

Evaluating Information Retrieval Systems for Children in an Educational Context

Suzanna Wentzel
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
s.d.wentzel@student.utwente.nl

ABSTRACT

Technology is increasingly being used in classrooms. Information retrieval systems (IRS) are one example of such technologies that children often use, to search for information. How do they or their teachers know which search engine they should use? Little knowledge exists about when and which searching systems used by children are good, since existing research mainly focuses on other groups of users than children. Therefore, this paper aims at creating a way of evaluating information retrieval systems for children. That is achieved by performing a literature research and complementing that with the results of a questionnaire. The final result is a metric for evaluating an IRS, which contains aspects that can either be true, false or unclear. Future research is needed to get a more detailed metric; one example of a direction for future work is creating detailed evaluations per aspect. Achieving a detailed evaluation of IRS can help teachers and parents guide their children to use the right search engines.

Keywords

Children, interaction, information retrieval systems, search engines, searching, learning, evaluation

1. INTRODUCTION

We are living in a time where digital technology is used more and more. This is also the case in classrooms [44]. Where 30 years ago children wrote an essay with the use of information booklets, children can now use search systems to retrieve information from the internet on topics for their assignments.

The result is that they have easier access to information, and more information is available. This also means that they can get overwhelmed easier. Now the questions arise: what is the best way to guide children in finding relevant information, which search engines should they use? Most people use Google nowadays [2], and children are a big part of that group. But is this the best engine they can use, or is another search engine more suitable? This research aims at defining a metric for evaluating information retrieval systems, such that those questions can be answered.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

31th Twente Student Conference on IT July 5th, 2019, Enschede, The Netherlands.

Copyright 2019, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

2. BACKGROUND INFORMATION

To understand the research questions, one needs to understand what ‘Information Retrieval Systems’ are. The term information retrieval (IR) is quite broad. It could mean looking up information in a book, but also retrieving information from a database or using a search engine to search for information. In the academic field, IR is defined as an aim to select relevant material (of mostly unstructured nature) from large collections of information (usually stored on computers) in response to user queries [34, 35].

Different types of information retrieval systems (IRS) exist, for instance personal information retrieval (information retrieval on a personal device, within consumer operating systems), domain-specific search and web search [35]. The type of IRS that this research will focus on is web search systems: retrieving information from the World Wide Web (search engines for example).

This research focuses on children at primary school (in group 7/8) who use IRS to gather information for their education. This specific target group is chosen because they start to learn how to write essays, in which they need to use IRS. In this context, one needs to understand how the IRS is used and how it fits into the process of children’s search behaviour. Jochmann et al. describe this process using a model which can be seen in Figure 1 [28].

The search starts with an information need of a child: in this example a child of nine years old has the assignment to write an essay about Ireland and wants to know more about the traditional food in Ireland. The child conceptualises this task and enters a query (“food Ireland”) in the IRS and follows certain searching strategies. The IRS matches the query and ranks the results, of which the child needs to decide and understand whether the results are useful. In this example it would be of use to know what makes an IRS good for children: is an IRS that shows all cafes and restaurants that serve local food in Ireland better than an IRS that gives only a short description of the local food?

3. RESEARCH QUESTIONS

RQ How can be evaluated how ‘good’ a certain information retrieval system is in comparison to other information retrieval systems, for children in group 7/8 (age 9-12) of primary school who have to write an essay?

RQ1 Which dimensions define an information retrieval system as ‘good’ for children in group 7/8?

RQ2 In what way can the dimensions in RQ1 be combined into a set of metrics for evaluating an information retrieval system?

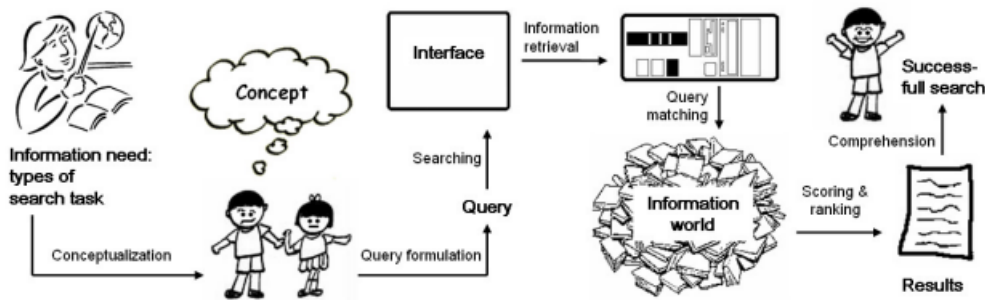


Figure 1. Model of a children's information retrieval process [28]

4. RELATED WORK

Currently, only a few researches focus on evaluating IRS for children, specifically IRS used for finding educational information is still in its infancy. More research on evaluating IRS is done with the focus on other contexts and specific aspects like relevance [5]. There has been research that analysed IRS for children on specific aspects, which can be included in this research. Furthermore, research which discusses the (challenges in) design of search systems can be interesting and useful. These works are described in the following subsections.

4.1 Analysis of IRS for children

In a recent paper from Huibers et al. it is explained how Wizenoz¹ evaluates their search system [23], based on the definition of relevance proposed by Cooper: on logical relevance and utility [13]. Furthermore, they state a definition of good: "In our opinion an educational information retrieval system for children can be described as good when it returns information that is readable, relevant and reliable for the child". That is a great starting point, but except for the relevance and utility, these statements are not (yet) substantiated with research. This research looks into these and more aspects.

4.2 Design of search systems

More research was done on the design of search systems for children. The first example is a work of Gossen on search engines for children. This work had the goal of designing appropriate search engines for children with a focus on the search user interface [18]. In this work, first the specifics of information retrieval for young users were analysed. Then, open issues and challenges were identified in user studies with children, using log file analysis and eye tracking, and using theories of human development. With this information, user interfaces were designed that address these conceptual challenges. Finally, these interfaces were evaluated by children. Chapter 3 of the work (Existing Algorithms and User Interface Concepts for Children) names a couple of dimensions that could influence how an IRS is perceived as good.

Another work, a survey [20] which describes an overview of the achievements in child-specific IRS, has some overlap with the previous mentioned study, but also contains a more specific human studies section. It summarises cognitive studies and its relevance to child-specific IRS.

These works on the design of search systems describe aspects that should be taken into account when designing search systems, but they do not mention any way of evaluating the search systems. The evaluation however could

be done on those same aspects it was designed on, since they take the limitations and abilities of children into account. This research looks into those aspects and ways to combine them, and contributes in this way to the evaluation of information retrieval systems.

5. METHODOLOGY AND APPROACH

To answer the central research question, different approaches and research methods were used per sub-question.

5.1 Gathering a list of dimensions (RQ1)

Two methods were used to answer RQ1: Literature was searched for dimensions on good in information retrieval systems for children. In addition to that, an online questionnaire was sent to teachers for their expert opinion on what they think is good for children in information retrieval systems, to see whether that matches what was found in literature.

5.1.1 Literature research

A literature research was performed to find dimensions that influence how an information retrieval system is perceived as 'good'. The resulting list of dimensions was written down.

5.1.2 Questionnaire

A questionnaire was created to get input from Dutch teachers of the 7th and 8th grade in primary school on what they think is important and good in search engines. The questionnaire was distributed via a Facebook group called "Bovenbouwwereld (gr5-8)" which consists only of primary school teachers in the 5th - 8th grade. In addition to that, it was distributed via acquaintances.

The questionnaire consisted of background questions and the questions which search engines they would recommend and why, and what they think is important in search engines and why. A control question was added to filter for nonsense answers.

5.1.3 Data analysis

The results of the questionnaires were analysed by determining the composition of the group respondents. Answers in why a teacher would recommend a search engine were written down in a list. Furthermore, a list of what is important in search engines used by pupils and why was written down.

The results of the questionnaire and literature research were combined and checked for overlap, and resulted in a list with unique dimensions, which is the answer to RQ1.

¹www.wizenoz.com

5.2 Creating the metrics

To answer RQ2, the results of RQ1 were necessary. This result was used to find at least one combination of dimensions, such that a metric was created. This metric was then tested on completeness (the metric considers all relevant aspects of the IRS) and soundness (the metric is usable in the real world) by using the metric to perform an estimated evaluation of multiple existing IRS.

6. RESULTS

By performing the methods as described in section 5, a resulting list of dimensions on good was found and a metrics was created.

6.1 Dimensions on good (RQ1)

6.1.1 Findings literature

A lot of aspects that make an IRS good (for children) could be found in literature. To categorise the aspects, dimensions were created. These dimensions were either found in literature or created by combining similar aspects. A detailed list of dimensions and their aspects (found in literature) can be found in Appendix A. From these dimensions, it was found that an information retrieval system for children can be described as good when:

[Relevant] *It returns relevant information.*

When a user has an information need, some information is relevant and some is irrelevant [13]. Therefore, to suit a child's information need, the system should return relevant information.

Showing the relevance of results means that results should be ranked [14], relevant information should be contained in the result [13], relevance cues are shown [18], the credibility/reliability of results is presented [3, 13] and results are visualised [21].

[Not irrelevant] *It does not return irrelevant information.*

Brown et al. have shown that children have difficulties with ignoring irrelevant features [9]. Examples of irrelevant information are advertisements and cluttering.

Not returning irrelevant information means that no advertisements are shown [36, 41] and no cluttering is present [9, 27].

[Understandable] *It shows understandable results.*

Gossen stated that search systems for young users should be child-appropriate, and thus not too complex [18]. Tan et al. and Bilal substantiated that by showing that personalisation of results by reading level can increase relevance and retrieval performance in search engines [38] and having misleading titles leads to complex navigational decisions [6].

Therefore, showing understandable results means that results should be readable [12, 18, 40, 38] and titles of results should be understandable [6].

[Emotions] *It aligns with the emotions of children.*

Emotions relate to the user experience, as Sluis et al. stated that the affective value is a direct measure of the user experience [40]. In addition to that, positive feelings are stimuli for persistence in using a search engine [7].

Aligning with the emotions of children means that the system has a positive affective value [7, 26, 40], interesting results are shown [40], some sort of fun

factor is present [11, 27] and the system is not slow [7, 10].

[Presentation] *It presents the results in a child-friendly way.*

The presentation of results is very important in order for children to find information. Choosing the right font size for example helps children with reading [37]. Furthermore, summaries [18], pictures [21] and text characteristics [40] help with understanding results.

Presenting results in a child-friendly way means that different media types are available [22, 40], results are clearly separated [18], queried keywords are highlighted in results [18], the font size is appropriate for children [18, 37], summaries are included in the results [18], a result has different text characteristics [40] and pictures are included in results [18, 21].

[Logical steps] *Children can understand/take logical steps.*

Due to memory overload, children can forget previous actions [18]. A back button makes it for instance easier to trace previously retrieved pages [6]. In addition to that, children like to start again at the homepage when starting a new search task [27], so a home button is required. Such logical steps are thus very important.

Therefore, allowing children to take logical steps means that the home page is directly reachable [27] and the system supports that children can go back one step with the browser's back button [6, 18, 27].

[Information need] *It can satisfy their information need.*

As was already shown in Figure 1 by Jochmann et al. [28], an IRS is used to suffice an information need. However, searching could hold children back in satisfying this information need, because they might have little domain knowledge or need to reformulate their queries [18]. Typing and spelling also limits children's abilities to find appropriate resources [6]. Therefore, children need some support in finding information.

Supporting children in satisfying their information need means that the system supports browsing [7, 24, 28, 29], keyword searching [6, 14, 18], faceted navigation [39], aggregated search/usage of verticals [15, 16], a helper function is present [6], children are assisted in formulating queries [15, 18, 43], queries are spell checked [6, 27] and the system supports natural language queries [27, 30, 42].

[Ethical] *It considers children as users in an ethical way.*

The use of children's personal data is a source of ethical and social concern [33]. Algorithms can affect the social upbringing of a child, if information is withheld or uncritically propagated [32]. Furthermore, persuasive design has great costs like the quality of relationships [31] and it is a child's right to be protected from harmful content [3].

Thus, considering children as users in an ethical way means that the privacy of children remains [32], there is no propagation or withholding of results [33, 32], no persuasive design is used in the interface [31] and the content is child-safe [3].

[Adaptable] *It is adaptable for all types of children.*

Children can have different abilities and disabilities and develop their skills while they get older. The IRS should strengthen the abilities and help with the disabilities. An adaptable search user interface increases the satisfaction of interacting with it [18], and collaborative searching takes away the focus on the query construction and allows focusing on the search task [8]. In addition to that, the system should also be adaptable for disabled children, since it is a child's right to have access to information and communication technologies [3, 1].

Being adaptable for all different types of children means that the user interface can be personalised [18], the system has an evolving search user interface [18], the system allows collaborative searching [8] and it is accessible for all children [1, 3].

[Skills] *It takes the (motor) skills of children into account.*

In order to use standard desktop computers, skills in using a mouse and keyboard are necessary. Many children have difficulties in using these devices, because they require high accurateness in movements [18].

Taking the (motor) skills of children into account means that the sizing of clickable elements is appropriate [24], scrolling is not necessary [6, 14, 36], only single point-and-click actions are required [25], it has audio support (reading out loud) [18] and alternative input methods are accepted [14, 18].

6.1.2 Results questionnaire

12 Dutch primary school teachers responded to the questionnaire for teachers. They teach in different areas of the Netherlands and all the schools those teachers work at have devices available where pupils can get access to Internet. The teachers teach at least classes in the 7th and the 8th grade (in the Dutch school system), some teach other classes as well. The respondents have different amounts of experience.

Pupils of 7 out of 12 teachers can choose a search engine themselves. On the question which search engine teachers would recommend to their pupils, 11 teachers answered Google. 3 teachers gave the reason that Google is easy to use. That the teachers use it themselves was also given as a reason 3 times. It was twice mentioned that it has a clear interface, because it shows clear titles and part of the results. Other reasons given were that children need Google in their life, so children need to be taught how to use it; it filters results and does not show weird pages; a lot can be found with it; Chromebooks are used in class and Chromebooks use Google; it shows which keywords were not found in results; children know it, using other search engines could cause confusion; the working of Google is clear and it is the most general search engine. One teacher said they are still searching for a search engine and that they would therefore not recommend any, and they think Google is comprehensible, but not for children.

On the question of what teachers think is important in search engines used by their pupils, some respondents named multiple features. Some respondents did not put the reasoning at the right question (not at "Why do you think that is important?" but at "What do you think is important on search engines that are used by your pupils?"), those reasons were still used. All features and the reasons given by teachers are described next.

Safety Safety was mentioned three times. Teachers think safety is important, because the children are primary school children: they are young and it considers useful information. They should not be confronted with for example porn.

Filters Filters are actually part of the safety and results fitting to age, but was named separately. A teacher thought this is important because they don't want children to see inappropriate content.

No inappropriate results This is also part of safety, but noted separately. It was deemed as important since not everything is suited for children.

Results fitting to age Multiple teachers mentioned this aspect, and deemed it important since it makes the search more effective and the children understand the information better. Again, they should not be confronted with porn.

Clear interface/overview A clear interface eases the search for an information need. One teacher specifically mentioned: "*Children need to learn how to find reliable information, they need to learn to find information using the correct keywords, and need to be able to handle a lot of results that do not always match their information need. How do you figure that out without reading the entire article, and how can you decide information is reliable? A clear overview helps with this.*" (translated from Dutch).

Queried keywords shown in results Teachers found this important because it eases the search for information, since children need to learn to find information using the correct keywords and handle results that do not always match their information need.

Boundaries, not too many results This was deemed important since that would be more effective, and children can understand the information better.

Direct information Having direct access to information (not having to go through advertisements) was mentioned as important because that is fast and clear.

Summary or part of article in result A teacher mentioned that this is important because it helps children learn how to find reliable information and handle results that do not match their information need.

Source clearly visible According to a teacher, children should be able to see if information is reliable, and that thus the source should be clearly visible.

Easy to use One teacher named that the system should be easy to use, since the users in this case are primary school children.

As few advertisements as possible This is important, because it is less distracting and calms the overview.

Immediately showing the right website Immediately showing the right website would be important because if that is the case, no weird websites are shown.

True and important sources at the top Having true and important sources at the top gives pupils a grip in finding information.

Determining value of results Being able to determine the value of results independently of the search engine is thought of as important since it contributes to developing media knowledge.

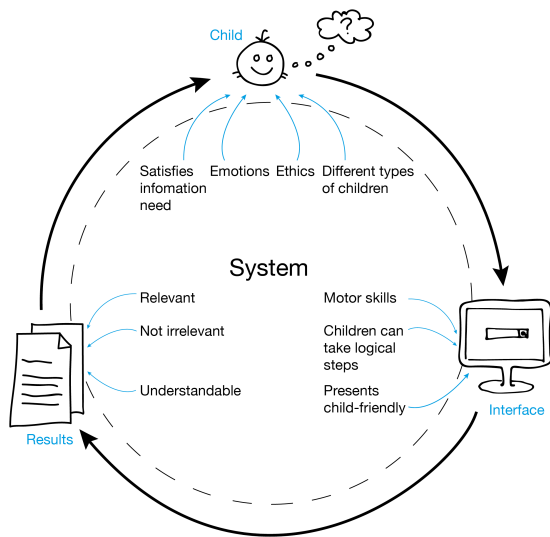


Figure 2. Evaluation paradigm for IRS for children

6.1.3 Merging literature & questionnaire

A lot of dimensions and aspects on the dimensions were found in literature. Respondents to the questionnaire confirmed some of those dimensions and aspects, for instance that it should return relevant information, and that there should be as few advertisements as possible. Also new aspects were mentioned by respondents, like not showing too many results and putting true and important sources at the top.

The findings in literature and the results of the questionnaire were combined in Table 1. When a dimension was mentioned by a teacher, the dimension is denoted by a *. When the column 'T' is marked by an X, a teacher mentioned that aspect in the questionnaire. References behind an aspect mean that the aspect was found in the referenced works, no reference means it was found only in the questionnaire.

Table 1 thus contains which and how dimensions define an IRS for children as good and is therefore the answer to **RQ1** *Which dimensions define an information retrieval system as 'good' for children in group 7/8?*. The paradigm in Figure 2 was created from the dimensions.

6.2 Metrics (RQ2)

All of the dimensions found in section 6.1.3 can be combined into a metric. In fact, Table 1 could already function as a metric, where the aspects are statements that can be tested on whether they hold or not or whether it is unclear if it holds.

This metric was used to estimate an evaluation of the following IRS:

- Google³, because of its wide use (also among children) and the recommendation of the respondents to the questionnaire.
- Web for Classrooms⁴, since it is an IRS which has children as specific target user and focuses on readability of documents.

²General Data Protection Regulation, <https://gdpr-info.eu/>

³www.google.com

⁴app.webforclassrooms.com

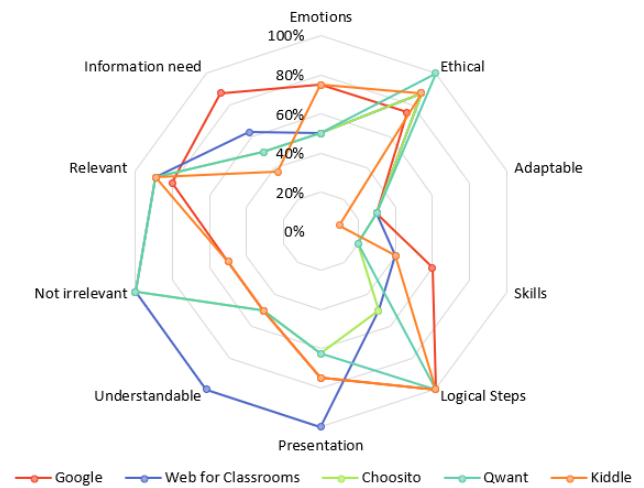


Figure 3. Estimated evaluation of IRS

- Choosito⁵, since it is another IRS which focuses on students and readability.
- Qwant Junior⁶, it is also an IRS focused on children and aims at respecting privacy and providing neutral results.
- Kiddle⁷, it is again an IRS focused on children and aims at respecting privacy.

The detailed evaluation of these search engines can be found in Appendix B. A 'Y' means that the aspect is true, a 'N' means that the aspect is not true and a 'U' means that it is unclear whether the aspect is true. These evaluations are summarised in Figure 3. The percentages of the IRS in each dimension are calculated in the following way:

$$score = \frac{\sum a_{present}}{a_{total}} * 100\%$$

The score per dimension is the percentage of aspects that were present taken from the total amount of aspects for that dimension. This means that all aspects are weighed the same in this first attempt of evaluating IRS with the metric. When it was unclear whether an aspect was present or not, it was taken as a half.

In Figure 3 one can see which search engines would perform best in a dimension. Google would for instance be best at supporting the satisfaction of the information need, while Web for Classrooms would be best at showing understandable results.

Since it was possible to evaluate the IRS using the metric, it can be concluded that the metric is a way of combining the dimensions found in RQ1 for evaluating an IRS. The metric is therefore an answer to **RQ2** *In what way can the dimensions in RQ1 be combined into a set of metrics for evaluating an information retrieval system?*

7. DISCUSSION

In this section, the limitations present while performing this research and the soundness and completeness of the metrics are discussed.

⁵www.choosito.com

⁶www.qwantjunior.com

⁷www.kiddle.co

Table 1. Dimensions that determine how an IRS is perceived as ‘good’ for children in group 7/8

Dimension	Aspects	T
It returns relevant information*	results are ranked [14]	X
	relevant information is contained in the results [13]	X
	relevance cues are shown [18]	
	the credibility/reliability of result is presented (by showing the source) [3, 13]	X
	true and important sources are at the top of the results	X
It does not return irrelevant information	no advertisements are shown [36, 41]	X
	no cluttering is present [9, 27]	
It shows understandable results	results are readable [12, 18, 40, 38]	
	titles of results are understandable [6]	
It aligns with the emotions of children	the system has a positive affective value [7, 26, 40]	
	interesting results are shown [40]	
	some sort of fun factor is present [11, 27]	
	the system does not have a slow response time [7, 10]	
It presents the results in a child-friendly way (in a clear overview*)	different media types are available [22, 40]	
	results are clearly separated [18]	
	queried keywords are highlighted in the results [18]	X
	queried keywords which were not found in results are shown	X
	the font size is appropriate for children (≥ 12 pt) [18, 37]	
	summaries or parts of the article are included in every result [18]	X
	a result has different text characteristics [40]	
	pictures/images/visualisations are included in every result [18, 21]	
It allows children to understand/take logical steps in the system (the operation is clear*)	the homepage is directly reachable [27]	
	the system supports that children can go back one step with the browsers’ back button [6, 18, 27]	
It supports children in satisfying their information need	the system supports browsing [7, 24, 28, 29]	
	the system supports keyword searching [6, 14, 18]	
	the system supports faceted navigation [39]	
	the system supports aggregated search/the usage of verticals [15, 16]	
	a helper function is present [6]	
	children are assisted in formulating queries [15, 18, 43]	
	queries are spellchecked [6, 27]	
	the system supports natural language queries [27, 30, 42]	
It considers children as users in an ethical way	the privacy of children remains (according to the GDPR ²) [32]	
	there is no propagation or withholding of results [33, 32]	
	no persuasive design is used in the interface [31]	
	the content is (made) child-safe (by using filters), such that it does not show inappropriate results [3]	X
It is adaptable for all different types of children	the user interface can be personalised [18]	
	the system has an evolving search user interface [18]	
	the system allows collaborative searching [8]	
	it is accessible to all children [1, 3, 4]	
	it shows results fitting to the child’s age	X
It takes the (motor) skills of children into account, and is therefore easy to use*	the sizing of clickable elements is appropriate (area $\geq 32^2$ pixels) [19, 24]	
	scrolling is not necessary [6, 14, 36]	
	only single point-and-click actions are required [25]	
	the system has audio support (reading out loud) [18]	
	alternative input methods are accepted and supported [14, 18]	

7.1 Limitations

The first limitation occurs in gathering requirements from the experts; the opinions of only 12 teachers in the Netherlands could be used. Opinions of parents, children, or teachers all over the world were not asked but should be included, such that ultimately a metric can be created that is complete and usable in every part of the world.

Another limitation in the methods is that the search engines were now evaluated by one assessor. To get a more representative evaluation, more assessors should evaluate the IRS with the presented metrics, since some of the aspects are subjective (like “*the system has a positive affective value*” and “*interesting results are shown*”).

Furthermore, to evaluate the individual aspects, statements were created that could be answered with either ‘yes’, ‘no’ or ‘unclear’. However, some of the aspects can be implemented in multiple ways. One example is assisting in formulating queries. Fails et al. show that existing query assistance tools do not meet children’s needs, and they give possible improvements [17]. This means that there are good and bad ways to provide such a query assistance tool. It would have been better to have very specific evaluations per aspect for grading each individual aspect, but that was out of scope for this research.

One last limitation is that some dimensions might be more important than other dimensions, but are now weighed the same. Jochmann found for instance that positive, hedonic expressions can be less important for children when evaluating a search interface than the usability [27]. Therefore, figuring out the right weights per dimension or aspect is important, but again out of scope for this research.

7.2 Soundness & Completeness

As can be seen in section 6.2, the metric is applicable to the real world, and is therefore sound (the metric is usable in the real world). Drawing conclusions about the completeness (the metric considers all relevant aspects of the IRS) is harder. Even though a lot of works were read, literature could have been missed. Furthermore, a contradiction was found in the responses of the teachers; they did not agree on the amount of results an IRS should give. This aspect was therefore left out, but might mean that the metric is incomplete.

Furthermore, both literature and respondents mentioned external influences for choosing and designing search engines, like the familiarity of teachers with a search engine, devices used in classrooms and the background of the pupils (native language, experience in using search engines, etc). The context of the search assignment can also have an influence: if a child needs to find an Indian restaurant, a search engine that only returns educational documents would probably not suffice the information need. These factors depend on the user or context entirely and were not included in this research, since this research focused on the information retrieval systems itself. However, since both literature and the respondents mentioned external influences, it is probably important to include them in the metric, which means that the metric is incomplete. Including the context in the findings of this research can therefore be an interesting direction for future work.

8. CONCLUSION

This paper presented a detailed list of aspects divided into dimensions that make an IRS for children in an educational context good. The list was gathered from literature and complemented by the results of a questionnaire for

teachers in the 7th and 8th grade. From this list a metric was created for evaluating IRS.

This metric consists of statements that can be true, false or unclear. It was tested on soundness and completeness by applying it to existing search engines. With this estimated evaluation we can conclude that the metric is a way to evaluate how ‘good’ a certain IRS is in comparison to other IRS, for children in group 7/8 of primary school.

However, the metric can not yet be considered as a complete method for evaluating IRS. For that to happen, further research needs to be done to get detailed tests or evaluations for all aspects, and the evaluation should be done by multiple assessors. Other issues worth investigating are the opinions of teachers worldwide, the opinions of parents and children, and including the context of the search assignment in the evaluation.

9. REFERENCES

- [1] Article 9 - accessibility. <https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities/article-9-accessibility.html>. Accessed: 2019-05-31.
- [2] Search engine market share. <https://netmarketshare.com/search-engine-market-share.aspx>. Accessed: 2019-05-05.
- [3] UNCRC: Article 17. <https://www.cypcs.org.uk/rights/uncrcarticles/article-17>. Accessed: 2019-05-29.
- [4] Web content accessibility guidelines (WCAG) 2.1. <https://www.w3.org/TR/WCAG21/>. Accessed: 2019-06-04.
- [5] R. Ali and M. S. Beg. An overview of web search evaluation methods. *Computers Electrical Engineering*, 37(6):835 – 848, 2011.
- [6] D. Bilal. Children’s use of the yahoooligans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51(7):646–665, 2000.
- [7] D. Bilal and J. Kirby. Differences and similarities in information seeking: children and adults as web users. In *Information Processing Management*, pages 649–670, 2002.
- [8] A. F. Blackwell, M. Stringer, E. F. Toye, and J. A. Rode. Tangible interface for collaborative information retrieval. In *CHI ’04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’04, pages 1473–1476, New York, NY, USA, 2004. ACM.
- [9] A. L. Brown and J. S. DeLoache. Skills, plans, and self-regulation. In *R. S. Siegler (Ed.), Children’s thinking: What develops?*, pages 3–35. Lawrence Erlbaum Associates, Inc, 1978.
- [10] J. Brutlag. Speed matters for Google web search. June 2009.
- [11] J. M. Carroll. Beyond fun. *Interactions*, 11(5):38–40, September 2004.
- [12] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de la Chica, and D. Sontag. Personalizing web search results by reading level. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM ’11, pages 403–412, New York, NY, USA, 2011. ACM.
- [13] W. S. Cooper. A definition of relevance for information retrieval. *Information Storage and*

- Retrieval*, (7):19–37, June 1971.
- [14] A. Druin, E. Foss, L. Hatley, E. Golub, M. L. Guha, J. Fails, and H. Hutchinson. How children search the internet with keyword interfaces. In *Proceedings of the 8th International Conference on Interaction Design and Children*, IDC '09, pages 89–96, New York, NY, USA, 2009. ACM.
- [15] S. Duarte Torres. *Information Retrieval for Children: Search Behavior and Solutions*. PhD thesis, University of Twente, Netherlands, February 2014.
- [16] S. Duarte Torres, I. Weber, and D. Hiemstra. Analysis of search and browsing behavior of young users on the web. *ACM Transactions on the Web (TWEB)*, 8, March 2014.
- [17] J. A. Fails, M. S. Pera, O. Anuyah, C. Kennington, K. L. Wright, and W. Bigirimana. Query formulation assistance for kids: What is available, when to help & what kids want. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, IDC '19, pages 109–120, New York, NY, USA, 2019. ACM.
- [18] T. Gossen. *Search Engines for Children: Search user interfaces and information seeking behaviour*. Springer Vieweg, Magdeburg, Germany, 2015.
- [19] T. Gossen, J. Hempel, and A. Nürnberger. Find it if you can: usability case study of search engines for young users. *Personal and Ubiquitous Computing*, 17(8):1593–1603, Dec 2013.
- [20] T. Gossen and A. Nürnberger. Specifics of information retrieval for young users: A survey. *Information processing and management*, (49):739–756, February 2013.
- [21] M. A. Hearst. *Search User Interfaces*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.
- [22] B. Homer, J. Plass, and L. Blake. The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, 24(3):786–797, May 2008.
- [23] T. W. C. Huibers and T. Westerveld. Relevance and utility in an educational search environment. In *KidRec '19: Workshop in International and Interdisciplinary Perspectives on Children Recommender and Information Retrieval Systems*, New York, USA, June 2019. ACM.
- [24] H. Hutchinson, A. Druin, B. B. Bederson, K. Reuter, A. Rose, and A. C. Weeks. How do i find blue books about dogs? the errors and frustrations of young digital library users. In *Proceedings of the 11th International Conference on Human-Computer Interaction (HCI 2005) (CD-ROM)*. Mahwah, NJ: Lawrence Erlbaum Associates, 2005.
- [25] K. M. Inkpen. Drag-and-drop versus point-and-click mouse interaction styles for children. *ACM Trans. Comput.-Hum. Interact.*, 8(1):1–33, 2001.
- [26] A. M. Isen and J. Reeve. The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29(4):295–323, December 2005.
- [27] H. Jochmann-Mannak. *Websites for children: Search strategies and interface design: Three studies on children's search performance and evaluation*. PhD thesis, Netherlands, January 2014.
- [28] H. Jochmann-Mannak, T. W. C. Huibers, and T. Sanders. Children's information retrieval: beyond examining search strategies and interfaces. In *The 2nd BCS-IRSG Symposium: Future Directions in Information Access*, number WoTUG-31 in eWic Series, pages 64–72, United Kingdom, September 2008. British Computer Society.
- [29] H. Jochmann-Mannak, L. Lentz, T. W. C. Huibers, and T. Sanders. Three types of children's informational web sites: an inventory of design conventions. *Technical communication*, 59(4):302–323, November 2012.
- [30] Y. Kammerer and M. Bohnacker. Children's web search with google: The effectiveness of natural language queries. In *Proceedings of the 11th International Conference on Interaction Design and Children*, IDC '12, pages 184–187, New York, NY, USA, 2012. ACM.
- [31] B. Kidron, A. Evans, and J. Afia. Disrupted childhood: The cost of persuasive design. Technical report, 5Rights, 2018.
- [32] N. Kucirkova, J. Fails, S. Pera, and T. W. C. Huibers. *Children and Search / Recommendations algorithms: What Adults Need to Know*. DigilitEY, 2018.
- [33] N. Kucirkova, J. Fails, S. Pera, and T. W. C. Huibers. *Children and search/recommendation algorithms: what adults need to know*. Brochure, 2018.
- [34] R. R. Larson. *Understanding information retrieval systems*. CRC Press, Boca raton, USA, 2012.
- [35] C. D. Manning, P. Raghavan, and H. Schütze. *An introduction to information retrieval*. Cambridge University Press, Cambridge, England, 2008.
- [36] S. Naidu. Evaluating the usability of educational websites for children. January 2005.
- [37] J. Nielsen. Children's websites: Usability issues in designing for kids. September 2010.
- [38] C. Tan, E. Gabrilovich, and B. Pang. To each his own: Personalized content selection based on text comprehensibility. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 233–242, New York, NY, USA, 2012. ACM.
- [39] K. A. Ullah and M. A. Iftikhar. Information search for children using faceted navigation. In *2013 International Conference on Open Source Systems and Technologies*, pages 28–33, December 2013.
- [40] F. van der Sluis and E. van Dijk. A closer look at children's information retrieval usage: Towards child-centered relevance. pages 3–10, July 2010.
- [41] N. Vanderschantz. *KidsQuestions: Assisting Children's Digital Information Seeking*. PhD thesis, University of Waikato, New Zealand, 2016.
- [42] N. Vanderschantz and A. Hinze. How kids see search: A visual analysis of internet search engines. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, HCI '17, pages 84:1–84:6, Swindon, UK, 2017. BCS Learning & Development Ltd.
- [43] N. Vanderschantz and A. Hinze. A study of children's search query formulation habits. In *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*, HCI '17, pages 7:1–7:4, Swindon, UK, 2017. BCS Learning & Development Ltd.
- [44] V. Vega and M. B. Robb. The common sense census: Inside the 21st-century classroom. Technical report, CA: Common Sense Media, 2019.

APPENDIX

A. DIMENSIONS ON GOOD

Table 2: Detailed dimensions on good found in literature

Dimension	Aspect	Importance
Relevant information	Containing relevant information	When a user has an information need, some information is relevant and some is irrelevant [13]. To suit the child’s information need, the system should return relevant information.
	Ranking of results	Children visit only the first result page and click on the first item of the result list [14].
	Relevance cues	Cues improve the information processing flow of children, they can estimate the relevance of the result in a better way [18].
	Credibility/reliability of information presented	When results are logically relevant, but the user has no faith in the accuracy of the results, that information is still useless for a user [13]. Furthermore, having access to reliable information is a child’s right [3].
	Visualisation of results	Visualisation and presentation of results is important as it affects the searcher’s judgement of the documents’ relevance, the human perceptual system is highly attuned to images, and visual representations can communicate some kinds of information more rapidly than text [21].
Not irrelevant information	No advertisements	Advertisements are very distracting for children, especially popups, and children could end up on a completely different website [36]. Furthermore, advertisements clutter pages, which hinders children in finding information [41].
	No cluttering	Children do not appreciate cluttered pages [27]. Furthermore, younger children have difficulties with ignoring irrelevant features, the less organised the stimulus and the greater the number of distracting elements it contains, the harder it will be for the child to ignore those irrelevant elements [9].
Understandable results	Readability of results	Personalisation of Web results by reading level can increase relevance and retrieval performance by search engines [38, 12]. Furthermore, adjusting the complexity of search results can be especially salient in the beginning of a search [40]. Gossen also stated that search systems for young users should be child-appropriate, and therefore not too complex [18].
	Understandability of titles in results	Having misleading titles in results leads to complex navigational decisions [6]. It is therefore important that titles of hyperlinks are understandable.
Connects with emotions	Affectivity	Positive affect fosters responsible behaviour and effective performance of tasks that need to be done [26]. In addition to that, Sluis et al. stated that the affective value is a direct measure of the user experience [40]. Positive feelings children experience are stimuli for persistence in using an engine [7]
	Interestingness	Interests can differ between groups of people. Especially for children who have less metacognitive skills or motivation, interest is an important part of child-centered relevance [40].
	Fun	Playful design provokes positive emotional expressions [27]. Carroll et al. also stated that fun should be included as a separate usability area: “People must want to use a system, and must continue wanting to use the system. Part of achieving this is making the system fun to use” [11].
	Speed	When a search engine has a slow response time, children can get distressed [7]. Adding to that, users are less engaged with systems when web search latency increases [10].
Presentation results	Availability different media types	Different media types put different demands on users [40]. For example, Homer et al. showed that learners who have a visual learning preference had less cognitive load when learning from video’s [22].
	Separation of results	Gossen et al. found that it was not clear that a list contains multiple results when they are not explicitly separated through UI elements, but only with white-space [18].
	Keyword highlighting	Having no keyword highlighting makes it harder for children to estimate how relevant each result is [18].

	Font size	Large font sizes help children read texts [37]. Furthermore it can help children judge the relevance of results by taking the reading competence of the children into account [18].
	Summary in result	Children are not yet experienced readers. To pay attention to this lower reading competence, the result should contain a short textual summary [18].
	Text characteristics	Text characteristics can help readers in finding and understanding texts [40]. With text characteristics, signalling devices (titles, headings), typography (bold, italic, font use, capitalisation) and structural elements (graphics, (sub)sections, table of contents, indexes) are meant.
	Picture in result	Visual representations can communicate some kinds of information more rapidly than text [21]. Furthermore, children find images important [18].
Logical steps	Reachability home page	When children start a new search task, they almost always go back to the home page to start from the beginning, problems arise when no clear home button is present [27].
	Ability to go one step back	Due to memory overload, children can forget previous actions, like which documents contained relevant information and which queries they already used [18]. A back button provides a linear path of previously retrieved pages and makes it easier for a user to trace their previous steps [6]. Children prefer to use the back button over bookmarking relevant results to return to good information [27].
Satisfies information need	Supporting browsing	Browsing offers structure and removes the need of choosing abstract terms as keywords, since it relies on recognition instead of recall [29, 28]. Furthermore, Hutchinson et al. and Bilal et al. found that children are more successful using browsing than keyword searching [24, 7].
	Supporting keyword searching	Druin et al. and Bilal et al. found that children prefer keyword searching over browsing [14, 6]. Gossen et al. argued that it is good to offer both browsing and keyword searching, because it enables children to search more flexibly and they learn and improve both techniques [18].
	Faceted navigation	Searches are more efficient and effective using faceted (facets are characteristic attributes that divide a domain) navigation [39].
	Aggregated search/use of verticals	Providing rich media from different genres, verticals can improve the Web experience of children [16]. Furthermore, children engage and explore more in aggregated pages, and vertical results were more likely to be clicked on the aggregated pages [15].
	Helper functions	Search instructions, search examples, browsing instructions, browsing examples and a context sensitive help wizard help with children's learning [6].
	Assisting in formulating queries	Searching could hold children back, because of the need of reformulation and little domain knowledge [18]. Assisting children in this reformulation helps children find their information need [43]. Query expansion can be highly beneficial to retrieve (focused) child-friendly content, since their average query length is smallest from all age groups and they have a small query vocabulary size [15]. Query suggestion helps children focus their search and alleviates the problem of finding the right keywords for the query [15].
	Spell checking	Typing and spelling limits children's ability to find appropriate resources [6], but typing and spelling is needed for most search engines to find relevant search results [27].
	Supporting natural language queries	Kammerer et al. found that natural language queries (in comparison to keyword queries) led to more explicit results and thus greater successes. It takes away the difficulties trying to apply keyword search knowledge [30, 27]. Another interesting advantage on using natural language queries found by Vanderschantz et al. is that using natural language queries less advertisements are shown in results [42].

Ethics	Privacy	Systems should comply to the GDPR in the EU and the COPPA ⁸ in the US. The use of children’s personal data is a source of ethical and social concern, especially if it is uncritically adopted for children’s learning, because most current algorithms follow adult design and are very often based on commercial models [32].
	Propagation/withholding results	Algorithms can affect the social upbringing of a child, especially if information is withheld or uncritically propagated [33]. Algorithms should be economically and politically independent [32].
	Persuasive design	Persuasive design strategies are deployed for commercial purposes to keep us online. It has great costs like the quality of relationships and the opportunity cost (loss of creativity, autonomy and memory) [31].
	Child-safe content	Article 17 of the UNCRC states that children and young people should be protected from media that would be harmful to them, which includes: pornography, media that depicts graphic violence and media that promotes irresponsible drug use [3].
Different types of children	Personalisability of search user interface	Personalisation of the SUI increases the satisfaction of children interacting with it. [18].
	Evolving search user interface	An evolving search user interface enables a flexible adaptation of the search user interface to address changing user characteristics in order to support diversity of users [18].
	Allowing collaborative searching	Collaborative searching takes away the focus on the query construction, and focuses on the search task itself [8].
	Accessibility	Article 17 of the UNCRC also states that efforts should be made to make sure everyone has access to media [3]. This includes young people whose freedoms are limited, or young people who may find the media difficult to access, such as some of those with disabilities or for whom English is not their first language. The UN Convention on the Rights of Persons with Disabilities aligns with this and recognises access to information and communications technologies, including the Web, as a basic human right [1].
Supporting motor skills	Sizing of (clickable) elements	Large target sizes (buttons and other widgets) allow children to make selections more quickly while small targets slow them down and can lead to frustration [24].
	Necessity of scrolling	When pages are long, children tend to not scroll down the page [36, 6], which means that information could get lost. Furthermore, when there are a lot of results, it becomes harder to choose one [36]. Druin et al. also found that vertical scrolling is not optimal for children [14].
	Single point-and-click vs drag-and-drop actions	Children prefer point-and-click actions over drag-and-drop actions, and children interact with point-and-click significantly faster with fewer errors than drag-and-drop [25].
	Supporting audio	An audio support (reading out loud) feature is needed by children who still learn to read, but it should be an option, because some children find audio support irritating [18].
	Alternative input methods	Typing and spelling slow children down, auto-complete does not always help because children are not looking at the screen at the right time [14]. Therefore, a search user interface for children should provide different possibilities for children to formulate their information need [18].

⁸Children’s Online Privacy Protection Act, <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>

B. EVALUATION OF IRS

Table 3. Estimated evaluation search engines

Dimension	Aspects	Google	Web for Classrooms	Choosito	Qwant Junior	Kiddle
Relevant	results are ranked [14]	Y	Y	Y	Y	Y
	relevant information is contained in the results [13]	Y	Y	Y	Y	Y
	relevance cues are shown [18]	Y	Y	Y	Y	Y
	the credibility/reliability of result is presented (by showing the source) [3, 13]	Y	Y	Y	Y	Y
	the truth and important sources are at the top of the results	N	U	U	U	U
Not irrelevant	no advertisements are shown [36, 41]	N	Y	Y	Y	N
	no cluttering is present [9, 27]	Y	Y	Y	Y	Y
Understandable	results are readable [12, 18, 40, 38]	N	Y	Y	N	N
	titles of results are understandable [6]	Y	Y	N	Y	Y
Emotions	the system has a positive affective value [7, 26, 40]	N	N	Y	N	Y
	interesting results are shown [40]	Y	Y	Y	Y	Y
	some sort of fun factor is present [11, 27]	Y	N	N	N	N
	the system does not have a slow response time [7, 10]	Y	Y	N	Y	Y
Presentation	different media types are available [22, 40]	Y	Y	Y	Y	Y
	results are clearly separated [18]	N	Y	N	N	N
	queried keywords are highlighted in the results [18]	Y	Y	N	Y	Y
	queried keywords which were not found in results are shown	Y	Y	N	N	N
	the font size is appropriate for children (≥ 12 pt) [18, 37]	Y	Y	Y	Y	Y
	summaries or parts of the article are included in every result [18]	Y	Y	Y	Y	Y
	a result has different text characteristics [40]	Y	Y	Y	Y	Y
	pictures/images/visualisations are included in every result [18, 21]	N	Y	Y	N	Y
Logical steps	the homepage is directly reachable [27]	Y	Y	Y	Y	Y
	the system supports that children can go back one step with the browsers' back button [6, 18, 27]	Y	N	N	Y	Y
Information need	the system supports browsing [7, 24, 28, 29]	N	N	N	N	N
	the system supports keyword searching [6, 14, 18]	Y	Y	Y	Y	Y
	the system supports faceted navigation [39]	Y	N	Y	N	N
	the system supports aggregated search/the usage of verticals [15, 16]	Y	Y	Y	Y	Y
	a helper function is present [6]	Y	Y	N	N	N
	children are assisted in formulating queries [15, 18, 43]	Y	N	Y	N	N
	queries are spellchecked [6, 27]	Y	Y	N	Y	Y
the system supports natural language queries [27, 30, 42]	Y	Y	N	Y	N	
Ethical	the privacy of children remains (according to the GDPR) [32]	U	Y	Y	Y	Y
	there is no propagation or withholding of results [33, 32]	U	U	U	Y	U
	no persuasive design is used in the interface [31]	Y	Y	Y	Y	Y
	the content is (made) child-safe (by using filters), such that it does not show inappropriate results [3]	Y	Y	Y	Y	Y
Adaptable	the user interface can be personalised [18]	Y	N	N	Y	N
	the system has an evolving search user interface [18]	N	N	N	N	N
	the system allows collaborative searching [8]	N	N	N	N	N
	it is accessible to all children [1, 3, 4]	U	U	U	U	U
	it shows results fitting to the child's age	N	Y	Y	N	N
Skills	the sizing of clickable elements is appropriate (area $\geq 32^2$ pixels) [19, 24]	N	Y	N	N	Y
	scrolling is not necessary [6, 14, 36]	N	N	N	N	N
	only single point-and-click actions are required [25]	Y	Y	Y	Y	Y
	the system has audio support (reading out loud) [18]	Y	N	N	N	N
	alternative input methods are accepted and supported [14, 18]	Y	N	N	N	N