Media Provision websites we can see that the Process-Pragmatic Criteria are of relative importance. In this case, that is because most of these information quality criteria are expected factors in a Media Provision context. Latency and Response Time have additional importance when the Media provided in a Media Provision context become more data-intensive, as is the case with (HD) movies, songs and large pictures. With respect to Currency and Timeliness we see the opposite of what we saw in Collaborative Content Creation websites. In Media Provision we see a focus on Currency, as a substitute of a focus on Timeliness. This is to be expected, since Currency is an objectively measurable criterion, which makes it easy to implement methods which rank information objects according to Currency. Since information objects cannot be updated on an ongoing basis, Currency is a good indicator for the expected Timeliness. The last focus of Media Provision websites is the Relevancy, which is present at all websites. This is displayed by the fact that all websites have implemented some kind of *Search Engine*, but in the case of hedonic websites, Relevancy becomes of smaller importance. We observed that Understandability was of no concern to all Media Provision websites. This was expected for photo and video sharing websites, but a discovery for blogging websites. It turns out that the ranking of information objects according to Understandability is not implemented, which may be caused by a lack of interest, or by a lack of technical solutions. In general we observe that the concern for information quality among all Media Provision websites is lower than among other categories.

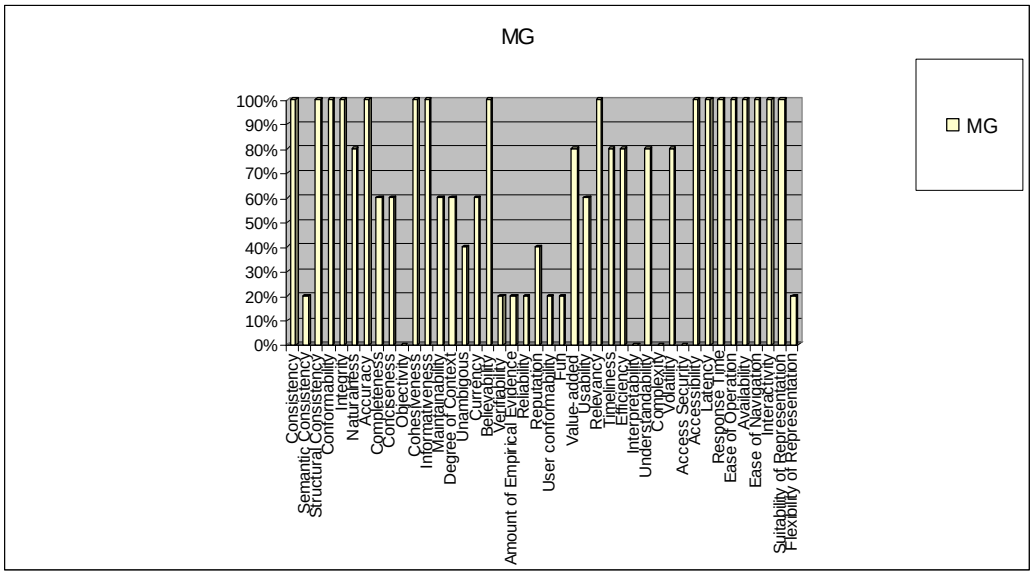

Figure 28: Occurrence of Information Quality Criteria in Metadata Generation websites

As we can see in Figure 28: Occurrence of Information Quality Criteria in Metadata Generation websites, Metadata Generation Websites have in general a high concern for information quality. We see expected behaviour in Metadata Generation, such as Consistency, which is needed to create comparable metadata. Accuracy, Cohesiveness and Informativeness are important to ensure that the metadata generated do indeed refer to a single information object. Integrity, which is defined as 'the extent to which the scope of the metadata is adequate' is of course a major concern to specific Metadata Generation websites, but at hybrid websites, were the generation of metadata is a nice add on, we observe that this might be lacking. We see little concern for Objectivity, which is caused by the fact that Metadata Generation websites gather opinions about real world objects. Objectivity may indirectly be achieved because extreme opinions balance out. We see again that Relevancy is a major concern to all websites, as are the Process-Pragmatic Criteria.

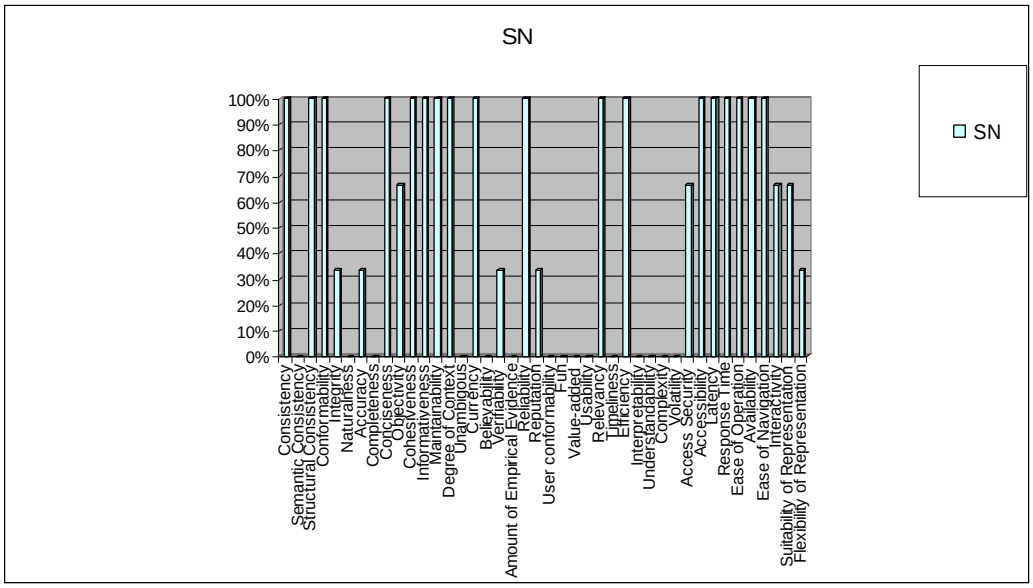

Figure 29: Occurrence of Information Quality Criteria in Social Networking websites

Due to the fact that in the research sample only three websites in social networking were analyzed, and that these websites displayed highly hybrid behaviour, it is hard to draw conclusion from this graph. We see that Consistency is of importance, but only as Structural Consistency. We see this in for example user profiles, which have a standardized lay-out, but the fields are free to fill in. We also observed a high concern for Relevancy and the Process-Pragmatics.

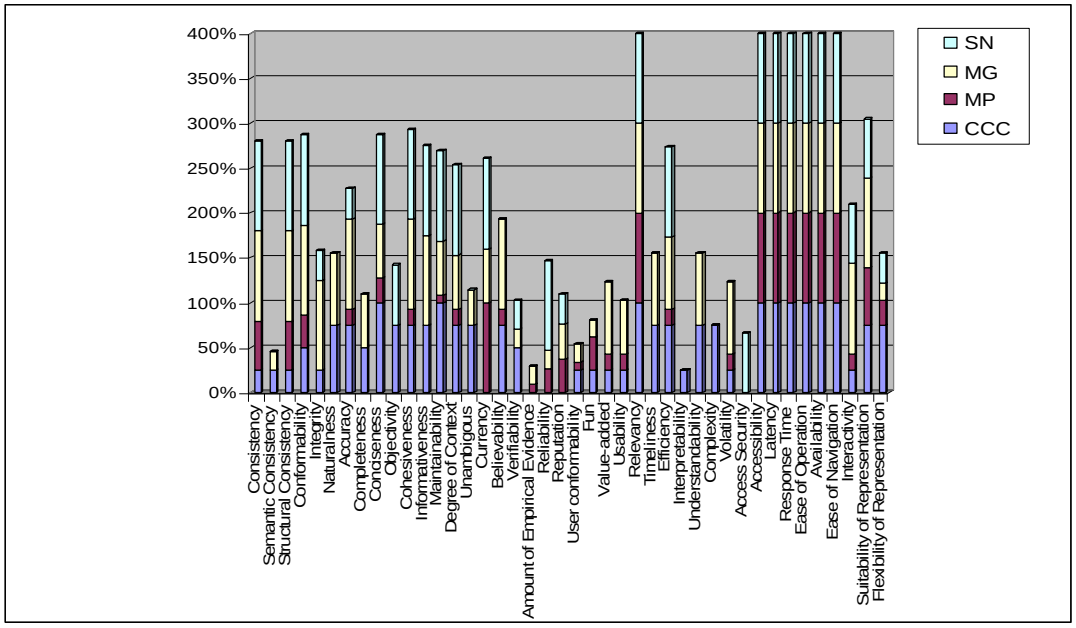

Figure 30: Cumulative Occurrence of Information Quality Criteria in websites

In General, we observe a high concern for the Process-Pragmatic Criteria and for Relevancy. We assume that this is because these criteria turn out to be Expected Factors in every context. But Relevancy might also be modelled as an Exciting factor,

making a highly relevant information object also very exciting. We also observe that in general the Semantic and Syntactic criteria are considered of more importance than the User-Pragmatic and Information-Pragmatic Criteria.

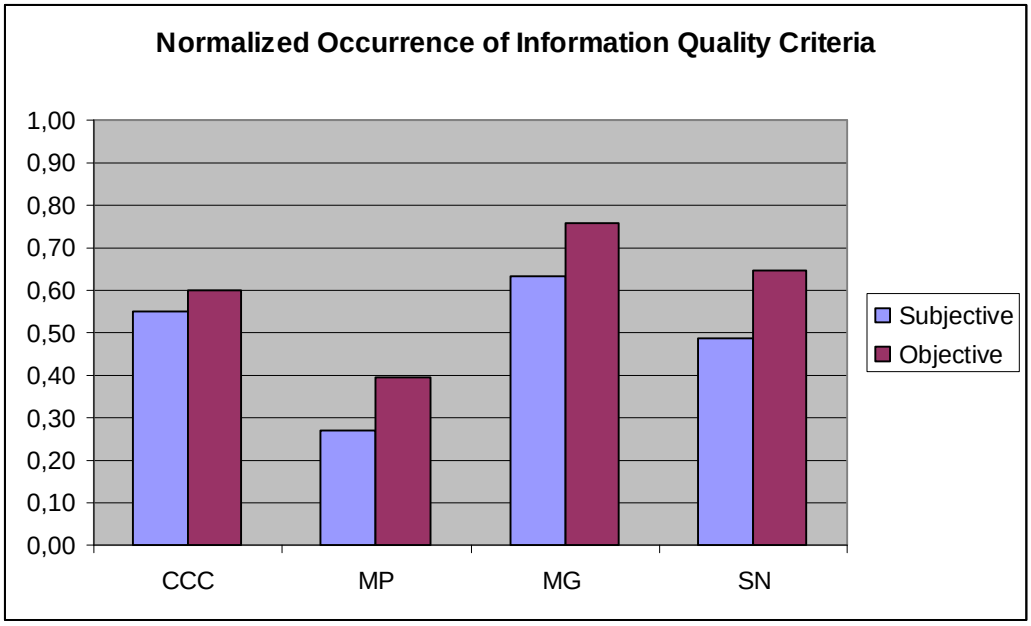

Figure 31: Normalized Occurrence of Information Quality Criteria

Besides the division in semiotic groups, there is another division, namely a division in subjective and objective criteria. A normalized graph of the different types of factors is displayed in Figure 31: Normalized Occurrence of Information Quality Criteria. We observed that the relative interest for Objective criteria was a bit larger at all different types of websites than the interest for Subjective criteria. This might be explained by the fact that Process-pragmatic criteria which are Expected factors are always objectively measurable (e.g. Response Time, Latency, Availability). Relevancy on the other hand is a subjective criterion of information quality.
When compared to other types of websites, Collaborative Content Creation has more interest for subjective information quality criteria in respect to the objective criteria. This is shown by the fact that the difference in interest between subjective and objective criteria is much smaller than at other websites (5% to 10-20%). This is likely caused by the fact that the improvement process at Collaborative Content Creation websites is executed by humans, who are more adept at assessing subjective criteria than machines are. Thus these websites have the ability to focus on these criteria and use that ability to the fullest.
As observed earlier, Media Provision websites have in general a lower interest for all information quality criteria than other types of websites. The lower interest might be explained by the fact that information objects can hardly be altered by the information repository, which is the opposite of Metadata Generation and Collaborative Content Creation.
Metadata Generation outperforms all other types of websites when it comes to the concerns for information quality. This is probably caused by the intense competition, that being that if another website decides to enter the Metadata Generation market for a certain kind of information object, it is highly likely that it will take a big market

share if it outperforms their competitors. Therefore Metadata Generation websites aim at Accuracy and Believability. It is also the only group with a focus on Integrity. Social Networking is performing quite moderate. Therefore it is hard to draw conclusions from these figures.

A short remark should be made about the weights of the information quality criteria. It is modelled here as that all information quality criteria have an equal weight. This is not necessarily true or even relevant. For example, if the perceived importance of Accuracy is much higher than of Enjoyability, then Figure 31: Normalized Occurrence of Information Quality Criteria, becomes more subjective.

Information Quality Criteria/Patterns	Declaration of Failure	Splitter	Mark-up tools	Partner Up	Rating Engine	Recommendation Engine	Trusted Contributors	Search Engine	Tag Engine	Upcoming	Version Control	Remember to Forget
Consistency	2	1	2	1	1	1	2	0	0	0	0	0
Semantic Consistency	2	1	1	1	1	1	2	0	0	0	0	0
Structural Consistency	2	1	2	1	1	1	2	0	0	0	0	0
Conformability	2	1	0	1	1	1	1	0	0	0	0	0
Integrity	2	1	0	1	1	1	1	0	0	0	0	0
Naturalness	2	1	0	1	1	1	1	0	0	0	0	0
Accuracy	2	1	0	1	1	1	1	0	0	0	0	0
Completeness	2	1	0	1	1	1	1	0	0	0	0	0
Conciseness	2	2	0	1	1	1	1	0	0	0	0	0
Objectivity	2	1	0	1	1	1	1	0	0	0	0	0
Cohesiveness	2	2	0	1	1	1	1	0	0	0	0	0
Informativeness	2	1	0	1	1	1	1	0	0	0	0	0
Maintainability	2	1	0	1	1	1	1	0	1	0	0	0
Degree of Context	2	1	0	1	1	1	1	0	0	0	0	0
Unambigous	2	1	0	1	1	1	1	0	0	0	0	0
Currency	2	1	0	1	1	1	1	0	0	2	0	2
Believability	2	1	0	1	1	1	1	0	0	0	0	0
Verifiability	2	1	0	1	1	1	1	0	1	0	2	0
Amount of Empirical Evidence	2	1	0	1	1	1	1	0	0	0	0	0
Reliability	2	1	0	2	1	1	1	0	0	0	0	0
Reputation	2	1	0	2	1	1	1	0	0	0	0	0
User conformability	2	1	0	1	1	1	1	0	0	0	0	0
Fun	2	1	0	1	2	2	1	0	0	1	0	0
Value-added	2	1	0	1	1	1	1	0	0	0	0	0
Usability	2	1	0	1	1	1	1	0	0	0	0	0
Relevancy	0	1	0	1	1	2	1	2	0	2	0	2
Timeliness	2	1	0	1	1	1	1	0	2	2	0	2
Efficiency	2	1	0	1	1	1	1	0	0	0	0	0
Interpretability	2	1	0	1	1	1	1	0	0	0	0	0
Understandability	2	1	0	1	1	1	1	0	0	0	0	0
Complexity	2	1	0	1	1	1	1	0	0	0	0	0
Volatility	2	1	0	1	1	1	1	0	0	0	2	0
Access Security	0	1	0	0	0	0	0	0	0	0	0	0
Accessibility	0	1	0	0	0	1	0	2	0	0	0	0
Latency	0	1	0	0	0	0	0	0	0	0	0	1
Response Time	0	1	0	0	0	0	0	0	0	0	0	1
Ease of Operation	0	1	0	0	0	0	2	1	0	0	0	0
Availability	0	1	0	0	0	0	0	0	0	0	0	0
Ease of Navigation	0	1	0	0	0	2	0	2	2	0	0	0
Interactivity	0	1	0	0	0	0	0	0	0	0	0	0
Suitability of Representation	0	1	0	0	0	0	0	0	0	0	0	0
Flexibility of Representation	0	1	0	0	0	0	0	0	0	0	0	0

Table 6: Cross matrix 'patterns - information quality criteria'

⇒ 2 means that the information quality criterion is a possible intended effect of the pattern.
⇒ 1 means that the information quality criterion is a possible secondary effect of the pattern.
⇒ 0 means that the information quality criterion is not positively affected by the pattern.

Table 6: Cross matrix 'patterns - information quality criteria' displays which information quality criteria might be affected by the implementation of a pattern. This does mean that the pattern in question should be duly implemented and the right format should be chosen. It is also that not each pattern is applicable in each context. The specific contexts in which an implementation is feasible are documented in the 'Context' section of each pattern. It appears in this matrix that the *Declaration of Failure* is the most powerful pattern, targeting the most possible problems. This is true, but the pattern is mostly applicable in Collaborative Content Creation contexts. The *Splitter* pattern has a lot of possible desirable side effects, but the side effects may be achieved by various implementations. A splitting on a repository level may have positive effects on Response Time and Availability, whereas a splitting information object level may have a positive effect on Response Time and Efficiency. However, the most noted effects are Conciseness and Completeness. It is also clear that the Process-pragmatic criteria are not often the target of the patterns. This is caused by the scope of the research, which focuses on Web 2.0, while problems with Process-pragmatic information quality criteria do occur in all websites. Therefore, patterns to solve these problems are outside of the scope.

7.3 Strategic foci of websites

This chapter describes what strategic foci websites have on the aspect of information quality criteria. We assume that successful websites have achieved their strategic information quality criteria. Therefore, an information quality criterion has to be present in the matrices of observed quality criteria for each website, to qualify as a strategic information quality criterion.

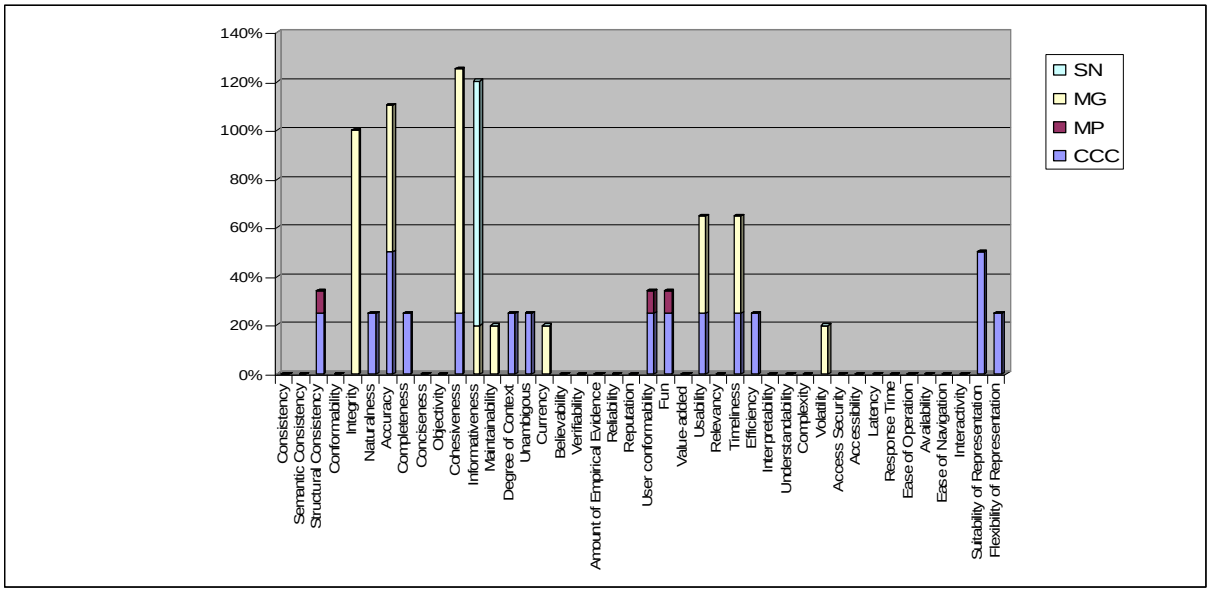

Figure 32: Cumulative Occurrence of strategic foci in websites

As we observe in Figure 32: Cumulative Occurrence of strategic foci in websites, Social Networking websites only have a focus on Informativeness. Whereas Collaborative Content Creation websites have more and often diverse strategic foci. The same applies for Metadata Generation websites, although we see a focus on

cohesiveness and integrity for all Metadata Generation websites. Our assumption is that in this graph we will only see Desired factors and Exciting factors. We see that the big stacks from the graphs, as cumulated in Figure 30: Cumulative Occurrence of Information Quality Criteria in websites, the Process Pragmatics and Relevancy have completely disappeared in the graph with strategic foci. This strengthens the idea that these criteria are indeed Expected Factors, and therefore not a strategic focus.

7.4 Discussion on Patterns

Based on the comparisons made, we can observe several patterns. *Trusted Contributors*, *Rating Engine*s, *Recommendation Engine*, *Search Engine*s and *Tag Engine*s are widespread among all types of Web 2.0 sites. However, implementations differ.

7.4.1 Declaration of Failure

Although this pattern was supposed to be effective mainly in Collaborative Content Creation environments (100%), it was found at other types of websites as well. In Media Provision environments, usage is low, but there is potential for growth. *Declaration of Failure* can indicate a lot of discrete quality problems with respect to e.g. lighting of a picture.

7.4.2 Splitter

Splitter patterns were only used in Collaborative Content Creation settings, although they might be valuable as well in other environments.

7.4.3 Mark-up Tools

Mark-up Tools and *Style Sheets* were often found in the Collaborative Content Creation and Media Provision environments.

7.4.4 Partner Up

Partner Up programs, although it would be expected that they would be dominant among Media Provision, are found among Metadata Generation and Social Networking as well. These programs are in these environments the result of the hybrid behaviour of websites. In Collaborative Content Creation environments *Partner Up* programs are unknown. There are perhaps opportunities in this field.

7.4.5 Upcoming

*Upcoming Section*s are not that widespread, but are unknown in Collaborative Content Creation Settings, although an *Upcoming Section* might boost popularity, overall quality and attention to an information object in Collaborative Content Creation environments. This is perhaps an opportunity.

7.4.6 Version Control

Version Control was found only at Collaborative Content Creation websites. At other websites, *Version Control* was a result of Collaborative Content Creation behaviour, displayed by websites which were not classified as such.

7.4.7 Remember to Forget

Remember to Forget was found at Metadata Generation websites and Media Provision websites, but not at Collaborative Content Creation or Social Networking. This is not because these websites do not strive to be timely, but because these websites have a hard time actually deleting information objects. At the Metadata Generation websites on which this pattern was not observed, it even imposed problems. The leading movie at IMDB.com and the leading board game at boardgamegeek.com are nigh unchanged for years, while tastes might have changed.

7.5 Discussion on Information Quality Criteria

As shown in Table 6: Cross matrix 'patterns - information quality criteria', there are some patterns which affect many information quality criteria. Although it appears that the process pragmatic criteria have no pattern that can be applied to improve performance on these criteria, this is not true. There are many patterns which can be applied, but these patterns are not specific Web 2.0, but are general web-patterns, such as the *Fat Client (Mahemoff 2006)*, which tackles problems with regard to Response Time.

7.5.1 Syntactics

Consistency is a major concern for all Metadata Generation and Social Networking websites, although this focus is more on Structural Consistency than on Semantic Consistency. Semantic Consistency is of little interest to most websites in the research population, perhaps this is caused by the relative effort that has to be directed for little improvement on this information quality criterion. Conformability is of interest to some Collaborative Content Creation and Media Provision websites, but of major interest to all Metadata Generation and Social Networking websites. This indicates that Conformability might be a hygiene factor to those websites. Integrity (whether the scope of the metadata is adequate) is of strategic interest to all Metadata Generation websites, but of limited interest to other websites. Naturalness was found to be interesting to Collaborative Content Creation and Metadata Generation websites.

7.5.2 Semantics

For Collaborative Content Creation semantic information quality criteria are found to be of relative importance. Five of these criteria can be classified as a possible strategic focus of one of the Collaborative Content Creation websites. It shows signs that Collaborative Content Creation websites have more freedom in choosing their strategic focus than other types of websites. The overall semantic information quality criteria are not the focus of the Media Provision websites. This might be a problem caused by the selection of the research population, but the freedom to upload and share makes it hard to exert control over these information objects. A scientific video website would emphasize Accuracy, but such a website was not included in the research population. Only the criterion Currency was found to be of interest to all Media Provision websites. Metadata Generation websites emphasize Accuracy, which is often also a strategic information quality dimension. In the research Cohesiveness was found to be of strategic importance to all Metadata Generation websites. This is because metadata generated about a cluster of concepts and phenomena is of little value. For Social Networking sites, Semantics appear to be of relative importance,

with Informativeness as a strategic focus of all websites in the research population. This might indicate that Informativeness is a hygiene factor for Social Networking sites.

7.5.3 User-Pragmatics

Believability was found to be a criterion of interest to most Collaborative Content Creation websites and to all Metadata Generation websites. In Media Provision environments, Verifiability was found to be neglected, whereas in other environments there were websites which displayed interest in Verifiability. Reliability is of importance to Social Networking websites. This category does not show many hints of patterns which could lead to conclusions. Of the websites in the research population, only three had a user-pragmatic criterion as strategic focus. One time User-conformability and two times Enjoyability.

7.5.4 Information-Pragmatics

In this category, the criterion Relevancy is important to all websites in the research population, although it is not of strategic importance to any website. This reflects the fact that all website aim to produce relevant information for their information consumers, but cannot strategically focus on making information objects relevant. When Usability is a criterion of interest, it is often a strategic focus. Timeliness is of importance to Collaborative Content Creation and Metadata Generation websites. In case of Metadata Generation, this can be caused by the nature of the object that is the source of the metadata generated. If that is highly volatile, the focus should be more on Timeliness. Some websites have timeliness as a strategic focus. Efficiency is found to be of importance in all environments but Media Provision. Understandability is of importance at Collaborative Content Creation and Metadata Genereation websites, but not at Media Provision and Social Networking websites.

7.5.5 Process-Pragmatics

Most Process-Pragmatics are hygiene factors to the information repositories that have attention for these information quality criteria. Only two criteria were found to be of strategic focus, Suitability of Representation and Flexibility of Representation. Access security was found of little importance, under the influence of free sharing and public licenses. Only at Social Networking websites this was considered important, due to the private information incorporated in these websites.

7.6 Validation of Methods and Results

The results as presented in this thesis are tested for rigor at each stage of the research. We will briefly summarize the steps taken to validate the results, and may conclude that the results are sufficiently valid to allow generalization to the entire population of Web 2.0 sites.

7.6.1 Validity of Information Quality Criteria

The validity of information quality criteria and the framework presented are governed by the thorough and systematic literature review as presented in Chapter 2.1: Justification. A remark should be made in this respect regarding the criteria user-conformability and enjoyability. These criteria have not been published in literature at the moment, and are not yet accepted by the research community. However, we have

confidence that these information quality criteria will be accepted shortly as a basis for research in information quality in Web 2.0.

7.6.2 Validity of Research Population

The research population is compromised of the top 50 websites, as is described in Chapter 3.2: Justification of the research population. This is enriched by adding new websites to the research population in an ad hoc fashion. But since the most successful websites are covered, the research population should be valid to be generalized.

7.6.3 Validity of Categorization according to Behaviour

The categorization according to the behaviour displayed by websites was initiated in a meeting with experts, as those four types of behaviour were found among websites. This was validated by a survey among users as stated in Chapter 3.1.5: Validation of the classification of websites. This resulted in a mean linearly weighted agreement between the raters and the researcher of $\kappa(wl)= 0,532$ and a mean quadratic weighted agreement between the raters and the researcher of $\kappa(wq)= 0,607$ (Vanbelle & Albert, 2009). This indicates a moderate to substantial agreement, according to Landis & Koch (1977). This gives confidence that the description of Web 2.0 as a set of behaviours, and the classification of websites according to those behaviours is feasible.

7.6.4 Validity of the Grounded Theory

The Grounded Theory was used in the research, with some adaptations since the research population was compromised of websites instead of humans. In the framework of the Grounded Theory, different approaches were used, as described in Chapter 4.4.1: Data collection. These included Observatory Research, Participatory Research and informal conversations with both users and specialists. The outcome of the Grounded Theory is valid, since of each pattern which was created, several implementations were observed.

7.6.5 Validity of Patterns

The Patterns in itself are valid because multiple implementations of each pattern have been observed. Since patterns imply that it is a piece of behaviour that is observed multiple times, this makes these patterns inherently valid. The question about the completeness of the list of patterns is acknowledged in Appendix E: Possible future patterns.

7.6.6 Validity of the AIM 4 IQ

The created approach for information quality has a solid rooting in the existing body of literature, and has been validated by the elaboration of fictional or non fictional scenarios. Since it is also a part of the outcome of this research, it should be validated further by action research or implementation of this approach in real websites, with real problems.

7.7 Rooting in Existing Literature

The pattern language created is not on its own. It is captured among quality frameworks and other patterns languages. From different perspectives, this pattern language has a place. Some of these perspectives are elaborated below.

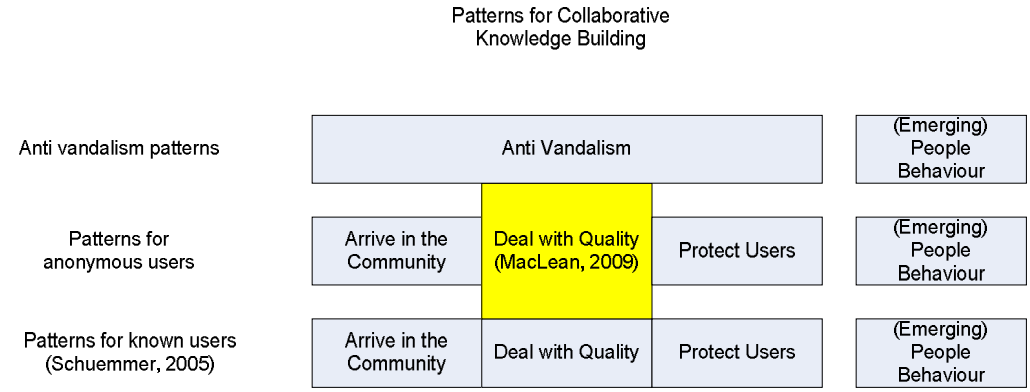

Figure 33: Patterns for Collaborative Knowledge Building

This figure describes patterns which can be used in the creation of processes in software and lay-out to facilitate and encourage desired behaviour of users. We see in Figure 33: Patterns for Collaborative Knowledge Building, that the 'Deal with Quality' for anonymous users is a problem which borders with 'Arriving in the Community' and 'Protecting Users'. But also that in a context where users are anonymous, both anti vandalism patterns and patterns for known users can provide valuable input. Since we have not found a pattern language for arriving in a Web 2.0 community, nor patterns for protecting users in Web 2.0, these are both areas of future research. Also, anti-vandalism practices were found during the research, but are a specific category, and therefore should be documented in a future research. Besides this framework, there is another layer of patterns, which describe the behaviour that people start to display when they are collaboratively working in a knowledge building environment, which in turn borders to the description of human behaviour in Psychology. Some of the links with the pattern language of Till Schuemmer (2005, page 131) are displayed in Figure 34: Links to Pattern Language of Schuemmer (2005). Note that *Remember to Forget*, which is inspired by Schuemmer has no links, since it was not even a full pattern in his pattern language. There is also a not displayed link between the *Letter of Recommendation* and the *Recommendation Engine*, which is that a high rating in the *Recommendation Engine*, is quite similar to a *Letter of Recommendation*, both being a statement that the information object is of high quality.

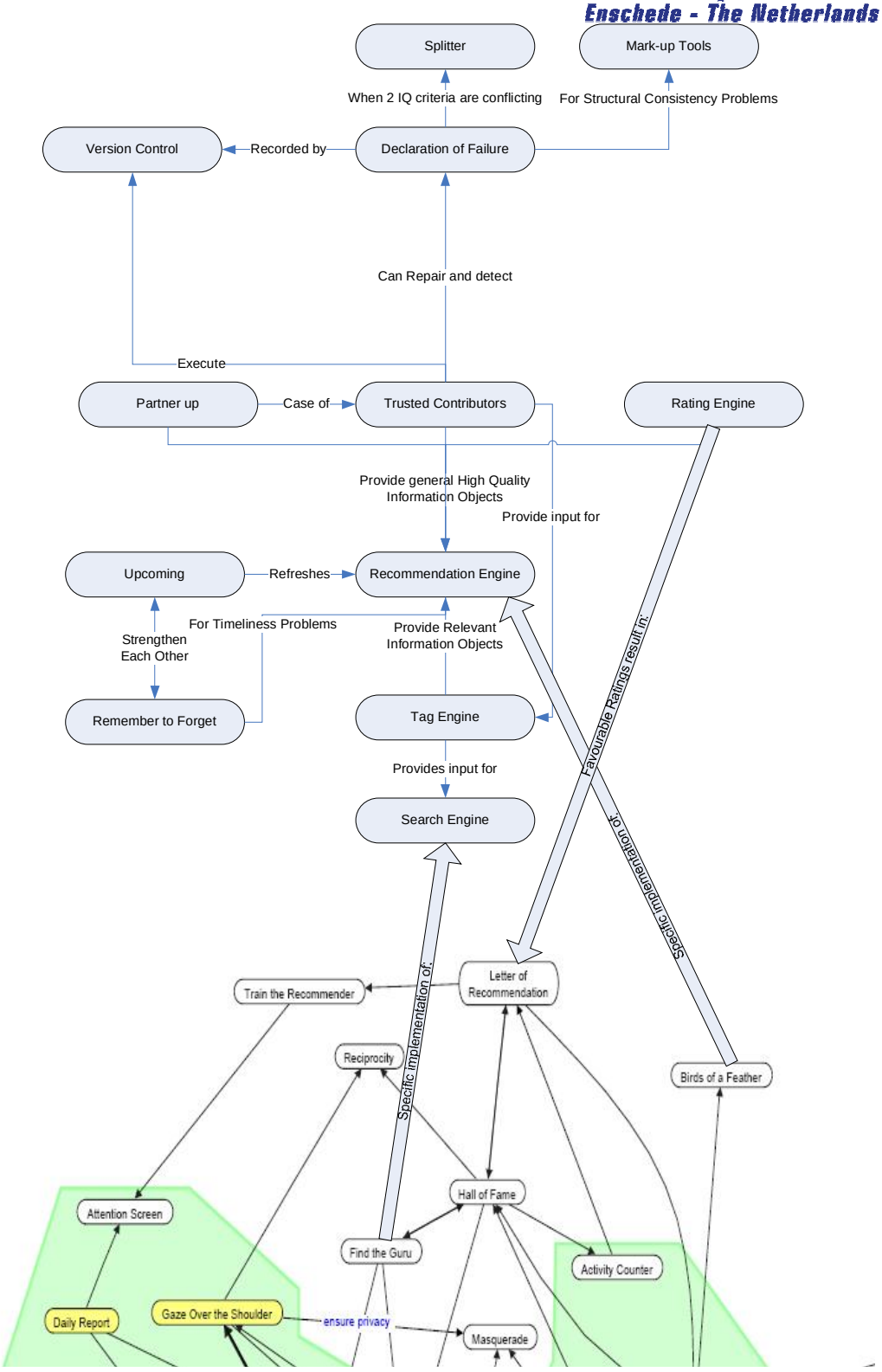

Figure 34: Links to Pattern Language of Schuemmer (2005)

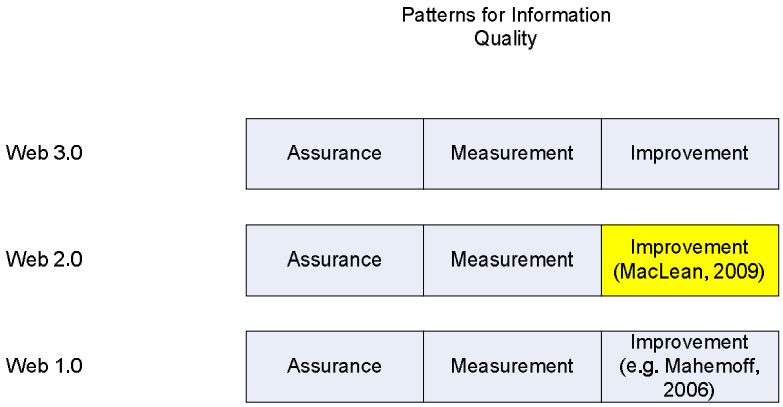

Figure 35: Patterns for Information Quality

In Figure 35: Patterns for Information Quality, it is shown that the pattern language constructed focuses only on improving information quality, and only in a Web 2.0 context. The absence of pattern language on information quality improvement in Web 1.0 proved an obstacle, since many of the Process-Pragmatic information quality dimensions were covered by design patterns for websites, but not documented as information quality patterns for the Web. As a result, examples as 'Fat Client (Mahemoff, 2006) were found, which have a positive result on Response Time, but which was only found as a design pattern. The measurement of information quality is an area of research in which much work has been done by Besiki Stvilia (Stvilia et al, 2008). But the work he has done is not documented as a pattern language, and therefore not displayed in this picture. In this graph there is a layer which is called Web 3.0, a still vague concept which is the future of the internet. It is displayed in this graph to underline the fact that in the future information quality patterns may be built on this existing body of knowledge.

8 Conclusions and Recommendations

8.1 Conclusions

Information Quality criteria differ in importance in different environments. As a result, websites may focus on any of the information quality criteria that are desired or exciting factors concerning their audience and information objects. This choice is strategic, since information is the only product of the company, the quality of that product is of course a strategic choice.

Media Provision sites tend to have a strong focus on Currency, which is the age of an information object, while Collaborative Content Creation sites seem to focus more on Timeliness, which is about whether an article is up to date.

Social Networking websites and Metadata Generation websites excel at achieving Informativeness, which is caused mainly by strict *Style Sheets* and limited options to choose in a rigid format. Therefore the information presented has a high cognitive density. The list-like display also achieves great structural and semantic consistency.

Naturalness seems to be a focus of Collaborative Content Creation sites, but can be neglected when special interest groups are the focus of the website.

Integrity (whether the scope of the metadata is adequate) is a strategic focus of all Metadata Generation websites in the research population. It is achieved by thorough knowledge of the field and the patterns: All Elements Present and Optional Elements when Needed. Other types of websites seem to have little interest in Integrity.

Accuracy is important for all Collaborative Content Creation and Metadata Generation websites, except Uncyclopedia, which focuses on Fun; which is hard to combine with accuracy. For Media Provision websites this seems of smaller interest.

Cohesiveness is a strategic focus for all Metadata Generation websites. This is logical, since generating Metadata about a broad range of subjects is impossible. For Collaborative Content Creation websites Cohesiveness is important as well.

Most of the work done in Collaborative Content Creation environments is manual labour, although easy tasks like grammar, spelling, linking and license checking can easily be done by bots. Bots can tackle clear often found problems, such as the linking of websites in different languages.

As a result of the dominance of manual labour, there is little difference between the approach of subjective and objective quality criteria. Both might be tackled with a Declaration of Failure.

Vandalism is a major threat to collaborative content creation sites, which have developed elaborate counter measures to battle vandalism. It is not covered in this

thesis, since it is not an information quality problem, although vandalism is about destroying information.

8.2 Recommendations

8.2.1 Research Recommendations

More research is needed, as always, to further analyze and investigate Web 2.0 on the level of methods and patterns. Numerous observations of candidate patterns have been made, but these candidates are smaller patterns than the patterns included in the pattern language.

A quantitative research approach may have concluded that the information objects provided on one website are better than the information objects as provided on another site. In this research we analyzed the methods to come to the high quality information objects. Future research should try to analyze whether websites which have implemented these patterns, do indeed outperform their competitors.

The AIM 4 IQ method as presented in Chapter 6 should be tested against a real website. This would provide additional validation for the presented method, as well as the missing link between theory and practice.

8.2.2 Practice Recommendations

Web 2.0 sites should investigate what exactly the strategic focus of the website is on the aspect of information quality. Information repositories should focus on a consistent approach, in which the information objects have the right quality.

Emerging Web 2.0 sites may investigate what the strategic focus of the main competitors is, and either try to outperform the competition, or focus on a ‘gap in the market’, creating information objects which have a high quality on a criterion which is not addressed by other websites.

8.3 Outlook

This chapter is highly speculative, and consists of things that may happen, or may not.

8.3.1 *Declaration of Failure*

Collaborative Content Creation will probably migrate towards photo and video, as the *Mark-up Tools* for these media become more user-friendly, much in the same trend as all software becomes more user-friendly. When that happens, the *Declaration of Failure* will be applicable in video or photo enhancement as well. It is conceivable that this will happen, because *Mark-up Tools* tend to become more user-friendly and advanced. When a sufficient number of people are familiar with the software to be used to enhance photos or video, the *Declaration of Failure* can be used to enhance these information objects as well. It is likely that the websites aimed at photo sharing will embrace this technology, but probably in a different section of the information repository, since not all users appreciate meddling with their pictures. This separate section will be aimed at collectively generating high quality pictures.

8.3.2 Waste Recycling

We see over the ages that people who are becoming more civilized have better ways to handle the waste that is produced. From dumping through burning to recycling, which is generating value from waste. In much the same way, most information objects have some (even if it is small) value. This, combined with the ever decreasing cost for information storage, may result in Wiki Waste Recycling. Then, instead of deleting content, it is moved to another wiki. This creates small band wikis, personalia wikis, not-quite-so-famous-football-player-wikis, and even slander wikis. Wikipedians stop deleting, start moving articles around, until these networks become once again interlinked and all knowledge will be available. By Recycling the garbage, (which is a high-level implementation of the *Splitter*-pattern) both worlds can be combined, getting relative irrelevant information objects out of the repository, and at the same time keeping them for those information consumers which are interested in them. We expect that, if this happens, it will be within 5-8 years. It will probably be initiated as a way to circumvent problems among Wikipedians, when some want to delete an article, whereas others want to keep it.

8.3.3 Social Truthfinding

In the western culture, which hopefully will evolve beyond modernism and post-modernism, opinions become more personal, less universal. Minority truths may emerge, as people find that beliefs and truths are mostly personal and hardly universal. Social truth finding is perhaps a new trend, a new road, in which several years from now, people link their personal profile to their goals, beliefs and hopes. This is speculative, and whether it will happen is unclear, it is the logical mash-up of minority opinion truths and social networking. We expect this to happen on the long term, but it is mostly a matter of one major social networking site, who embraces this concept. Perhaps this is a gap in the market as well.

8.3.4 How-to Wikis

We saw the emergence of Wikihow.com, now a top 500 website in many countries. Yet, we believe that there is room for such a website in other languages, much as with Wikipedia, marktplaats.nl and many other websites which benefit from network effects. Therefore, the battle for the Dutch wiki-how-to-manual may start. Wikihow.com, Wikipedia, and perhaps others. This will happen for many countries though, as people start to form new communities which focus on this aspect. The emergence of such a community might benefit from internal struggles in the Wikipedia community, resulting in members leaving the community, and forming a new team to create a how to manual. It will not compete with Wikipedia on the information object side, but might be a fierce competitor on the side of recruiting new active members. We expect this to happen within 1-2 years, as the first signs of these sites are already appearing.

9 References

[Alexa, 2009]
http://www.alexa.com/site/ds/top_sites?cc=NL&ts_mode=country&lang=none on 3 February, 2009.

[Alexander, 1977]
C. Alexander, S. Ishikawa, and M. Silverstein. *A pattern language*. Center for Environmental Structure Series. Oxford University Press, 1977.

[Alexander, 1979]
Christopher Alexander. *The timeless way of building*. Oxford University Press, New York, 1979.

[Allan, 2003]
George Allan. A critique of using grounded theory as a research method. *Electronic Journal of Business Research Methods*, 2(1), July 2003.

[Court Ruling 2005]
LJN: AZ2550, Rechtbank 's-Gravenhage , zittingsplaats Amsterdam , AWB 05/8898

[Court Ruling 2006]
LJN: AZ5686, Hoge Raad , R06/013HR

[Dedeke, 2000]
A. Dedeke. A conceptual framework for developing quality measures for information systems. In *Proceedings of 5th International Conference on Information Quality*, pages 126–128, 2000.

[Dick 2000]
Bob Dick. Grounded theory: A thumbnailsketch, Retrieved 10 February 2009. www.scu.edu.au/schools/gcm/ar/arp/grounded.html

[Eppler & Wittig, 2000]
M. J. Eppler and D. Wittig. Conceptualizing information quality:a review of information quality frameworks from the last ten years. In *Proceedings of the 2000 Conference on Information Quality*, pages 83–96, 2000.

[Gamma et al, 1995]
Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. *Design Patterns*. Addison-Wesley Professional, January 1995.

[Giles 2005]
Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, December 2005.

[Glaser and Strauss, 1967]
Barney G. Glaser and Anselm L. Strauss. *Discovery of Grounded Theory: Strategies for Qualitative Research.* Aldine, June 1967.

[Knight & Burn, 2005]
S. A. Knight and J. Burn. Developing a framework for assessing information quality on the world wide web. *Informing Science*, 8:159–172, 2005.

[Landis & Koch, 1977]

J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, March 1977.

[Lee, 2005]

Daniel H. Lee. Contextual it business value and barriers:an e-government and e-business perspective. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, pages 1–10, 2005.

[Mahemoff, 2006]

Micheal Mahemoff. *Ajax Design Patterns*. O'Reilly Media, first edition, June 2006.

[Meszaros & Doble, 1996]

Gerard Meszaros and Jim Doble. A pattern language for pattern writing. In *The 3rd Pattern Languages of Programming conference*, pages 1–33, 1996.

[Millward Brown Optimor, 2009] Millward Brown Optimor. Google is the world's first $100 billion brand, number one for the third year running. April 2009. Retrieved 4 May 2009. http://www.millwardbrown.com/Sites/Optimor/Content/KnowledgeCenter/BrandzRanking.aspx

[Naumann & Rolker, 2004]

Felix Naumann and Claudia Rolker. Assessment methods for information quality criteria. In *Proc. IQ2000 MIT Conference on Information Quality*, pages 148–162, 2000.

[O'Reilly, 2009] Tim O'Reilly. What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html, accessed 4 May 2009

[Parker et al, 2006]

M. B. Parker, V. Moleshe, R. De_la_harpe, and G. B. Wills. An evaluation of information quality frameworks for the world wide web. In *8th Annual Conference on WWW Applications, 6-8th September 2006, Bloemfontein, Free State Province, South Africa.*, 2006.

[Price et al, 2008]

Rosanne Price, Dina Neiger, and Graeme Shanks. Developing a measurement instrument for subjective aspects of information quality. *Communications of the AIS*, 22(3):49–74, January 2008.

[Price & Shanks 2005]

Rosanne Price and Graeme Shanks. A semiotic information quality framework: development and comparative analysis. *Journal of Information Technology*, 20(2):88–102, June 2005.

[Price & Shanks, 2005b]

Rosanne J. Price and Graeme Shanks. Empirical refinement of a semiotic information quality framework. In *Proceedings of the 38th Hawaii International Conference on System Sciences*, pages 1–10, 2005.

[Rao & Osei-Bryson, 2007]

L. Rao and K. M. Osei-Bryson. Towards defining dimensions of knowledge systems quality. *Expert Systems with Applications*, 33(2):368–378, 2007.

[Salganik et al, 2006]

Matthew J. Salganik, Peter S. Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311:854–856, February 2006.

[Schuemmer, 2005]

Till Schuemmer. *Patterns for Building Communities in Collaborative Systems*. PhD thesis, 2005.

[Schwartz & Russo, 2004]

Ruth B. Schwartz and Michele C. Russo. How to quickly find articles in the top is journals. *Commun. ACM*, 47(2):98–101, 2004.

[Shanks & Corbitt, 1999]

G. Shanks and B. Corbitt. Understanding data quality: Social and cultural aspects. In *Proceedings of the 10th Australasian Conference on Information Systems*, pages 785–797, 1999.

[Spiegel.de 2009]

Scandal season and name changes in german politics. Retrieved May 4, 2009 http://www.spiegel.de/international/germany/0,1518,607269,00.html

[Strong et al, 1997]

D. M. Strong, Y. W. Lee, and R. Y. Wang. Data quality in context. *Communications of the ACM*, 40(5):103–110, 1997.

[Stvilia et al, 2007]

B. Stvilia, L. Gasser, M. B. Twidale, and L. C. Smith. A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, 58(12):1720–1733, 2007.

[Stvilia et al, 2008]

B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001, 2008.

[Su & Jin, 2007]

Y. Su and Z. Jin. Assuring information quality in knowledge intensive business services. In *2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007*, WiCOM 2007, pages 3238–3241, Department of Business Strategy and Policy, Tsinghua University, 100084 Beijing, China.

[Vanbelle & Albert, 2009]

S. Vanbelle and A. Albert. A note on the linearly weighted kappa coefficient for ordinal scales. *Statistical Methodology*, 6(2):157–163, March 2009.

[Wang & Strong, 1996]

Richard Y. Wang and Diane M. Strong. Beyond accuracy: what data quality means to data consumers. *J. Manage. Inf. Syst.*, 12(4):5–33, 1996.

[Zultner & Mazur, 2006]

Richard E. Zultner and Glenn H. Mazur. The kano model: Recent developments. In *The Eighteenth Symposium on Quality Function Development*, December 2006.

10 Appendix A: Crosstables

These are the raw data used to construct the graphs in chapter 7

10.1 Websites-patterns

This table contains the data used to construct Figure 25: Usage of Patterns in Web 2.0 sites

	Website/Patterns	Declaration of Failure	Splitter	Mark-up tools	Partner Up	*Rating Engine*	Recommendation Enginge	Trusted Contributors	Search Engine	Tag Engine	Upcoming	Version Control
CCC	Wikipedia	1	1	1	0	0	1	1	1	0	0	1
CCC	wikihow.com	1	1	0	0	1	1	1	1	1	0	1
CCC	wikia.com	1	1	1	0	1	1	1	1	1	0	1
CCC	Uncyclopaedia	1	1	1	0	1	1	1	1	1	0	1
	Average Usage	1,00	1,00	0,75	0,00	0,75	1,00	1,00	1,00	0,75	0,00	1,00
MP	Youtube	1	0	1	1	1	1	1	1	1	1	0
MP	Marktplaats.nl	0	0	0	1	1	1	1	1	1	0	0
MP	Blogger.com	0	0	1	1	0	1	1	1	1	0	0
MP	Flickr	0	0	1	0	1	1	1	1	1	1	0
MP	Dumpert.nl	0	0	0	1	1	1	1	1	1	0	0
MP	Ebay.com	0	0	0	1	1	1	1	1	1	1	1
MP	Web-log.nl	0	0	1	0	0	0	1	1	1	0	0
MP	Photobucket	0	0	1	0	1	1	1	1	1	1	0
MP	Wordpress.com	0	0	1	1	0	1	1	1	1	0	0
MP	Vimeo.com	0	0	1	0	1	1	1	1	1	1	0
MP	Icanhascheezburger.com	0	0	1	0	1	1	1	1	0	1	1
	Average Usage	0,09	0,00	0,73	0,55	0,73	0,91	1,00	1,00	0,91	0,55	0,18
MG	IMDB.com	0	0	0	1	1	1	1	1	1	0	0
MG	Zoover	0	0	0	0	1	1	1	1	1	0	0
MG	Boardgamegeek	0	0	0	0	1	1	1	1	1	1	0
MG	Digg.com	0	0	0	1	1	1	1	1	1	1	0
MG	Last.fm	1	0	0	1	1	1	0	1	1	1	1
	Average Usage	0,20	0,00	0,00	0,60	1,00	1,00	0,80	1,00	1,00	0,60	0,20
SN	Myspace.com	0	0	0	1	1	1	1	1	1	0	0
SN	Hyves.nl	0	0	1	1	1	1	1	1	1	1	0
SN	Facebook	0	0	0	0	0	1	1	1	1	0	0
	Average Usage	0,00	0,00	0,33	0,67	0,67	1,00	1,00	1,00	1,00	0,33	0,00

Table 7: Cross matrix 'websites-patterns'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.
⇒ 1 means that an implementation of the pattern is observed
⇒ 0 means that implementation of the pattern is not observed

10.2 Websites – Information Quality Criteria

These tables were used to construct Figure 30: Cumulative Occurrence of Information Quality Criteria in websites

	Websites/IQ Criteria	Consistency	Semantic Consistency	Structural Consistency	Conformability	Integrity	Naturalness
CCC	Wikipedia	1	1	1	1	1	1
CCC	wikihow.com	0	0	0	1	0	1
CCC	wikia.com	0	0	0	0	0	0
CCC	Uncyclopaedia	0	0	0	0	0	1
	Average Usage	0,25	0,25	0,25	0,50	0,25	0,75
MP	Youtube.com	0	0	0	1	0	0
MP	Marktplaats.nl	1	0	1	1	0	0
MP	Blogger.com	1	0	1	0	0	0
MP	Flickr	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	0
MP	Ebay.com	1	0	1	1	0	0
MP	Web-log.nl	1	0	1	0	0	0
MP	Photobucket	0	0	0	0	0	0
MP	Wordpress.com	1	0	1	0	0	0
MP	Vimeo.com	0	0	0	1	0	0
MP	Icanhascheezburger.com	1	0	1	0	0	0
	Average Usage	0,55	0,00	0,55	0,36	0,00	0,00
MG	IMDB.com	1	0	1	1	1	1
MG	Zoover	1	0	1	1	1	1
MG	Boardgamegeek	1	1	1	1	1	0
MG	Digg.com	1	0	1	1	1	1
MG	Last.fm	1	0	1	1	1	1
	Average Usage	1,00	0,20	1,00	1,00	1,00	0,80
SN	Myspace.com	1	0	1	1	0	0
SN	Hyves.nl	1	0	1	1	0	0
SN	Facebook	1	0	1	1	1	0
	Average Usage	1,00	0,00	1,00	1,00	0,33	0,00

1 = Information quality criterion of interest

0 = Information quality criterion of limited interest.

Table 8: Cross matrix 'websites - information quality criteria (Syntactics)'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.

⇒ 1 means that the information quality criterion is considered an important criterion to the website.

⇒ 0 means that the information quality criterion is of limited interest to the website.

	Websites/IQ Criteria	Accuracy	Completeness	Conciseness	Objectivity	Cohesiveness	Informativeness	Maintainability	Degree of Context	Unambigous	Currency
CCC	Wikipedia	1	1	1	1	1	1	1	1	1	0
CCC	wikihow.com	1	0	1	1	1	1	1	1	1	0
CCC	wikia.com	1	1	1	1	1	1	1	1	1	0
CCC	Uncyclopaedia	0	0	1	0	0	0	1	0	0	0
	Average Usage	0,75	0,50	1,00	0,75	0,75	0,75	1,00	0,75	0,75	0,00
MP	Youtube.com	0	0	0	0	0	0	0	0	0	1
MP	Marktplaats.nl	1	0	1	0	1	0	0	1	0	1
MP	Blogger.com	0	0	0	0	0	0	0	0	0	1
MP	Flickr	0	0	0	0	0	0	0	0	0	1
MP	Dumpert.nl	0	0	0	0	0	0	0	0	0	1
MP	Ebay.com	1	0	1	0	1	0	1	1	0	1
MP	Web-log.nl	0	0	0	0	0	0	0	0	0	1
MP	Photobucket	0	0	0	0	0	0	0	0	0	1
MP	Wordpress.com	0	0	0	0	0	0	0	0	0	1
MP	Vimeo.com	0	0	0	0	0	0	0	0	0	1
MP	Icanhascheezburger.com	0	0	1	0	0	0	0	0	0	1
	Average Usage	0,18	0,00	0,27	0,00	0,18	0,00	0,09	0,18	0,00	1,00
MG	IMDB.com	1	1	1	0	1	1	0	0	0	0
MG	Zoover	1	1	1	0	1	1	1	1	1	0
MG	Boardgamegeek	1	1	0	0	1	1	1	1	1	1
MG	Digg.com	1	0	0	0	1	1	0	0	0	1
MG	Last.fm	1	0	1	0	1	1	1	1	0	1
	Average Usage	1,00	0,60	0,60	0,00	1,00	1,00	0,60	0,60	0,40	0,60
SN	Myspace.com	0	0	1	0	1	1	1	1	0	1
SN	Hyves.nl	0	0	1	1	1	1	1	1	0	1
SN	Facebook	1	0	1	1	1	1	1	1	0	1
	Average Usage	0,33	0,00	1,00	0,67	1,00	1,00	1,00	1,00	0,00	1,00

1 = Information quality criterion of interest

0 = Information quality criterion of limited interest.

Table 9: Cross matrix 'websites - information quality criteria (Semantics)'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.
⇒ 1 means that the information quality criterion is considered an important criterion to the website.
⇒ 0 means that the information quality criterion is of little interest to the website.

	Websites/IQ Criteria	Believability	Verifiability	Amount of Empirical Evidence	Reliability	Reputation	User conformability	Fun
CCC	Wikipedia	1	1	0	0	0	0	0
CCC	wikihow.com	1	0	0	0	0	0	0
CCC	wikia.com	1	1	0	0	0	1	0
CCC	Uncyclopaedia	0	0	0	0	0	0	1
	Average Usage	0,75	0,50	0,00	0,00	0,00	0,25	0,25
MP	Youtube.com	0	0	0	0	1	0	1
MP	Marktplaats.nl	1	0	1	0	1	0	0
MP	Blogger.com	0	0	0	1	0	0	0
MP	Flickr	0	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	1	1
MP	Ebay.com	1	0	0	1	1	0	0
MP	Web-log.nl	0	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	1	0	0	0
MP	Vimeo.com	0	0	0	0	1	0	1
MP	Icanhascheezburger.com	0	0	0	0	0	0	1
	Average Usage	0,18	0,00	0,09	0,27	0,36	0,09	0,36
MG	IMDB.com	1	1	0	0	1	0	0
MG	Zoover	1	0	0	0	0	0	0
MG	Boardgamegeek	1	0	1	0	0	0	0
MG	Digg.com	1	0	0	0	0	1	0
MG	Last.fm	1	0	0	1	1	0	1
	Average Usage	1,00	0,20	0,20	0,20	0,40	0,20	0,20
SN	Myspace.com	0	0	0	1	0	0	0
SN	Hyves.nl	0	0	0	1	1	0	0
SN	Facebook	0	1	0	1	0	0	0
	Average Usage	0,00	0,33	0,00	1,00	0,33	0,00	0,00

1 = Information quality criterion of interest

0 = Information quality criterion of limited interest.

Table 10: Cross matrix 'websites - information quality criteria (User Pragmatics)'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.
⇒ 1 means that the information quality criterion is considered an important criterion to the website.
⇒ 0 means that the information quality criterion is of little interest to the website.

	Websites/IQ Criteria	Value-added	Usability	Relevancy	Timeliness	Efficiency	Interpretability	Understandability	Complexity	Volatility
CCC	Wikipedia	0	0	1	1	1	0	1	1	1
CCC	wikihow.com	1	1	1	1	1	1	1	1	0
CCC	wikia.com	0	0	1	1	1	0	1	1	0
CCC	Uncyclopaedia	0	0	1	0	0	0	0	0	0
	Average Usage	0,25	0,25	1,00	0,75	0,75	0,25	0,75	0,75	0,25
MP	Youtube.com	0	0	1	0	0	0	0	0	0
MP	Marktplaats.nl	1	1	1	0	1	0	0	0	1
MP	Blogger.com	0	0	1	0	0	0	0	0	0
MP	Flickr	0	0	1	0	0	0	0	0	0
MP	Dumpert.nl	0	0	1	0	0	0	0	0	0
MP	Ebay.com	1	1	1	0	1	0	0	0	1
MP	Web-log.nl	0	0	1	0	0	0	0	0	0
MP	Photobucket	0	0	1	0	0	0	0	0	0
MP	Wordpress.com	0	0	1	0	0	0	0	0	0
MP	Vimeo.com	0	0	1	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	1	0	0	0	0	0	0
	Average Usage	0,18	0,18	1,00	0,00	0,18	0,00	0,00	0,00	0,18
MG	IMDB.com	1	1	1	1	1	0	1	0	0
MG	Zoover	1	1	1	1	1	0	1	0	1
MG	Boardgamegeek	1	1	1	1	1	0	1	0	1
MG	Digg.com	0	0	1	1	0	0	0	0	1
MG	Last.fm	1	0	1	0	1	0	1	0	1
	Average Usage	0,80	0,60	1,00	0,80	0,80	0,00	0,80	0,00	0,80
SN	Myspace.com	0	0	1	0	1	0	0	0	0
SN	Hyves.nl	0	0	1	0	1	0	0	0	0
SN	Facebook	0	0	1	0	1	0	0	0	0
	Average Usage	0,00	0,00	1,00	0,00	1,00	0,00	0,00	0,00	0,00

1 = Information quality criterion of interest

0 = Information quality criterion of limited interest.

Table 11: Cross matrix 'websites - information quality criteria (Information Pragmatics)'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.
⇒ 1 means that the information quality criterion is considered an important criterion to the website.
⇒ 0 means that the information quality criterion is of little interest to the website.

	Websites/IQ Criteria	Access Security	Accessibility	Latency	Response Time	Ease of Operation	Availability	Ease of Navigation	Interactivity	Suitability of Representation	Flexibility of Representation
CCC	Wikipedia	0	1	1	1	1	1	1	1	1	1
CCC	wikihow.com	0	1	1	1	1	1	1	0	1	1
CCC	wikia.com	0	1	1	1	1	1	1	0	1	1
CCC	Uncyclopaedia	0	1	1	1	1	1	1	0	0	0
	Average Usage	0,00	1,00	1,00	1,00	1,00	1,00	1,00	0,25	0,75	0,75
MP	Youtube.com	0	1	1	1	1	1	1	0	1	0
MP	Marktplaats.nl	0	1	1	1	1	1	1	0	1	1
MP	Blogger.com	0	1	1	1	1	1	1	0	0	1
MP	Flickr	0	1	1	1	1	1	1	1	0	0
MP	Dumpert.nl	0	1	1	1	1	1	1	0	1	0
MP	Ebay.com	0	1	1	1	1	1	1	0	1	1
MP	Web-log.nl	0	1	1	1	1	1	1	0	0	0
MP	Photobucket	0	1	1	1	1	1	1	1	1	0
MP	Wordpress.com	0	1	1	1	1	1	1	0	1	0
MP	Vimeo.com	0	1	1	1	1	1	1	0	1	0
MP	Icanhascheezburger.com	0	1	1	1	1	1	1	0	0	0
	Average Usage	0,00	1,00	1,00	1,00	1,00	1,00	1,00	0,18	0,64	0,27
MG	IMDB.com	0	1	1	1	1	1	1	1	1	0
MG	Zoover	0	1	1	1	1	1	1	1	1	1
MG	Boardgamegeek	0	1	1	1	1	1	1	1	1	0
MG	Digg.com	0	1	1	1	1	1	1	1	1	0
MG	Last.fm	0	1	1	1	1	1	1	1	1	0
	Average Usage	0,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,20
SN	Myspace.com	0	1	1	1	1	1	1	1	1	1
SN	Hyves.nl	1	1	1	1	1	1	1	1	0	0
SN	Facebook	1	1	1	1	1	1	1	0	1	0
	Average Usage	0,67	1,00	1,00	1,00	1,00	1,00	1,00	0,67	0,67	0,33

1 = Information quality criterion of interest

0 = Information quality criterion of limited interest.

Table 12: Cross matrix 'websites - information quality criteria (Process Pragmatics)'

⇒ CCC stands for Collaborative Content Creation websites.
⇒ MP stands for Media Provision websites.
⇒ MG stands for Metadata Generation websites.
⇒ SN stands for Social Networking websites.
⇒ 1 means that the information quality criterion is considered an important criterion to the website.
⇒ 0 means that the information quality criterion is of little interest to the website.

10.3 Strategic foci of websites

These tables contain the raw data used to create Figure 32: Cumulative Occurrence of strategic foci in websites

	Websites/IQ Criteria	Consistency	Semantic Consistency	Structural Consistency	Conformability	Integrity	Naturalness
CCC	Wikipedia	0	0	1	0	0	0
CCC	wikihow.com	0	0	0	0	0	1
CCC	wikia.com	0	0	0	0	0	0
CCC	Uncyclopaedia	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,25	0,00	0,00	0,25
MP	Youtube.com	0	0	0	0	0	0
MP	Marktplaats.nl	0	0	0	0	0	0
MP	Blogger.com	0	0	0	0	0	0
MP	Flickr	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	0
MP	Ebay.com	0	0	0	0	0	0
MP	Web-log.nl	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	0	0	0
MP	Vimeo.com	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	1	0	0	0
	Average Usage	0,00	0,00	0,09	0,00	0,00	0,00
MG	IMDB.com	0	0	0	0	1	0
MG	Zoover	0	0	0	0	1	0
MG	Boardgamegeek	0	0	0	0	1	0
MG	Digg.com	0	0	0	0	1	0
MG	Last.fm	0	0	0	0	1	0
	Average Usage	0,00	0,00	0,00	0,00	1,00	0,00
SN	Myspace.com	0	0	0	0	0	0
SN	Hyves.nl	0	0	0	0	0	0
SN	Facebook	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00

1 =Strategic Focus

0 = No strategic focus

Table 13: Cross matrix 'websites - strategic information quality criteria (Syntactics)'

	Websites/IQ Criteria	Accuracy	Completeness	Conciseness	Objectivity	Cohesiveness	Informativeness	Maintainability	Degree of Context	Unambigous	Currency
CCC	Wikipedia	1	0	0	0	0	0	0	0	0	0
CCC	wikihow.com	0	0	0	0	0	0	0	0	1	0
CCC	wikia.com	1	1	0	0	1	0	0	1	0	0
CCC	Uncyclopaedia	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,50	0,25	0,00	0,00	0,25	0,00	0,00	0,25	0,25	0,00
MP	Youtube.com	0	0	0	0	0	0	0	0	0	0
MP	Marktplaats.nl	0	0	0	0	0	0	0	0	0	0
MP	Blogger.com	0	0	0	0	0	0	0	0	0	0
MP	Flickr	0	0	0	0	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	0	0	0	0	0
MP	Ebay.com	0	0	0	0	0	0	0	0	0	0
MP	Web-log.nl	0	0	0	0	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	0	0	0	0	0	0	0
MP	Vimeo.com	0	0	0	0	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MG	IMDB.com	1	0	0	0	1	0	0	0	0	0
MG	Zoover	1	0	0	0	1	0	1	0	0	0
MG	Boardgamegeek	1	0	0	0	1	0	0	0	0	0
MG	Digg.com	0	0	0	0	1	1	0	0	0	1
MG	Last.fm	0	0	0	0	1	0	0	0	0	0
	Average Usage	0,60	0,00	0,00	0,00	1,00	0,20	0,20	0,00	0,00	0,20
SN	Myspace.com	0	0	0	0	0	1	0	0	0	0
SN	Hyves.nl	0	0	0	0	0	1	0	0	0	0
SN	Facebook	0	0	0	0	0	1	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00

1 =Strategic Focus

0 = No strategic focus

Table 14: Cross matrix 'websites - strategic information quality criteria (Semantics)'

	Websites/IQ Criteria	Believability	Verifiability	Amount of Empirical Evidence	Reliability	Reputation	User conformability	Fun
CCC	Wikipedia	0	0	0	0	0	0	0
CCC	wikihow.com	0	0	0	0	0	0	0
CCC	wikia.com	0	0	0	0	0	1	0
CCC	Uncyclopaedia	0	0	0	0	0	0	1
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,25	0,25
MP	Youtube.com	0	0	0	0	0	0	0
MP	Marktplaats.nl	0	0	0	0	0	0	0
MP	Blogger.com	0	0	0	0	0	0	0
MP	Flickr	0	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	1	0
MP	Ebay.com	0	0	0	0	0	0	0
MP	Web-log.nl	0	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	0	0	0	0
MP	Vimeo.com	0	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	0	0	0	0	1
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,09	0,09
MG	IMDB.com	0	0	0	0	0	0	0
MG	Zoover	0	0	0	0	0	0	0
MG	Boardgamegeek	0	0	0	0	0	0	0
MG	Digg.com	0	0	0	0	0	0	0
MG	Last.fm	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SN	Myspace.com	0	0	0	0	0	0	0
SN	Hyves.nl	0	0	0	0	0	0	0
SN	Facebook	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00

1 =Strategic Focus

0 = No strategic focus

Table 15: Cross matrix 'websites - strategic information quality criteria (User-Pragmatics)'

	Websites/IQ Criteria	Value-added	Usability	Relevancy	Timeliness	Efficiency	Interpretability	Understandability	Complexity	Volatility
CCC	Wikipedia	0	0	0	0	0	0	0	0	0
CCC	wikihow.com	0	1	0	1	1	0	0	0	0
CCC	wikia.com	0	0	0	0	0	0	0	0	0
CCC	Uncyclopaedia	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,25	0,00	0,25	0,25	0,00	0,00	0,00	0,00
MP	Youtube.com	0	0	0	0	0	0	0	0	0
MP	Marktplaats.nl	0	0	0	0	0	0	0	0	0
MP	Blogger.com	0	0	0	0	0	0	0	0	0
MP	Flickr	0	0	0	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	0	0	0	0
MP	Ebay.com	0	0	0	0	0	0	0	0	0
MP	Web-log.nl	0	0	0	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	0	0	0	0	0	0
MP	Vimeo.com	0	0	0	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MG	IMDB.com	0	0	0	0	0	0	0	0	0
MG	Zoover	0	1	0	1	0	0	0	0	0
MG	Boardgamegeek	0	1	0	0	0	0	0	0	0
MG	Digg.com	0	0	0	1	0	0	0	0	1
MG	Last.fm	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,40	0,00	0,40	0,00	0,00	0,00	0,00	0,20
SN	Myspace.com	0	0	0	0	0	0	0	0	0
SN	Hyves.nl	0	0	0	0	0	0	0	0	0
SN	Facebook	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

1 =Strategic Focus

0 = No strategic focus

Table 16: Cross matrix 'websites - strategic information quality criteria (Information-Pragmatics)'

	Websites/IQ Criteria	Access Security	Accessibility	Latency	Response Time	Ease of Operation	Availability	Ease of Navigation	Interactivity	Suitability of Representation	Flexibility of Representation
CCC	Wikipedia	0	0	0	0	0	0	0	0	1	0
CCC	wikihow.com	0	0	0	0	0	0	0	0	1	1
CCC	wikia.com	0	0	0	0	0	0	0	0	0	0
CCC	Uncyclopaedia	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,50	0,25
MP	Youtube.com	0	0	0	0	0	0	0	0	0	0
MP	Marktplaats.nl	0	0	0	0	0	0	0	0	0	0
MP	Blogger.com	0	0	0	0	0	0	0	0	0	0
MP	Flickr	0	0	0	0	0	0	0	0	0	0
MP	Dumpert.nl	0	0	0	0	0	0	0	0	0	0
MP	Ebay.com	0	0	0	0	0	0	0	0	0	0
MP	Web-log.nl	0	0	0	0	0	0	0	0	0	0
MP	Photobucket	0	0	0	0	0	0	0	0	0	0
MP	Wordpress.com	0	0	0	0	0	0	0	0	0	0
MP	Vimeo.com	0	0	0	0	0	0	0	0	0	0
MP	Icanhascheezburger.com	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
MG	IMDB.com	0	0	0	0	0	0	0	0	0	0
MG	Zoover	0	0	0	0	0	0	0	0	0	0
MG	Boardgamegeek	0	0	0	0	0	0	0	0	0	0
MG	Digg.com	0	0	0	0	0	0	0	0	0	0
MG	Last.fm	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
SN	Myspace.com	0	0	0	0	0	0	0	0	0	0
SN	Hyves.nl	0	0	0	0	0	0	0	0	0	0
SN	Facebook	0	0	0	0	0	0	0	0	0	0
	Average Usage	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

1 =Strategic Focus

0 = No strategic focus

Table 17: Cross matrix 'websites - strategic information quality criteria (Process-Pragmatics)'

11 Appendix B: Personal Reflection

The impact of the research on my personal behaviour and professional acting was tremendous. Where I used to have a lot of different topics to focus on, during my Master Thesis, I could focus on a single subject, create one enormous thesis, instead of small parts of a thesis. I had to decide everything myself, although I enjoyed a great amount help from a lot of people, on reflecting on the subject. For which I must once again express my gratitude.

11.1 Personal issues

I did get to know myself better during the research. That I could easily get lost in the information that was provided, while I should be watching the quality of the information.

Also, while writing my research, doing everything from writing to reflecting and talking to other people, I stumbled upon a most lovely girlfriend, which impacted my personal life even further. As it is difficult to tell which part impacted me in which way, it is needed to mention the fact in this reflection.

11.1.1 Becoming a good person

When I started studying, I had several goals. The first goal was not graduating, not making it in time, not having a good time. It was 'becoming a good person'. And at the time I probably had no idea what the depth of that philosophy was. For people do not become good people by chance. It requires a constant ethical reflection on what you do, on the results and on the motivation why you do it. And now, at the end of my studies at the university, I have to reflect. Did I indeed become a good person?

I am not really sure. I think so, at least that I have become an acceptable person; someone who has high moral standards, so that he is accepted by society. And perhaps I did venture even beyond that, developed skills and habits which are more than society could expect from me. Yet, I do not think the development towards a good person ends here, it is an endless journey. 'Acceptable' is a certain minimum level, 'good' is a direction, not a point at which you can arrive. The word 'love' needs a mention in this context, in the meaning of 'agape' and 'caritas', and complete love for the world and its inhabitants is not something to be achieved in this lifetime.

On the other hand, I think I did the good things, the right and honourable things at times, in a way that I wouldn't have done when I started studying.

11.2 Ethical issues

The ethical issues involved in this research are limited. But they are limited within a very broad spectrum; truth finding. While concerning yourself with information quality, you also get to the point where you can see that information quality criteria obstruct each other. Everybody agrees that you can lose some information quality on the criterion of completeness, to gain a lot on the criterion of conciseness. However, can you lose a bit on the criterion of accuracy, to gain a lot on the criterion of believability? Not all criteria of information quality are created equal, nor should they be.

11.2.1 Financial issues

Somewhere during the research I was notified that there was a certain subsidy which I could apply for. Although the subsidy was not exactly intended for the use for master students who are writing there thesis, we could probably get some funding out of it. The ethical issue was here: Should I take the money, get the additional motivation out of it, and become a bit wealthier? However, I decided against it, since it was not intended for this purpose. This is of course a pity, since I definitely could use the money. But I simply cannot enjoy money that I shouldn't have. It is against my believe that you have to excel, you have to be exalted in your ethical issues. Otherwise, you will just be one of the grey mass, not one of the shining stars, which stand out in a dark and broken world.

11.2.2 Free information for everyone.

The philosophy behind Wikipedia is 'Free information for everyone'. It is more a philosophical principle than that it is a slogan or working business principle. It is a belief, and I must say that I started to believe in the communist approach to information and information objects. Since the first copy of an information object is expensive, but each copy afterwards is nearly free, I started to believe that it should be done by the government and not by information companies. Copyrights are exit, free information for everyone. As such I believe that the creation of information for the masses should be nationalized, that means that enjoyable information, like music, video and the like should be done by the government, at least a part of it. Then we can get rid of copyrights and legal issues, and take another step towards free information for everyone.

12 Appendix C: Acknowledgements

People are not alone in this world, and without help, they achieve nothing. Their lives are empty and their days are filled with endless void. Thus I want to thank a lot of people for their essential contribution to this thesis. Above all, the Giver of Life, who makes the sun rise every day, over good and bad, over rich and poor. Without life, none of this would have been possible.

First of all, my parents, for making this study possible and helping me, financially, morally and for their efforts to make me study.
Roland, for his numerous hours and expertise, as well as presenting this interesting research topic. Markus for his input and interest in the research, although he was not my official supervisor, his input was highly appreciated an of course Maya as my second supervisor for her input, especially on the area of the methodology. The staff of IS&CM for their input in the field.
The people who took their time and talked with me about their experiences in the field, they are all users with rich experiences which could give me the answers needed to observe patterns and behaviour expected by websites. The Wikipedians, for their most interesting behaviour.

And a lot of other people to thank. Thank you!

13 Appendix D: Questionnaire about behaviour of websites.

In this questionnaire you are asked to assess what type of behaviour is displayed by some websites. There are 4 categories of behaviour (as presented in 12.1, 12.2. 12.3. and 12.4) which a website may or may not display. Please indicate:
A: what behaviour is dominant with a D.
B: what behaviour is present, but not dominant with a P.
C: what behaviour is not present with an X.

13.1 Collaborative Content Creation

Collaborative Content Creation behaviour is the creation of information objects, cooperatively, by two or more persons with possibly different knowledge and skills and supplying these information objects to users.

13.2 Media Provision

Media Provision behaviour is the provision of information objects (like photos, videos and articles), which are created by a single user, to other users.

13.3 Metadata Generation

Metadata Generation behaviour is the collecting of tags, ratings and opinions about information objects or real-world objects and providing this information to other users.

13.4 Social Networking

Social Networking behaviour is the communication among users and the information about users and their respective relations.

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket				
	Boardgamegeek				
	Last.fm				
	Uncyclopaedia				
	Dumpert.nl				
	Marktplaats.nl				
	Facebook				
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com				
	Youtube.com				
	Zoover				
	Ebay.com				
	Hyves.nl				
	Digg.com				
	Icanhascheezburger.com				
	wikihow.com				
	Wikipedia				
	Web-log.nl				
	Myspace.com				

D for Dominant

P for Present, but not dominant

X for not Present

13.5 Results questionnaire about behaviour of websites.

These tables contain the answers to the questionnaire, as described in Appendix D.

13.6 Researcher

	Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
Vimeo.com	0	D	1	1
Blogger.com	0	D	1	1
Photobucket	0	D	1	1
Boardgamegeek	0	1	D	1
Last.fm	1	1	D	1
Uncyclopaedia	D	1	1	1
Dumpert.nl	0	D	1	1
Marktplaats.nl	0	D	1	1
Facebook	0	1	1	D
wikia.com	D	1	1	1
Flickr	0	D	1	1
Wordpress.com	0	D	1	1
IMDB.com	1	1	D	1
Youtube	0	D	1	1
Zoover	0	1	D	1
Ebay.com	0	D	1	1
Hyves.nl	0	1	1	D
Digg.com	0	0	D	1
Icanhascheezburger.com	0	D	1	1
wikihow.com	D	1	1	1
Wikipedia	D	1	1	1
Web-log.nl	0	D	1	1
Myspace.com	0	1	1	D

D for Dominant

P for Present, but not dominant

X for not Present

13.7 Respondent 1

	Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
Vimeo.com				
Blogger.com				
Photobucket				
Boardgamegeek				
Last.fm			D	
Uncyclopaedia				
Dumpert.nl				
Marktplaats.nl	0	D	1	0
Facebook				
wikia.com				
Flickr				
Wordpress.com				
IMDB.com				
Youtube				
Zoover				
Ebay.com				
Hyves.nl	0	1	1	D
Digg.com				
Icanhascheezburger.com				
wikihow.com				
Wikipedia	D	1	1	1
Web-log.nl				
Myspace.com				

D for Dominant

P for Present, but not dominant

X for not Present

13.8 Respondent 2

	Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
Vimeo.com				
Blogger.com				
Photobucket				
Boardgamegeek				
Last.fm	0	D	D	0
Uncyclopaedia				
Dumpert.nl	0	D	1	0
Marktplaats.nl	0	0	1	0
Facebook				
wikia.com				
Flickr	1	D	1	0
Wordpress.com				
IMDB.com	1	1	D	0
Youtube	0	1	1	0
Zoover				
Ebay.com				
Hyves.nl				
Digg.com				
Icanhascheezburger.com				
wikihow.com				
Wikipedia	D	D	1	1
Web-log.nl				
Myspace.com				

D for Dominant

P for Present, but not dominant

X for not Present

13.9 Respondent 3

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket	0	D	0	0
	Boardgamegeek	1	0	D	0
	Last.fm	1	1	D	1
	Uncyclopaedia				
	Dumpert.nl				
	Marktplaats.nl	0	1	0	0
	Facebook				
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com				
	Youtube	0	D	1	1
	Zoover				
	Ebay.com	0	1	0	0
	Hyves.nl	0	1	0	D
	Digg.com				
	Icanhascheezburger.com				
	wikihow.com				
	Wikipedia	D	0	0	0
	Web-log.nl				
	Myspace.com	0	1	0	D

D for Dominant

P for Present, but not dominant

X for not Present

13.10 Respondent 4

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket				
	Boardgamegeek				
	Last.fm	1	0	D	1
	Uncyclopaedia	1	0	0	0
	Dumpert.nl				
	Marktplaats.nl	0	1	0	0
	Facebook	D	D	1	D
	wikia.com				
	Flickr				
	Wordpress.com	0	D	D	0
	IMDB.com	1	0	D	0
	Youtube	0	D	D	0
	Zoover				
	Ebay.com	0	D	1	0
	Hyves.nl	1	1	1	D
	Digg.com				
	Icanhascheezburger.com				
	wikihow.com				
	Wikipedia	D	0	0	0
	Web-log.nl				
	Myspace.com				

D for Dominant

P for Present, but not dominant

X for not Present

13.11 Respondent 5

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket				
	Boardgamegeek	0	1	D	1
	Last.fm	0	1	D	1
	Uncyclopaedia	D	1	0	0
	Dumpert.nl				
	Marktplaats.nl	0	D	0	0
	Facebook				
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com	0	0	D	1
	Youtube	0	D	1	1
	Zoover				
	Ebay.com	0	D	1	0
	Hyves.nl	0	1	0	D
	Digg.com	0	0	D	0
	Icanhascheezburger.com				
	wikihow.com				
	Wikipedia	D	1	0	0
	Web-log.nl				
	Myspace.com	0	1	0	D

D for Dominant

P for Present, but not dominant

X for not Present

13.12 Respondent 6

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket	1	D	1	1
	Boardgamegeek				
	Last.fm			D	
	Uncyclopaedia				
	Dumpert.nl				
	Marktplaats.nl	0	D	D	D
	Facebook				
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com	1	D	1	1
	Youtube	D	D	D	1
	Zoover				
	Ebay.com	1	D	D	D
	Hyves.nl	1	D	D	D
	Digg.com				
	Icanhascheezburger.com	D	D	1	0
	wikihow.com				
	Wikipedia	D	D	1	0
	Web-log.nl				
	Myspace.com	0	D	1	D

D for Dominant

P for Present, but not dominant

X for not Present

13.13 Respondent 7

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com	1	D	1	1
	Photobucket	0	D	0	1
	Boardgamegeek	D	1	D	1
	Last.fm				
	Uncyclopaedia				
	Dumpert.nl	0	D	1	1
	Marktplaats.nl	0	1	1	1
	Facebook	0	D	1	D
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com	D	1	D	1
	Youtube	1	D	1	1
	Zoover				
	Ebay.com	0	D	1	1
	Hyves.nl	0	1	1	D
	Digg.com				
	Icanhascheezburger.com				
	wikihow.com				
	Wikipedia	D	0	0	1
	Web-log.nl				
	Myspace.com				

D for Dominant

P for Present, but not dominant

X for not Present

13.14 Respondent 8

		Collaborative Content Creation	Media Provision	Metadata Generation	Social Networking
	Vimeo.com				
	Blogger.com				
	Photobucket	0	D	1	1
	Boardgamegeek	0	0	D	1
	Last.fm				
	Uncyclopaedia				
	Dumpert.nl	0	D	1	1
	Marktplaats.nl	0	0	D	0
	Facebook	0	1	1	D
	wikia.com				
	Flickr				
	Wordpress.com				
	IMDB.com	0	0	D	0
	Youtube	0	D	1	1
	Zoover				
	Ebay.com	0	0	D	0
	Hyves.nl	0	1	1	D
	Digg.com				
	Icanhascheezburger.com	0	D	1	1
	wikihow.com				
	Wikipedia	D	1	0	0
	Web-log.nl				
	Myspace.com	0	1	1	D

D for Dominant

P for Present, but not dominant

X for not Present

14 Appendix E: Possible future patterns

During the course of the research, numerous observations were made, which could be elaborated into patterns. For different reasons these patterns were omitted in the pattern language. These (proto-)patterns are briefly described below and can be the guideline for further research.

14.1 404-error

The 404-error is perhaps the most basic implementation of the *Declaration of Failure* . It declares a failure on Availability, Latency or Response Time. The error in itself is not that interesting, but the possibilities to redirect or request for improvement are interesting areas of improvement. It was omitted in the research because it is not specific Web 2.0, but can be applied in all internet environments.

14.2 Are you sure you're human?

This pattern is the collection of all verification methods that are in use to verify that a user is indeed human. It is an anti-vandalism pattern, which is now a sub-pattern of *Trusted Contributors*. The reason to make this pattern independent is because it targets a specific problem: Vandalism by bots. It was omitted in the research because it could be embedded in a more general pattern.

14.3 House style

A house style enhances the structural consistency in an information repository. It was omitted in the research because it is not specific Web 2.0, but can be applied in all internet environments.

14.4 Human error prevention

Human errors form a source of frustration and may damage the experience a user has with a website. With techniques to prevent human errors, such as suggesting search terms when a user has made a spelling error, the experience of the website can be improved. It increases ease of operation of a website. It was omitted in the research because it is not specific Web 2.0, but can be applied in all internet environments.

14.5 Light version

Light Version addresses the need for a better response time. It has links with AJAX-pattern Fat Client. A website which has a Light Version can display a link to a light version in the loading screen. It becomes increasingly important with the rise of mobile technologies. It was omitted in the research because it is not specific Web 2.0, but can be applied in all internet environments.

14.6 Limited Options

Limited Options can improve Structural and Semantic Consistency in Social Networking websites. It is the counterpart of an anti-pattern, Too Much Freedom, which destroys value by letting everybody do what they want, and although the individual information object might benefit from it, the overall repository is damaged because of reduced consistency. Limited Options reduces the freedom of an individual information object to improve the value of the information repository.

14.7 Logo

The Logo is present on all websites, and it has an impact on information quality. If a logo of the information provider is present on the information object, it transfers the Reputation and Reliability of the information repository to the information object, thus improving the information quality. It was omitted in the research because it is not specific Web 2.0, but can be applied in all internet environments.

14.8 Make it Fun!

This pattern was one of the most promising patterns, which aimed at collecting meta-information about information objects, by making the task of providing meta-information fun in itself. Google Image Labeler was using a game to collect tags about pictures. It was omitted from the research because no more than a single implementation could be found.

14.9 Place for Meta-Discussion

A Place for Meta-Discussion is often seen among Web 2.0 sites, although it is present at many other sites as well. It is unclear what it does at the moment, because it should remove the meta-discussion from the main information object (in Collaborative Content Creation environments), but research at Wikipedia also indicates that the Meta-Discussion cannibalizes the effort done on the main information objects. It is incorporated as an implementation of the *Splitter*-pattern, but should be an independent pattern perhaps. When a website has *A Place for Meta-Discussion* is not regarded as an active implementation of the *Splitter* pattern in the research.

14.10 This is repulsive

The pattern 'This is repulsive' is a typical anti-vandalism pattern. It lets users tag content submitted by other users as repulsive, thus alerting the administrators of a website that the content is repulsive and perhaps even illegal. It is found among all Web 2.0 sites, but removed from the research. That is because it is essentially an implementation of the *Tag Engine* and the *Declaration of Failure* , but on the aspect of legality and good taste, but not on a criterion of information quality.

14.11 Wikignome

In Collaborative Content Creation settings is this one of the most valuable *Trusted Contributors*, who adds a lot of knowledge of the processes to the information object. It is incorporated in the *Trusted Contributors*-pattern, since it has to implement. Wikignomish behaviour is emergent behaviour, which cannot be implemented as such.

14.12 Wisdom of Crowds

A high-level pattern, this pattern assumes that by gathering more feedback and ratings, the overall information quality will improve. It is the core principle behind all Web 2.0 sites which can boast on high traffic ratings. It was omitted in the research because it was neigh impossible to implement.