### UNIVERSITY OF TWENTE.





# On the way to understand the structure of entrepreneurship research

Dagmara Magdalena Fijoł Master Thesis Business Administration

Thesis supervisors:

Dr. Rainer Harms, University of Twente, Enschede, The Netherlands Dr. Jonathan D. Linton, University of Ottawa, Ottawa, Canada

> University of Twente School of Management and Governance Enschede January 31, 2012

### **Management Summary**

Entrepreneurship as a field of research is relatively young, fragmented and dynamically evolving. Researchers with various backgrounds are active in entrepreneurship research, and although they do contribute to the development of the field, lack of a common approach and lack of agreement about the definitions, causes barriers for the scholarly development of entrepreneurship research and makes it difficult to identify key topics in the field. To understand better what are the core topics of today entrepreneurship research, an assessment of the structure is needed.

Previous research on the entrepreneurship itself has established that the field is fragmented, having relevant topics spread across various disciplines, lacking consistency in topics over time (Gartner et al., 2006), and not having convergence (Grégoire et al., 2006).

Previous studies on this subject were based on citation-based analysis. Nevertheless this method is subject to many limitations, including the differential influence exercised by various number of researchers across sub-disciplinary fields (Theoharakis & Hirst, 2002), perfunctory mention citations (Kotler, 1972), not used citations, citations that are used in order to please potential reviewers (Tellis, Chandy, & Ackerman, 1999), to name just a few. To avoid this situation and provide objective overview of the most dominant topics in the field, self-organizing maps (SOM) are used in this research.

Self-organizing maps are used in this study to identify and map the key topics published in the entrepreneurship literature between 2006 and 2010. This method uses terms directly from articles and organizes them according to their frequency of appearance in a given documents, therefore it is entirely objective. Another advantage is that this method gives a better representation of the key terms present in the literature, than co-citation analysis, because it catches all topics regardless of them being cited or not. Furthermore, SOM analysis based on a given year will make it possible to identify core topics per year and make some comparisons between them. This way one can make conclusions about visible trends and changes in them, and make some assumptions about the developments of the entire field.

Initial results indicated that most dominant topics in the filed are following: 'innovation', 'venture capital', 'firm performance', 'competitive advantage', 'product development', 'entrepreneurial activity', 'product innovation', 'economic growth', 'technology transfer', 'entrepreneurial firms', 'venture creation', 'risk taking', 'nascent entrepreneurs', 'social networks', 'entrepreneurial orientation', 'financial performance', 'entrepreneurial opportunities', 'new technology', 'innovation process', social entrepreneurship', and 'corporate entrepreneurship'. These are the most dominant themes across all analyzed years, and even though there are slight frequency fluctuations, these are the key subjects in entrepreneurship research.

Self-organized maps do not only provide the list of most dominant topics, the method also allows to cluster similar data together. Clusters are based on numbers of articles, so that not only the topic can be identified, but also the number of articles in which the topics was published. This means that one can analyze cluster's size and its key subjects.

The sizes and topics of clusters differ per year, but terms like: 'firm', 'form', 'family', 'entrepreneur', 'venture', 'business', 'technology', 'entrepreneurship', 'network', 'social entrepreneurship', 'growth', 'age', 'capital', 'ties', and 'innovation' seem to be stable elements of clusters in every given year.

Clustering of data means that research is focused on selected areas of topics, however in case of an entrepreneurship research the number of clusters in every year is very big (67 clusters in 2010, 56 clusters in 2006 and 2007, 48 clusters in 2009, and 47 clusters in 2010). Even though the number of cluster is slightly decreasing with every year, it does not mean that the research is narrowing – the amount of clusters in every year is still too big, meaning that entrepreneurship research is broad and remains fragmented.

Simpson's diversity index used to measure the degree of similarity in data collected from entrepreneurship research supported the findings reported on the maps. Simpson's index ranges from 0 to 1, where the greater the value, the greater the sample diversity. This index represents the probability that two samples randomly selected from a given dataset will belong to different groups. (Simpson E. H., 1949)

In case of terms present in entrepreneurship literature. the Simpson's index calculated for each year is close to 1, what means that topics published in entrepreneurship literature are really diverse and two randomly selected topics most probably will not belong to the same category. This proves that entrepreneurship research remains very fragmented, broad, and topics do not seem to be connected.

Further, a group analysis was performed for four levels of analysis – *person, team, venture,* and *environment*. Group *Person* included topics like entrepreneur, female entrepreneur, owner, founder, behavior, skills, experience, education, etc. Group *Team* involved topics like team, family, employees, *Venture* included topics like organizational form, performance, growth, production, innovation, strategy, and finally *Environment involved* topics like risk, opportunity, network, ties, clusters, market, etc. Group analysis has shown that the most dominant group of topics is related to the person with the share of 37% in 2006, 52% in 2007, 44% in 2008, 47% in 2009 and 42% in 2010. Even though the interest slightly changes over years, this group of topics does remain the most dominant.

The least dominant are topics related to environment with shares between 8% in 2008, 9% (2009), 11% (2006, 2010) and 12% in 2007. The interest in environment related subjects is constant and remains at the stable level.

Topics related to team and venture receive a comparable level of attention. Interest in venture related topics oscillates between minimal 22% in 2007 and 27% in 2010. The interest slightly grew and it will most likely continue to grow.

Topics related to team has made 26% of all published content in 2006, then it dropped to 14% in 2007, grew again in 2008 to 23%, and from 2009 onwards its declining again. All in all, it seems that in the period between 2006 and 2010 the focus is on the person.

Based on the text and SOM analysis, several conclusions about the entrepreneurship research and its structure arise. Main and most apparent conclusion is that the entrepreneurship research field remains very broad. It is possible to identify several key themes that can be considered as constant and fixed elements for entrepreneurship research, however the field resembles more of a patchwork than a solid, homogenous structure.

Results compared to previous research indicated that key themes identified over 20 years ago are still actual, however not complete any more. The field actually broadened and included new shares of topics. In the study of Morris, Lewis and Sexton (1994) several most dominant themes in the entrepreneurship research were identified. The core topics at the were: creation, innovation, pursuit of opportunities, risk management, uncertainty, pursuit of profit, personal advantages, new production methods, management, coordination of resources and value creation. This research adds some new positions like: finance, social entrepreneurship, environment awareness , organization theory, strategy, and marketing.

All in all, this paper brings valuable results and enhances the body of knowledge. Obtained results have allowed to make some conclusions about the general structure of the field and most dominant themes in the entrepreneurship research. Based on the data it is not possible to make any predictions about the future trends, but it is very likely that the field will be evolving and real-time situations like financial-crisis, politics, and national regulations will influence the direction of the research. Entrepreneurship as a research field may be young and immature, but nobody can claim anymore that it is illegitimate or not independent.

### Preface

This paper reports the study on the structure of entrepreneurship research field. However, to me this paper is much more. I do not only report the research that I have performed – with this paper I graduate and close an important part.

This paper is an example of an international and virtual cooperation – most of research activities were performed in Japan, Poland and Netherlands, and they were being supervised by dr. R. Harms of University of Twente in the Netherlands and Prof. J. D. Linton of University of Ottawa in Canada.

Here I would like to thank both of my supervisors for their valuable contribution. Dr. Harms gave a direction for this research, read every single word in this report and provided very helpful comments for which I am very grateful. Prof. Linton supervised methodological part of this research, he provided the knowledge necessary to understand and operate the method and made this project possible. Here I would like to thank Prof. Linton for all his help.

I am very thankful to Prof. M. J. Embrechts of Rensselaer Polytechnic Institute, Troy, New York. He not only let me use the software that he has created, he also provided personal support for any issues that I have faced during this research. And although it would be nice to say that the method was simple and there were no issues, it's not true. There were many situations where I desperately needed help from Prof. M. J. Embrechts. This method was new to me and I did not always know if I used correct procedures, or how I should read the results. Prof. M. Embrechts would answer all my questions, even if they were the same over and over. I am very grateful and can only hope that I was not too annoying with my questions.

Last by not least, I would like to thank my husband, Wojtek, who for last months constantly listened to me saying that I can't conduct this research, that this can't be done. Here is some good news for you – it's done! I would like to thank you for all your support and comments. You know better than anyone that without your help this paper would be something completely different.

I hope that readers of this paper will enjoy the reading and find something valuable or inspiring for themselves. I certainly did!

Hengelo, 30<sup>th</sup> January 201

# Table of Contents

Management Summary I			
Preface		IV	
1. Intro	oduction	. 1	
1.1.	Problem indication	. 2	
1.2.	Research questions	. 3	
1.2.2	2. Research questions	. 3	
1.2.3	3. Approach to answer the research question	. 3	
1.3.	Significance of the research	4	
2. Liter	rature review	. 5	
2.1.	Defining entrepreneurship	. 5	
2.2.	Entrepreneurship as the research domain	. 8	
2.2.2	1. Why is it important to research entrepreneurship?	. 8	
2.2.2	2. Entrepreneurship research – domain demarcation	11	
2.3.	Developments of the field	12	
2.3.1	1. Four phases of entrepreneurship research	13	
2.4.	Current state of the entrepreneurship research	15	
3. Text	mining	16	
3.1.	Self-Organizing Maps	16	
3.1.1	1. Introduction	17	
3.1.2	2. Architecture	17	
3.1.3	3. Main principle	19	
3.1.4	4. Main algorithm	19	
3.2.	SOM application potential	19	
3.3.	Entrepreneuship and text analysis	20	
3.4.	Entrepreneurship by SOM	21	
4. Met	hodology	22	
4.1.	The software	22	
4.2.	Data collection	22	
4.3.	The procedure	23	
4.3.2	1. Generating a dictionary	23	
4.3.2	2. Text alignment	23	
4.4.	Data analysis	25	
4.5.	How to read the map	25	
5. Resu	ılts	30	
5.1.	Most dominant topics published in 2006	30	
5.2.	Most dominant topics published in 2007	34	

5.3.	Most dominant topics published in 2008	. 38
5.4.	Most dominant topics published in 2009	. 42
5.5.	Most dominant topics published in 2010	. 46
5.6.	Articles vs. clusters	. 50
5.7.	Convergence by Simpson's diversity index	. 51
5.8.	How did the key topics changed over time	. 52
5.9.	Group analysis	. 55
Cono Refe	clusions and discussion	. 57 . 61
Арре	endix	. 65
I. II. V.	Most dominant topics found in the selected entrepreneurship literature in 2006 Most dominant topics found in the selected entrepreneurship literature in 2007 Most dominant topics found in the selected entrepreneurship literature in 2008 Most dominant topics found in the selected entrepreneurship literature in 2009	. 66 . 67 . 68 . 69
	5.3. 5.4. 5.5. 5.6. 5.7. 6.8. 5.9. Cond Refe Appe 1. 1. 1. V.	<ul> <li>Most dominant topics published in 2008</li></ul>

### List of tables

Table 1. Classical views on entrepreneurship	6
Table 2. Diversity index for every analyzed year	51
Table 3. Changes in key entrepreneurship topics	54

# List of figures

Figure 1. Possible delineations of the entrepreneurship phenomenon	8
Figure 2. Introductory framework	9
Figure 3. Hypothesis testing model	10
Figure 4. Fragment of self-organizing map of business and management journals	17
Figure 5. Schematic view of a feedforward single-layer artificial neural network	18
Figure 6. 2-dimensional map of neurons	18
Figure 7. Dictionary creating process	23
Figure 8. Alignment procedure	24
Figure 9. Key topics in the clusters in 2006	33
Figure 10. Key topics in the clusters in 2007	37
Figure 11. Key topics in the clusters in 2008	41
Figure 12. Key topics in the clusters in 2009	45
Figure 13. Key topics in the clusters in 2010	49
Figure 14. Number of articles in the clusters	50
Figure 15. Changes in topics related to innovation	52
Figure 16. Changes in topics related to finance and growth	52
Figure 17. Changes in topics related to entrepreneurship	53
Figure 18. Shares of particular topics groups within the total number of publications for 2006	56
Figure 19. Shares of particular topics groups within the total number of publications for 2007	56
Figure 20. Shares of particular topics groups within the total number of publications for 2008	56
Figure 21. Shares of particular topics groups within the total number of publications for 2009	56
Figure 22. Shares of particular topics groups within the total number of publications for 2010	56

# List of Maps:

<ul> <li>Map 2. Most dominant topics published in 2006</li> <li>Map 3. Clusters identified in the entrepreneurship literature in 2006</li> <li>Map 4. Kohonen map for most dominant topics published in 2007</li> <li>Map 5. Most dominant topics published in 2007</li> <li>Map 6. Clusters identified in the entrepreneurship literature in 2007</li> </ul>	. 30
<ul> <li>Map 3. Clusters identified in the entrepreneurship literature in 2006</li> <li>Map 4. Kohonen map for most dominant topics published in 2007</li> <li>Map 5. Most dominant topics published in 2007</li> <li>Map 6. Clusters identified in the entrepreneurship literature in 2007</li> </ul>	31
<ul> <li>Map 4. Kohonen map for most dominant topics published in 2007</li> <li>Map 5. Most dominant topics published in 2007</li> <li>Map 6. Clusters identified in the entrepreneurship literature in 2007</li> </ul>	. 32
Map 5.Most dominant topics published in 2007Map 6.Clusters identified in the entrepreneurship literature in 2007	34
Map 6. Clusters identified in the entrepreneurship literature in 2007	35
	. 36
Map 7. Kohonen map for most dominant topics published in 2008	. 38
Map 8. Most dominant topics published in 2008	39
Map 9. Clusters identified in the entrepreneurship literature in 2008	. 40
Map 10. Kohonen map for most dominant topics published in 2009	. 42
Map 11. Most dominant topics published in 2009	43
Map 12. Clusters identified in the entrepreneurship literature in 2009	44
Map 13. Kohonen map for most dominant topics published in 2010	46
Map 14. Most dominant topics published in 2010	47
Map 15. Clusters identified in the entrepreneurship literature in 2010	48

### 1. Introduction

Entrepreneurship as a field of research is relatively young, fragmented and dynamically evolving. Researchers with various backgrounds are active in entrepreneurship research, and although they do contribute to the development of the field, lack of a common approach and lack of agreement about the definitions, causes barriers for the scholarly development of entrepreneurship research and makes it difficult to identify key topics in the field. To understand better what are the core topics of today entrepreneurship research, an assessment of the structure is needed.

Previous research on the entrepreneurship itself has established that the field is fragmented, having relevant topics spread across various disciplines, lacking consistency in topics over time (Gartner et al., 2006), and not having convergence (Grégoire et al., 2006).

Previous studies on this subject were based on citation-based analysis. Nevertheless this method is subject to many limitations, including the differential influence exercised by various number of researchers across sub-disciplinary fields (Theoharakis & Hirst, 2002), perfunctory mention citations (Kotler, 1972), not used citations, citations that are used in order to please potential reviewers (Tellis, Chandy, & Ackerman, 1999), bias in favor of popular authors (May, 1967), or those who write reviews (Woodward & Hensman, 1976), and finally methodological articles and authors in established fields with many researchers tend to be cited more often (Margolis, 1967). As a result of that the citation and co-citations analysis can potentially miss out the important, but not yet popular topic. Therefore, the text-mining method used in this study is fully automated and objective, thus it addresses most of the above stated concerns.

In this research self-organizing maps (SOM) are used to identify and map the key topics published in the entrepreneurship literature between 2006 and 2010. This method uses terms directly from articles and organizes them according to their frequency of appearance in a given documents, therefore it is entirely objective. Another advantage is that this method gives a better representation of the key terms present in the literature, than co-citation analysis, because it catches all topics regardless of them being cited or not. Furthermore, SOM analysis based on a given year will make it possible to identify core topics per year and make some comparisons between them. This way one can make conclusions about visible trends and changes in them, and make some assumptions about the developments of the entire field.

This study is to analyze 1671 articles from 9 entrepreneurship journals published between 2006 and 2010. Results presented in this paper give a good overview of the dominant topics in the field and enhances academic understating of the development of the entrepreneurship field.

### **1.1.** Problem indication

The first known entrepreneurship course was taught at Harvard University by Myles Mace in 1947, and since then academic interest in entrepreneurship has grown to include more than 2200 courses offered at 1600 colleges and universities, 44 English-language refereed academic journals, 100+ entrepreneurship centers, 277 endowed positions, and over 1200 members in the Entrepreneurship Division of the Academy of Management. It seems that there is a lot of activity in this field, but little intellectual cohesion among these efforts. (Brush et al., 2003)

Brush et al (2003) argue that lack of cohesion in entrepreneurship field happens for several reasons. Main reason is that there is a small number of universities and colleges that offer Ph.D. programs in entrepreneurship, and therefore the majority of scholars who are active in the entrepreneurship field have various disciplinary backgrounds. Because of that entrepreneurship research is considered to be interdisciplinary and consisting out of several subfields (Linton, Himel, & Embrechts, 2009). In the past several studies were conducted on developments of the field, core authors and literature (Cornelius, Landstrom, & Persson, 2006; Gregoire, Noel, Dery, & Bechard, 2006) - however very little research has been done on the structure of this field itself. Lack of understanding of the structure has some functional consequences for scholars active in entrepreneurship research, as well as researchers or students interested in it.

Most apparent difficulty regards defining entrepreneurship research domain. At this moment research domain is not defined, entrepreneurship research is very broad and indistinctive, many of entrepreneurship topics fit into already existing disciplines, but are often marginal to them, thus are not covered completely. Entrepreneurship knowledge is spread across various disciplines and there is a clear lack of completeness and structure.

Entrepreneurship research remains pretty much fragmented because of the continuous debates about the definition of the field and the absence of the coherent and widely accepted conceptual framework. (Zahra, S. A. & Wright, M.; 2011) The fragmentation causes barriers that make it difficult to see the entrepreneurship as the complete and mature field of science.

Finally, entrepreneurship is a fast developing field of research with several scientific journals especially dedicated. Nevertheless, these journals are categorized in the field of business, management or innovation what makes it difficult to judge which of them are publishing core papers on entrepreneurship and influence scholars the most. Another issue is the huge and continuously growing amount of published material available to researchers. In 1996 there were 140,000 titles available only in the field of management<sup>1</sup> (Sandelands, 1996). Nowadays it is even more complicated to navigate between journals. The most difficult part for authors is deciding which journal is the most relevant and suitable for their research interests. Researchers should be engaged in research that will enhance the body of knowledge and will help to develop the field further, but choosing which journal is the most relevant and appropriate for researcher active in entrepreneurship research has never been more difficult.

<sup>&</sup>lt;sup>1</sup> Entrepreneurship journals are most often to be found in the 'management' category

To summarize, problems defined for this research can be listed as follows:

- 1) Entrepreneurship is not defined,
- 2) Research domain is not defined,
- 3) Entrepreneurship research is very broad and indistinctive,
- 4) Many of entrepreneurship topics fit into already existing disciplines, but are often marginal to them, thus are not covered completely,
- 5) Entrepreneurship knowledge is spread across various disciplines,
- 6) There is clear lack of completeness and structure.

In order to address these problems a map of entrepreneurship journals will be developed to present the structure of entrepreneurship research. The mapping method will help identifying the core topics of entrepreneurship research, trends and emerging topics, as well as it will allow to detect changes in topics that happened over time. It will help to discover boundaries of entrepreneurship research and understand its structure. Next to that, it will allow the illustration of relationship of chosen journals to each other, but also will identify the core topics covered by the journals and that eventually will help researchers to judge journals' suitability and appropriateness (Linton, et al., 2009).

#### **1.2.** Research questions

This paragraph contains the research questions that are addressed in this Master Thesis.

#### 1.2.1. Main Research Objective

Overall objective of this research is to learn about the structure of entrepreneurship research, to find out what are the core topics existent in the current entrepreneurship research, and to understand how these topics changed over time, and finally to predict future trends in entrepreneurship research domain. The results of this research will not only put some light on current state of entrepreneurship research, but also will present typically published content in entrepreneurship literature and will help understanding the relation between selected journals. The objective will be reached by means of literature review and text mining of which final result will be incorporated into Kohonen self-organizing map and later analyzed and discussed with relation to previous research.

#### **1.2.2.** Research questions

Above mentioned problems lead to a central question: <u>What is the structure of entrepreneurship</u> <u>research</u>? Some additional questions that arise as well are following:

- 1) Which (sub)fields are existent within entrepreneurship research literature?
- 2) What are the core topics covered by entrepreneurship research literature?

#### **1.2.3.** Approach to answer the research question

All questions are answered by means of literature study and analysis of content published between year 2006 and 2010 in selected journals. Main method used in this research is text mining. Text

mining, often referred to as data mining, is a process of automated retrieval of novel, interesting and high quality information from one single document or from many documents. Retrieved information is organized, incorporated into self-organized map, and then analyzed. The results are reported in Chapter 5.

### **1.3.** Significance of the research

This research aims at understanding the structure of entrepreneurship research. The results can be very useful for entrepreneurship researchers and practitioners in various ways. First of all entrepreneurship is an evolving field of research that still lacks common agreement regarding the definition of the phenomenon and boundaries of the domain. This research is based on data retrieved from entrepreneurship literature, and although it will not provide any definitions that entrepreneurship lacks at this moment, it will put some light on core topics and interests of researchers active in this field. Thus, it can help understanding what are the boundaries of entrepreneurship research domain, what are the emerging topics relevant to this field, and how the research domain evolved over time. Additionally, this method can give insight into multidisciplinary nature of the field and give a representation on how close some subfields are. It can also help to understand which journals can act as bridges connecting these different subfields. Furthermore, new technique used in this research will allow to identify the typically published content in selected entrepreneurship journals, thus results of this research offer new researchers, students as well as junior faculty, knowledge which journals are and which are not in fit with their research interests. Finally, Self Organizing Maps are a useful tool for analyzing large amount of information without having the need of finding individuals with the right knowledge and risking the bias. It is also quite useful technique for considering bodies of knowledge where experts might not be available, and therefore it is important analytical tool for both librarians and information scientists (Linton et al, 2009).

### 2. Literature review

This chapter will present some of the theoretical views on entrepreneurship as a phenomenon and field of research. Although there is many literature available on the entrepreneurship itself, not so much research was done on entrepreneurship as a research domain. Nevertheless, some available theories and perspectives are presented below.

### 2.1. Defining entrepreneurship

This chapter needs to start with defining the phenomenon of entrepreneurship. Entrepreneurship is a complex matter and if we want to understand its underlying story a clear idea of what entrepreneurship is has to be developed. Entrepreneurship is very complex and it is difficult to capture it in one simple definition. Scholars of many fields have tried to grasp the definition of entrepreneurship, however differences between approaches and opinions of economists, psychologists and management specialists make it impossible to agree on one common definition.

For instance, economists tend to focus more on the cost and production functions, risk-bearing abilities of entrepreneurs and differences in perceptions of the risks associated with entrepreneurship. Psychologists also tackle issues related to entrepreneurial risk, but the generally focus more on individual differences of entrepreneurs, such as self-efficacy, need for achievement, locus of control, tolerance and ambiguity. (Cardon, Wincent, Singh, & Drnovsek, 2009; McMullen & Shepherd, 2006a; Rindova, Barry, & Ketchen, 2009; Westhead, Wright, & McElwee, 2011)

While studying available literature on the topic, it is clear that there are a lot of ideas what entrepreneurship is, and there is still no common agreement on the definition. Davidsson (2004) provides an overview of several views on what entrepreneurship consists of:

- New entry (Lumpkin & Dess, 1996)
- The creation of new enterprise (Low & MacMillan, 1998)
- The creation of new organizations (Gartner, 1988)
- A purposeful activity to initiate, maintain and aggrandize a profit-oriented business (Cole, 1949)
- Taking advantage of opportunity by novel combinations of resources in ways which have impact on the market (Wiklund, 1998)
- The process by which individuals either on their own or inside organizations pursue opportunities without regard to the resources the currently control (Stevenson & Jarillo, 1990)
- The process of creating something different with value by devoting the necessary time and effort; assuming the accompanying financial, psychological, and social risks, and receiving the resulting rewards of monetary and personal satisfaction (Hisrisch & Peters, 1989)

Not only phenomenon of entrepreneurship is difficult to capture. Also the role of entrepreneur became a subject of research and here are several views developed as well. In the paper of C. van Praag (1999): "Some classic views on entrepreneurship", six views on the entrepreneurship are presented and discussed. Van Praag (1999) focused mainly on entrepreneur's position within the economic system, position within the firm, entrepreneurial task, entrepreneurial personality,

entrepreneurial drive, and supply and demand on the market. Her findings are presented in the Table 1.

The six views presented by van Praag (1999) exhibit some similarities and differences. The biggest differences are to be found in the perceived role of an entrepreneur within the economic system. Cantillion had seen entrepreneur as an independent owner and trader, while for Shumpeter and Kirzner an entrepreneur equals innovator. But of course 'innovator' is not everything that an entrepreneur is – entrepreneur is also an employer, arbitrageur, coordinator, organizer and gap-filler, leader, speculator, manager, source of information, and more (Kirzner, 1983).

	Richard Cantillon	Jean-Baptiste Say	Alfred Marshall	Joseph Schumpeter	Frank Knight	Israel Kirzner
1. Position within the economic system	Central (arbitragers): Responsible for exchange and circulation	Central coordinator in production and distribution	Coordinate both supply and demand	Innovator	Bearing the uncertainty and making decisions in economic progress	Discover and exploit profit opportunities
2. Position within the firm	Independent owner	Coordinator / manager	Overall control: both manager and employer	Leader and innovator, not necessarily owner or director	Direction and control when uncertainty is evolved	Innovator
3. Definition of entrepreneurial task	Balancing supply and demand: economic equilibrating	Add value to raw materials: production of wealth	Seek opportunities, develop, produce and sell	innovate and lead and make a profit out of it	Bearing the real uncertainty	Perceive profit opportunities in an earlier stage than others
4. Entrepreneurial personality and ability needed	Dearing and well educated	Many (moral) qualities and most important: knowledge	Intelligence and specialized abilities to perform all tasks and be a natural leader of men	Seek, find and implement new innovations	Extended entrepreneurial abilities, good luck and good fortune	No special abilities but the knowledge to find knowledge
5. Returns and entrepreneurial drive	Non-contractually arranged (possible) profit/income	Profit of the added value, sell product for more than bought	Direct (high) private benefits and esteem	Position in class / society and scarce motivating forces	Residual income of the added value (profit) but also prestige and job satisfaction	Earn a profit on early discoveries
6. Supply and demand in the market	Natural balance: supply and demand rule the amount of entrepreneurs	Limited to several entrepreneurs	Survival of the fittest due to the high prices	Temporary: no lasting position of that person	Balanced	Be ahead of competition in supply

 Table 1. Classical views on entrepreneurship.

After researching entrepreneurial role, researchers have become interested in specifics and main topics of the field itself. Gartner (1990) studied entrepreneurship literature and found the following eight themes to emerge in entrepreneurship research:

- The entrepreneur himself
- Innovation
- Organization creation
- Creating value
- Profit or non-profit
- Growth
- Uniqueness
- The owner-manager

Similar results were found by Morris, Lewis and Sexton (1994) who analyzed the content of journal articles and books. The researchers have found following common keywords:

- Start, form, create
- New business
- Innovation, new product, new market
- Pursuit of opportunities
- Risk taking, risk management, uncertainty
- Pursuit of profit, personal advantages,
- New production methods
- Management
- Coordination of resources
- Value creation

The above presented lists are evidence not only of the fact that entrepreneurship is an interesting subject of research, but also that it gives researchers some difficulties in finding one, common definition. Davidsson suggests (2004) that different entrepreneurship definitions actually address two relatively distinct social realities. First one refers to the phenomenon of self-employment, where some people rather than working for somebody else under an employment contract, prefer to work on their own and become self-employed, or team owner-managers of an independent business. This implies a different risk/reward structure with a much wider span of possible financial outcomes, and a more fluid boarder between work and leisure. Quite often these new business formations involve some element of innovation which is often necessary for a firm to survive. However, as Davidsson (2004) point out, it is well known fact that most of independent firms are relatively stable operation and low to medium value-added industries. Some of the business will grow in size providing its owners with new challenges regarding the different types management or transfer the ownership. In this social reality the term 'entrepreneurship' involves topics like 'self-employment', 'small business management', 'stages of development models', and 'family business'. In other words: anything that concerns independently owned firms and their owner-managers constitutes entrepreneurship.

The second social reality that appears as a major underlying theme in entrepreneurship definitions is that the development and renewal of any society, economy or organization requires micro-level actors who have the initiative and persistence to make change happen. (Davidsson, 2004) Davidsson (2004) argues that institutions as well as market and organizational structures may facilitate or hinder change and development, but these structures do not create any change or change themselves – it is humans who make the change. Thus in the end it is the unique knowledge, perceptions and goals of individuals that drive to take an action and create novelty. In order for these new initiatives to have an impact, individuals need to create the value or save resources. Thus in this approach the term 'entrepreneurship' involves topics like start-up, innovation, corporate venturing, organizational rejuvenation, and change agency. (Davidsson, 2004; Schumpeter, 1934; Sharma & Chrisman, 1999)

A problem with many definitions and views on entrepreneurship is that they are in fact a mixture of two alternatives described above. Figure 1. below presents the possible delineations of the entrepreneurship phenomenon:

A. Micro-level novel initiative



B. Independent

Figure 1. Possible delineations of the entrepreneurship phenomenon (Davidsson, 2004)

The two realities described above refer to or independent business or micro-level novel initiatives, however focusing on one reality only may not be enough since entrepreneurship is connoted with a broad array of things that are not necessarily belong to independent business or micro-level novel initiative only. Thus, as Davidsson (2004) suggests, choosing the intersection of quadrate A and B seems like a more complete alternative since it combines the both approaches. Therefore Davidsson (2004) proposes a 'micro-level' novel initiative' view, but only with connection with a market context, thus excluding non-market activities such as non-profit enterprises, and internal and organizational change. Only activities undertaken by independent business that introduce novelty to a market are included in Davidsson's definition of entrepreneurship. Davidsson favors the Kirzner's (1973) view on entrepreneurship that sees entrepreneurship as competitive behavior that drives market processes, and although it may seem that other definitions are unreasonably excluded, it is not the case. In Davidsson's view not every business owner is considered an entrepreneur – only individual that introduces a novel idea (innovation) and stimulates the market to change is a true entrepreneur.

It has to be reminded that no one can claim to have the right definition of the entrepreneurship. Here different approaches and definitions are presented only to demonstrate that entrepreneurship is a complex matter and researchers still search for one true answer what entrepreneurship really is. Nevertheless, lack of common definitions creates some practical problems for researchers active in entrepreneurship research field. Main problem is that entrepreneurship research has vague and very broad boundaries, thus it is very difficult for researchers to come up with common entrepreneurial theories that are needed before practical implications for entrepreneurs can even be made.

### 2.2. Entrepreneurship as the research domain

In this section entrepreneurship as a research domain is discussed. First it has to be explained why it is important to study entrepreneurship, and then demarcations of the domain can be presented.

#### 2.2.1. Why is it important to research entrepreneurship?

Entrepreneurship was not always seen as an action of transforming the innovations into economic

goods – it was actually seen as robbery: 'one man's pain is another man's gain' (Van Praag, 1999). Around the 18<sup>th</sup> century theories about entrepreneurship started to be developed and since then the entrepreneurship have been a subject of a constant interest among researchers. Why did researchers become so interested in entrepreneurship? The answer to this question seems to be very straightforward – entrepreneurship has an effect on economy. Entrepreneurship can trigger economical growth by creating more employment or by implementing new technologies that help increasing productivity and that is why the phenomenon of entrepreneurship deserves every attention that it gets.

Nevertheless linking entrepreneurship to economical growth is not so easy, because not every form of entrepreneurship can trigger growth. The influence on economical growth depends heavily on the form of entrepreneurship. If every start-up business is regarded as an entrepreneurial activity, even the agricultural start-up, then most probably entrepreneurship will not lead to economic growth (Acs, 2007). In the traditional analyses of economic development the main focus was on big corporations whereas small enterprises that bring innovations and stimulate competition were often neglected (Acs et al, 2004). Large corporations indeed affect the growth in economy, mainly due establishing new plants that create more employment or innovations that help increasing productivity. This growth does not have to be necessarily caused by domestic corporations; also international organizations can influence the economy growth in the country where they operate (Acs and Armington, 2004). But can individual entrepreneurship lead to economical growth?

In principle, an individual's decision whether or not to engage in entrepreneurial activity is influenced by several external factors like national policies, and internal factors like skills and experiences of a prospective entrepreneur. These two conditions together affect the entrepreneurial process, and if they are combined effectively they lead to innovation processes and increased competition on the market, which in turn lead to economic development (Acs, 2007). This view is also supported by Carree et al. (2005) who proposed a framework that links various intermediate variables and linkages, and explains how entrepreneurship influences economic growth. According to Carree et al. (2005) the examples of these intermediate links are: innovation, variety of supply, competition, specific efforts and energy of entrepreneurs, etc. There are also some conditions for entrepreneurship, for instance personal traits, cultural background, education and experiences. In order to have a better overview of the concept, Carree et al (2005) propose a following framework:



Figure 2. Introductory framework. (Source: Carree et al. 2005)

Carre et al (2005) point out that entrepreneurship is the activity that takes place on the individual level, while economic growth is relevant at level of firms, regions, industries and nations. Thus linking entrepreneurship to economic development is in fact linking individual level to collective levels. This framework clearly shows that entrepreneurship is dependent on personal, cultural and institutional conditions and only successful combination of them can lead to innovative ideas and increased competition (Carree et al. 2005). There is no doubt that innovation is an important factor contributing to the economic growth, and that innovation is also associated with entrepreneurship.

One of the first theories of economic development was Shumpeter's original theory that described entrepreneur as an innovator - the main driver of economic growth. The positive relationship between innovation and economic development is well established, mainly due to the fact that innovation is relatively easy to measure. Innovation is often measured by number of patents or expenditures on research and development. Theories about economic growth most of the times are based on neoclassical traditions where development is driven by changes in capital and labor inputs or technological change (Solow, 1956; Romer, 1990). However, these studies did not provide any evidence or explanations what is the link between entrepreneurship and economic growth. This link was established by Davidsson (2003) who connected entrepreneurship and innovation. Other researchers have supported his thesis by explaining that entrepreneurship can contribute to economic growth through introducing innovations, bringing the change, creating competition and enhancing rivalry (Wennekers et al, 1999; Carree et al, 2003).

These findings are also supported by research of Wong et al (2005) who tried to measure the impact of entrepreneurship, defined as a new business creation, on economic growth. To measure this impact Wong et al. have used the TEA (Total Entrepreneurship Activity) rates from the Global Entrepreneurship Monitor program. There are three types of TEA rates: opportunity, necessity and high growth potential. Wong has used following model to test his hypotheses:



Figure 3. Hypothesis testing model. (Source: Wong et al, 2005)

The results of the research of Wong et. al. indicate that innovation positively affects GDP growth. There is no evidence to suggest that higher levels of overall TEA can be associated with higher GDP growth rates, only high potential TEA is found to have a positive effect on economic growth. What is important to note is that high potential TEA characterizes ventures that are growing, innovative, international and ventures that have market creation impact. Thus these results indicate that existence of firms does not have an impact on the economic growth, but only entrepreneurship in conjunction with innovation can contribute to the growth. Although there is some critique on this study, especially regarding the GEM methodology, the findings of Wong et al (2005) are consistent with earlier studies' results that treat innovation as the key to economic growth (Karlsson et al. 2004).

Concluding, even though not every form of entrepreneurship has an effect on economical growth it is still worth to study entrepreneurial processes in order to better understand the process of creating new venture and entrepreneurial activities and motivations involved.

#### 2.2.2. Entrepreneurship research – domain demarcation

As it was mentioned in the previous section entrepreneurship research has very broad boundaries. Davidsson (2003) points after Gartner (2001) major topics that are included in entrepreneurship research, and these are the following: new venture ideas and strategies, ecological influences on venture creation, the acquisition and management of venture capital and venture teams, self employment, the owner-manager, management succession, corporate venturing and the relationship between entrepreneurship and economic growth. When looking at these topics it is very easy to get an impression of 'all-inclusive' domain.

In the past there were some attempts to organize and delineate entrepreneurship research domain. Shane and Venkataraman (2000) have seen the need not only to define the entrepreneurship phenomenon, but also to define precisely the scholarly domain. The authors have suggested that entrepreneurship research field should be focused on examination on how, by whom and with what effects opportunities to create future goods and services are discovered, evaluated and exploited. This of course involves the examination of sources of opportunities, the processes of discovery, evaluation and exploitation of opportunities, and individuals who discover, evaluate and exploit them (Venkataraman, 2000; Davidsson, 2003).

Thus, three sets of research questions are particularly interesting here: 1) why, when and how entrepreneurial opportunities come into existence, 2) why, when and how some people discover and exploit these opportunities, and 3) why, when and how are different modes of action used to exploit these opportunities (Davidsson, 2003).

The trouble with this approach towards defining entrepreneurship research domain is that general primacy is given to the individual and 'opportunity', that is not defined very precisely. Other approach was presented by Gartner (1988), who suggested that entrepreneurship research should focus on the behaviors of entrepreneurs during the process of creating new venture. However, this approach is also criticized as the one that is too narrow, since it does not include the discovery processes (Davidsson, 2003).

Davidsson (2003) suggests somehow more complete approach towards defining the entrepreneurship domain and its subfields. His ideas are more like propositions and thoughts in ongoing debate, rather than ready solutions, but author proposes a complete approach, thus it is worth mentioning here.

Davidsson (2003) in his proposition of domain delineation makes use of the ideas of Gartner (1988) and Shane & Venkataraman (2000). First author states that entrepreneurship research should study the behavior in the process of emergence, but it should delineate between behavior, process and emergence. Further, based on ideas of Shane & Venkataraman (2000) author suggests that entrepreneurship research should also distinguish between two sub-processes: discovery and

exploitation. Moreover, Davidsson (2003) argues that entrepreneurship should not only study the emergence of new independent organization, but also emergence of new market offerings through different modes of exploitation. And finally, Davidsson (2003) points out that entrepreneurship research should study the variety of outcomes on different levels and should adopt the assumption that economy is characterized by heterogeneity and uncertainty.

Thus, what Davidsson (2003) offers as entrepreneurship research demarcation is the following:

- 1) Uncertainty and Heterogeneity
- 2) Processes of emergence; behaviors in the processes of discovery and exploitation
- 3) Creating new business ventures and its outcomes on different level analysis

Concluding, even though the goal of entrepreneurship research is to understand the phenomenon of entrepreneurship, it is often limited to success cases of creation of new ventures. To make the story more complete Davidsson (2003) proposes following domain delineation:

"Starting from assumption of uncertainty and heterogeneity, the domain of entrepreneurship research encompasses the study of processes of (real or induced, and completed as well as terminated) emergence of new business ventures, across organizational context. This entails the study of the origin and characteristics of venture ideas as well as their contextual fit; of behaviors in the interrelated processes of discovery and exploitation of such ideas, and of how the ideas and behaviors link to different types of direct and indirect antecedents and outcomes on different level of analysis."

Although this approach is complete, it makes the research domain quite broad and less distinctive. Furthermore, many of the above questions would fit within some of existing disciplines or subdisciplines, but none of these existent disciplines would cover entrepreneurship topics completely (Davidsson, 2003). Thus, lack of common agreement on definitions of entrepreneurship phenomenon and research domain, together with the fact that most of the entrepreneurship topics fit with other disciplines (but often are marginal to them) is actually the real and most important problem of entrepreneurship researches community today.

### 2.3. Developments of the field

'The field of entrepreneurship, as a disciplinary research field, is not more than 25 years old, but has developed substantially over that short period of time' (Cornelius et al, 2006). However its maturity is often questioned. Cornelius et al (2006) believe that entrepreneurship have moved from management subdiscipline of management studies to a separate field, although is not yet fully mature. Nevertheless, it shows all the signs of maturing field, from its increasing internal orientation and establishment of key areas to advanced and theoretical approach with specific-terms (Cornelius, et al., 2006)

According to Cornelius et al. (2006) there has been a debate on how new fields are created and how they develop. Authors state after Hansson (1993) that there are two approaches to knowledge creation, the 'technical' and the 'theoretical'. Hansson argued that young field of research are characterized by technical approach, where researchers are focused on the object of study (as

opposed to theories and methodologies) to find the knowledge that can be applied in a practical situation. Much of the research is empirical, but due to a lack of conceptual platform, the knowledge is fragmented. According to Hansson mature fields have, on the other hand, strong theoretical approach to knowledge and practical application is less important. The aim of mature disciplines is to understand, explain and sometimes even predict, rather than simply describe (Cornelius, et al., 2006).

Cornelius et al. (2006) explains that entrepreneurship research has applied a technical approach to knowledge for many years, describing the phenomenon of entrepreneurship and making use of empirical approach. The research focus shifted in time from identifying psychological characteristics of entrepreneurs, toward an assessment of cognitive and behavioral aspects of entrepreneur with increased focus on context and on the entrepreneurial process (Cornelius, et al., 2006). The quality of studies has improved, and new theoretical perspectives were introduced, thus the conclusion of Cornelius et al. is that the current state of entrepreneurship research is somewhere between technical or applied, and theoretical or mature approach to knowledge development (Cornelius, et al., 2006; Davidsson, Low, & Wright, 2001a, 2001b)

According to Cornelius et al. (2006) maturing disciplines have following characteristics:

- An increasing internal orientation with researchers citing the work of other entrepreneurship researchers
- A stabilization of topics within the field
- An identifiable research community that is lead by core researchers that highly influenced the development of the field

Both Cornelius et al. (2006) and Gregoire et al. (2006) in their research are trying to illustrate the developments of entrepreneurship as a research discipline. In Cornelius et al. study bibliometric data from the Social Science Citation Index is used in order to determine the developments of entrepreneurship research. The approach was not unique and was previously used in several other studies (Reader & Watkins, 2001), however the approach of Cornelius et al. is more revolutionary. Authors have taken a snapshot of the key researchers in the field in the different periods of time and then conducted a co-citation analysis of refereed academic articles that included the word "entrep\*" in the title, key words, or abstract between 1986 and 2005. Similar approach was applied in the research of Gregiore et al, (2006). Authors also conducted a co-citation analysis of co-citation networks that emerge from the 20,184 references listed in the 960 full-length articles published in the Babson College Entrepreneurship Research Conference's *Frontiers of Entrepreneurship Research* (FER) proceedings between 1981 and 2004 (Gregoire, et al., 2006).

The results of these two studies provide a detailed analysis of given time periods in which core topics and authors are identified. There are four time periods, thus four phases of entrepreneurship research that put some light on developments of the field.

#### 2.3.1. Four phases of entrepreneurship research

#### 1981-1986: A Focus on the Person

Results of Gregoire et al. study indicate that in the literature in the period between 1981 and 1983 a focus was on the person of entrepreneur and his/her characteristics. Most work in that period aims at identifying the characteristics of entrepreneurs, and also tend to consider both personality traits

(attitudes, preferences) and non-psychological variables, like education and age. Moreover, the results indicate that researchers considered differences between various categories of entrepreneurs. Other trends were motivation of research and development entrepreneurs, technology innovators, their characteristics and their need for achievement and power. But taken together, all this work is unified in terms of need for establishing the core dimensions of one's personality (Gregoire, et al., 2006).

#### 1987-1992: The Emergence of New Topics

In this period the topics related to person are still popular, but new topics started to emerge as well. Researchers became interested in motivations that could explain why one decides to engage in entrepreneurial activity. Boulton and Carland (1984) argued that the motivation of entrepreneurs can be treated as primary causes for their behaviors. Shapero, Sokol (1982) and Gartner (1985) researched the contextual factors that motivate entrepreneurs. Authors of that period also focused on the personality traits and individual motivators. Some work on venture creation and entrepreneurs' chances for success also appeared (Gregoire, et al., 2006).

Toward the second half of the '80s, entrepreneurship began to move beyond the personal characteristics of entrepreneurs and started to show interest in other conceptual activities. New interest areas oscillate around firm organization, venture performance, influence of industry structure, firm-level strategy, and other aspects of competitive advantage on performance. Entrepreneurship research of the period appears to have converged around tighter groups of key references. Cornelius et al point out they are Aldrich H.E., Portes A., Light I., Waldinger R., just to name a few. These researchers can be considered a foundation or core authors of that period. Other authors who published their work on entrepreneurship were specialized in other disciplines, like management, marketing, etc, what indicates that entrepreneurship as a field of research was not yet mature (Gregoire, et al., 2006).

#### 1993–1998: A Subfield of Strategy

In this period a focus changed toward new-venture-performance, but the absence of co-citation relationships indicate that from the point of view entrepreneurship scholars' citations activities, each of the work constitutes own 'conceptual island'. Co-citation evidence indicates that instead of leading the field to converge upon a cluster of key references, the competitive strategy in entrepreneurship research took place (Gregoire, et al., 2006).

Furthermore, as results of Gregoire at al. study indicate that during this period there were two streams of research focusing on venture capital. The first one centered on the decision models of venture capitalists (represented by co-citation among MacMillan et al., 1985; Tyebjee & Bruno, 1984; MacMillan et al., 1987), and a second one centered on the actual contributions of venture capitalists to the growth and performance of the venture they fund (represented by co-citation among MacMillan et al., 1989; Gorman & Sahlman, 1989; Sapienza, 1992).

Other features of that period are disappearance of references about the personality traits and individual characteristics of entrepreneurs, stable interest in venture capital, increasing dominance of strategy and the resource based perspective, and finally co-citation relationships, although less dense, are still present. Concluding, entrepreneurship field in years 1993-1998 was in state of flux, with weaker points of convergence (Gregoire, et al., 2006).

#### 1999–2004: New Beginnings

In the last studied period there are much more co-citations relationships within groups of text focusing on the similar subject. There are several networks present, but for the first time the most frequently cited work is entrepreneurship specific. In general, when compared to other periods, entrepreneurship research published in the FER during the 1999–2004 centered around topics like "opportunity identification and exploitation, organizational emergence, the relationships between social capital, interorganizational learning and innovation, the implications of entrepreneurship as a firm-level dimension, and the syndication network of venture capitalists financing arrangements" (Gregoire, et al., 2006).

#### 2.4. Current state of the entrepreneurship research

The results of study conducted by Gregoire et al. (2006) and Cornelius et al. (2006) lead to conclusion that over studied period of time (25 years) there has been convergence in the field, however the levels of this convergence remained low. But it is interesting to observe the developments of the field that moved from its infancy toward being more mature. Entrepreneurship as a discipline is self-reflective, what means that there is an increasing amount of research with focus on state of entrepreneurship research, its developments and its future. The number of outsiders who are specialized in other disciplines, but are cited by entrepreneurship researchers has decreased over time. Furthermore, the research community recognizes and identifies with a large number of core authors, who greatly influenced the development of the field. Also increasing complexity of research areas indicates the maturity of the field (Cornelius, et al., 2006).

Nevertheless, entrepreneurship cannot be considered a fully mature field. Relative novelty of the field is evident with the lack of entrepreneurship theories. Entrepreneurship research is focused on many topics, often from other disciplines, like finance, psychology and management, the discipline has broad boundaries and is not distinctive. On the other hand, there is need for entrepreneurship-specific theories and these need to be developed in the future, therefore it can be predicted that this field will be still evolving. In order to monitor this movement and change, it is necessary to state up to date with the entrepreneurship literature.

### 3. Text mining

Main method used in this research is text-mining. This method is automated and fast retrieval of novel and interesting information from one document or set of documents and identification of valuable patterns. The text-mining method is similar to the data-mining method, but text-mining concerns collections of documents, where the contents are readable and the meaning is obvious. However, it doesn't mean that data-mining and text-mining are two distinct concepts. Both are based on samples of past examples. The composition of samples is different (numbers vs. text), but many of the learning methods are similar. And that is because text samples are processed and transformed into a numerical representation (Weiss, 2005).

A primary focus of text-mining methods is classification. The concept of classification can be extended to data where connections are not easily observable. Classification is basically assigning a document to the correct category. Originally, this form of action was considered indexing, but nowadays so many documents are available online, the applicability of this task has broadened. Some of the most obvious tasks are related to email: for example automatic forwarding of emails to appropriate company department or detecting a spam. The spreadsheet model with one column corresponding to the correct answer is the common classification model for data, and the transformed text data can be combined with standard numerical data-mining data (Weiss & Indurkhya, 1998).

Text-mining is widely used for retrieval of data. A basic concept of information retrieval is measuring similarity: a comparison is made between documents. For such comparison even a small set of words input into a search engine can be considered as a document that can be matched to others. Thus, measuring similarity is fundamental to information retrieval (Weiss, 2005).

Last, but not least, are clustering and organizing of documents. Text mining is very useful for clustering and organizing unstructured data. This is especially useful for analyzing high-dimensional data where connections are not easy to identify (Weiss, 2005).

Text-mining is most often used for classification, retrieval of information and clustering and organizing documents. This is obviously not a complete list of text-mining functions. Text-mining software's that are available today are very advanced, thus their application and analyzing potential is much bigger.

### 3.1. Self-Organizing Maps

In this section the concept of self-organizing map (SOM) is presented. In general SOM is developed by means of software, so very detailed explanation of technical procedures of this method or indepth understanding is perhaps not needed, nevertheless it is worth to present the concept itself and some basic mechanisms involved in the process of creating the map.

#### 3.1.1. Introduction

The self organizing map is an effective software tool for the visualization of high-dimensional data (Kohonen, 1998). It allows an orderly mapping of a high-dimensional distribution onto a regular lowdimensional grid. This tool converts complex and non-linear statistical relationships between data into simple geometric relationships on a low-dimensional display (Kohonen, 1998). In other words, humans are not able to visualize high-dimensional data, so this technique solves this problem by reducing dimensions and producing a map of one or two dimensions where similar data is simply clustered together.



Figure 4. Fragment of self-organizing map of business and management journals. (Linton et al. 2009, pp.221)

Kohonen networks are the basic types of self-organizing networks. It is this self-organizing ability what makes these networks very useful, for instance for adaptation to the unknown input data. The working way of network resembles the learning method that is used by human brain – there are no ready formulas, brain learns during the normal functioning. Term 'Kohonen networks' are used to name the entire group of networks, where the type of a learning process is competitive and self-organized. This learning process is based on providing signals at the entry of the network, and choosing in the competitive process the 'winning neuron', which is the most similar to the input vector (M. Embrechts & Archiniegas, 2000; Kohonen, 1998).

#### 3.1.2. Architecture

First component of self-organizing map is n-dimensional data which is later projected into one or two dimensional map that is easier for visual understanding of relation between data. Second component are the weight vectors. Each vector has basically two components: 1) the data, and 2) its natural location (position on the map) (Kohonen, 1998).

The structure of the network is fundamental. The single neuron is rather simple and cannot do much on its own, but many neurons together allow conducting more complicated operations. The architecture of the Kohonen networks is feedforward and usually based on a single layer, and each neuron must have as many inputs as the whole system. This way all neurons can compete. The Figures 5 and 6 below presents the structure of feedforward single-layer network.



Figure 5. Schematic view of a feedforward single-layer artificial neural network.



Figure 6. 2-dimensional map of neurons.

#### 3.1.3. Main principle

Self-organizing of the network takes place in three stages: 1) construction, 2) learning, and 3) identification. In the first stage the network receive input signals which stimulate neurons, in the second stage competitive learning process takes place. Learning process is based on mechanism which identifies the degree of similarity between every neuron's weight and input signal. More, it assigns the unit with the beat match – the winner. In the last stage the network adapts weights of winning vector and his neighbors, based on response strength. Then a map is organized and similar data is grouped together (Kohonen, 1988, 1990, 1998; Oja & Kaski, 1999).

#### 3.1.4. Main algorithm

The SOM learning principle can be also represented mathematically. The basic algorithm can be written in a compressed form as:

$$m_1(t+1) = +\alpha(t)s(t)\delta_{ci}[x(t) - m_1(t)]$$

Where x(t) are vectors belonging to some predefined class, and model vectors  $m_1(t)$  are labeled by symbols corresponding to the predefined class.  $\alpha(t)$  is the scalar-valued learning-rate factor, and  $\delta_{ci}$  is the Kronecker delta (Kohonen, 1998).

Furthermore, Kohonen defined two types of proximity, rectangular and Gaussian, however this neighborhood function is most often taken to be the Gaussian (Kohonen, 1998):

$$h_{c(x),i} = \alpha(t)exp\left(-\frac{\|\mathbf{r}_1 - \mathbf{r}_c\|}{2\sigma^2(t)}\right)$$

Where  $0 < \alpha(t) < 1$  is the learning-rate factor, which decreases in time,  $r_1$  and  $r_c$  are the vector locations on the display grid, and  $\sigma(t)$  corresponds to the width of the neighborhood function, which also decreases in time (Kohonen, 1998).

Above information constitute the theory Kohonen networks. More detailed information is presented in various publications of Kohonen.

#### 3.2. SOM application potential

The application potential of self-organized maps is huge, therefore this method had been already used in many studies and various disciplines. Most often used applications are classification and categorization. Practical applications of document classification range from spam filtering, document authentication, plagiarism detection, document searching, homeland security, and fault diagnosis (M. J. Embrechts, Linton, Bogaerts, Heyns, & Evangelista, 2007).

SOM has been successfully used in chemistry, fluidics, hydrology, neurology, physics, and more. However, in those fields SOM is not based on text analysis, but numerical data. The application of SOM in those disciplines is used for classifications, but not only. It is often used for forecasting natural phenomenon like rain (Srivastava, Panda, & Mondal, 2010), or developing a model for earthquakes (Yamamoto, Yoshino, & Kawamura, 2010), for controlling artificial locomotion (Santana & Araujo, 2010), and many more.

In the last years SOM got more popular in social sciences as well. Embrechts et al. (2007) presented a recent study where readability fingerprints where used for document classification. Document fingerprints are based on readability indexes that define how easy the text is to understand for particular population. Readability indexes have been developed for different groups of readers, for instance for students in the lower grade levels, university students, foreign students with English as their second language, military pilots and soldiers. Similarly, different readability indices has been used to for computer manuals, military instruction manuals and electronic documents.

In the study of Embrechts et al. (2007) readability fingerprints were applied to encode 71 books in 5 different languages: English, Latin, Dutch, German and French. Then a self-organized map was developed, where languages were clustered together based on their similarity.

Maps like that can also be used to organize various texts, speeches (transcripts), languages, topics, newspapers, emails, books, journals, and many more.

### 3.3. Entrepreneuship and text analysis

The developments of the field were previously studied by several researchers (Ratnatunga et al 1997, Cornelius et al. 2006, Gregoire et al. 2006) The most common method used in the previous studies was a citation analysis. In Cornelius et al. study bibliometric data from the Social Science Citation Index is used in order to determine the developments of entrepreneurship research. The approach was not unique and was previously used (Reader & Watkins, 2001), however the approach of Cornelius et al. is more revolutionary. Authors have taken a snapshot of the key researchers in the field in the different periods of time and then conducted a co-citation analysis of refereed academic articles that included the word "entrep\*" in the title, key words, or abstract between 1986 and 2005. Similar approach was applied in the research of Gregiore et al, (2006). Authors also conducted a co-citation analysis of co-citation networks that emerge from the 20,184 references listed in the 960 full-length articles published in the Babson College Entrepreneurship Research Conference's *Frontiers of Entrepreneurship Research* (FER) proceedings between 1981 and 2004 (Gregoire, et al., 2006).

The results of these two studies provide a detailed analysis of given time periods in which core topics and authors are identified. The above described methods are very effective, quick (special software is available) and allow a large scale research. However it needs to be stressed that citation analysis is subject to many limitations, including the differential influence exercised by various number of researchers across sub-disciplinary fields (Theoharakis & Hirst, 2002), perfunctory mention citations (Kotler, 1972), not used citations, or citations that are used in order to please potential reviewers (Tellis, Chandy, & Ackerman, 1999). This list of criticism is not complete, some other authors add some more critique: citation analysis are biased in favor of popular authors (May, 1967), or those who write reviews (Woodward & Hensman, 1976), and finally an argument that methodological articles and authors in established fields with many researchers tend to be cited more often (Margolis, 1967). Therefore, citation and co-citations analysis can potentially miss out the important, but not yet popular topic.

Text-mining and Self Organizing Maps are techniques that are fully automated, therefore they are objective and quick. A basic concept of text-mining is measuring similarity, but text mining also is very useful for clustering and organizing unstructured data. This is especially useful for analyzing high-dimensional data where connections are not easily visible (Weiss, 2005). Text-mining allows also to build list of words, or terms that are most frequently found in a given document, or group of documents. This way an objective list of the most popular words or terms in the group of articles can be identified. This method is more reliable than a citation-analysis, because citation analysis can potentially exclude topics that are not cited, while text mining allows to identify all the topics regardless of them being cited or not. Thus, even emerging topics will show on the list, therefore researchers can not only analyze past trends, they can analyze current trends as well, and possibly point out to new directions of the field development.

### 3.4. Entrepreneurship by SOM

Kohonen maps are probably the most known type of artificial neural network, and although their structure is rather simple (feedforward, single-layer), application potential is really huge. Self-organizing maps have been successfully used in other disciplines, however they were never applied to entrepreneurship research. Text-mining and self-organizing maps used to analyze entrepreneurship literature will allow discovering the core topics of the field and detect how these topics changed in time. Furthermore, a closer look at journals is also taken, what results in map of journals and their typically published contents. All taken together, this new technique of analysis of entrepreneurship literature will put some light on the research structure and will give researchers and students a valuable and useful knowledge about the field in which they are active.

### 4. Methodology

#### 4.1. The software

Text-mining software used in this research is a state-of-the-art tool developed by Prof. Mark J. Embrechts of School of Engineering, Rensselaer Polytechnic Institute in Troy, New York. This software allows description of text, categorization and classification, text fingerprinting, auto-summarization, creating dictionaries of single or multiple words, frequency counts, discovery of trends, building self-organizing maps, and more.

### 4.2. Data collection

First, a list of scientific journals dedicated to entrepreneurship research has been selected by consulting 'Business' and 'Management' categories of the Social Sciences Citation Index. Nine journals have been selected. This list includes the most respected and influential journals that can be found in all important and relevant search engines and databases. Nine journals is of course not exhaustive, but the number of journals was limited for two reasons, first limited time for conducting this research, and second, availability – only journals subscribed by University of Twente could be considered. Below is the list of nine entrepreneurship journals included in the study.

- 1) Journal of Business Venturing (JBV)
- 2) Small Business Economics (SBE)
- 3) Entrepreneurship: Theory and Practice (ETP)
- 4) Entrepreneurship and Regional Development (ERD)
- 5) Journal of Small Business Management (JSBM)
- 6) International Small Business Journal (ISBJ)
- 7) Family Business Review (FBR)
- 8) Technovation (Tech)
- 9) Strategic Entrepreneurship Journal (SEJ)

All articles from between 2006 and 2010 inclusively are downloaded from the Web of Science for each of the journal. Each of the articles has been given a code that enables to track the article back to the original journal, issue number and year. All special issues together with introduction articles are included. However, it needs to be noted that the number of issues per year and number of articles per issue varies across different journals. Some journals have 12 issues with 11 to 23 articles per issue, while others have only 6 issues per year, and 5 or 6 articles per issue. In total there are 1 671 articles collected.

### 4.3. The procedure

#### 4.3.1. Generating a dictionary

All articles that are in pdf format are converted into txt format with use of software. Then a flat file consisting of all files is created. After that, a 'library' is made for every word or word pair that appears at least once in the examined file. The frequency of every single word of word pair is established by use of text-mining software. Next, the frequency of words and words pairs is counted for the set of all of 1 671 articles. Simple common words like I, a, and, the etc. are not included in the analysis. Most frequently occurring words or set of words are then examined and analyzed and the words that are not specific to business or entrepreneurship, like geographic names or general words like research, study etc. are excluded. Also terms that could be as well specific entrepreneurship terms, as journal names, are excluded (i.e. learning organization, strategic management, family business). Excluding these terms will help to avoid 'false positive' effect, where for instance term 'strategic management' is found to be very frequent in the literature, but is also mentioned repeatedly in the references. Then a 'dictionary' for common terms for entrepreneurship will be developed out of most frequently occurring words. This dictionary consists out of 637 terms.



Figure 7. Dictionary creating process.

Next to the general dictionary of all terms found in entrepreneurship literature, dictionaries for most common terms found in given years and given journals are developed as well. In total there are 15 dictionaries developed.

#### 4.3.2. Text alignment

To mine text, it needs to be converted into a form that text-mining procedures can use. This usually involves generating features in a spreadsheet format, because classical text-mining looks at highly structured data. Nevertheless, text-mining procedure are highly developed, therefore most time spent on text-mining project is for data preparation (Weiss, 2005).

The first step in handling text is to break the stream of characters into words, or more specifically *tokens*. This is fundamental for further analysis. Without identifying tokens, it is difficult to extract higher-level information from the documents. Then, all words are stemmed to the root. The intent of these stemmers is to reach the root form of the word without inflectional or derivational prefixes and suffixes. For example, the word 'denormalization' would be reduced to the stem 'norm'. The end result of such aggressive stemming is to reduce the number of words (types) in texts, thereby making

distributional statistics more reliable. Overall, stemming will result in large reduction in dictionary size, but on the other hand, the smaller the dictionary, the more intelligence in its composition is needed to capture the most and the best words (Weiss, 2005).

Next, words mapped to a single root (like: system, firm, internal, etc) are normalized and scaled. The process is often referred as alignment procedure. The graphic below presents the alignment procedure:



Figure 8. Alignment procedure.

To simplify, every article is compared to the 'dictionary' in order to determine how many times each word was used in the article. These counts are divided by the total number of words in the article what results in the vectors describing the occurrence of commonly occurring words and word pairs. Once the process is completed for all articles examined, all vectors describing the topics are entered in the Kohonen Self-Organizing map (SOM) program, where similar vectors are placed close to each other (Deboeck & Kohonen, 1998; Kohonen, 1988, 1990, 1998; Linton, et al., 2009). This way a map of articles is presented, where similar contents, cluster together, while dissimilar ones are placed far from each other. Although the map seems to be 2-dimensional, in fact it has the shape of a donut, thus the cells at the edges are actually connected.

For more detailed analysis maps with predefined classes are created. This way map is divided into specified categories, like years of publication, or different journals. The procedure of creating those maps is almost the same as described above, the only difference is that the files are associated with a certain category, like for instance a year of publication. Thus, all articles published in year 2006 are associated with class 1, articles published in year 2007 are associated with class 2, and so on. All classes are edited by hand. As a result a map of several colors representing each class is created where files are not only clustered in cells, but different colors also indicate the association to a given class. Therefore, more detailed analysis can be performed.

### 4.4. Data analysis

Both, dictionaries and maps are subject of detailed analysis. Dictionaries, containing the most frequently used words and terms, are great source of information about core topics in the field. Furthermore, dictionaries are created for given journals, as well as for given time periods, therefore in-depth analysis of main focus and changes in focus is possible. Analysis of development of the entrepreneurship research field, especially in the light of earlier research on this subject, is performed. In addition, these dictionaries provide knowledge on particular interest of journals. Formally speaking, they all focus on entrepreneurship, but the field is really complex, and journals dedicated to entrepreneurship in fact differ in their typically covered topics.

Kohonen maps, on the other hand, are more sophisticated tool of analysis. They cluster similar subjects or journals together and allow monitoring trends and changes in them. This tool is especially useful, when there is a need of analyzing a lot of complex data. The representation of data in this case is more straight-forward and connections are more easily visible.

### 4.5. How to read the map



1. Each number on the map represents one article. Different cell colors represent different categories (here: journals).



2. Similar data clusters together – it is either placed in one cell, or in the neighboring cells.



3. The bigger the cluster, the more frequent the topic. In principle clusters show most dominant topics.


4. After analyzing the frequency lists, key topics are manually placed on the map, so that they give a clear representation what the cells exactly represent. The blue cells represent the clusters, however the weight of the topics is not shown.



5. Topics that were key subjects in at least 2 articles, or were placed in the neighboring cells form clusters. Final version of map represents only clusters of most dominant topics in a given year.

The map is a 9 x 18 grid of alveolar cells. Numbers in a given cell represent one article, or multiple articles in case of cells with more than 1 number . Similar articles cluster together, either in one cell, or in the group of cells. Each number on the map represents one article, however for identifying the key topics the original articles are not analyzed. What is analyzed, is the so called 'frequency list'. In the process of creating the Kohonen map, the software creates frequency lists, where all the words and words pairs are listed together with the absolute and relative frequency of appearance in the given text.

Relative frequencies in this case mean that the software calculates how important is a given term to the document, and also to entire collection of documents. In other words, importance is proportional to a number of times word appears in document, but is offset by how frequent the word appears in the corpus of all documents.

To calculate the number of times the word appears in the document the following formula is used:

$$tf_i = \frac{n_i}{\sum_k n_k}$$

To calculate how important the word is to the collection of documents, a more complex formula is being used. Idf means the importance of term by dividing the number of all documents by the number of documents containing the term.

$$idf_i = \log \frac{|D|}{|\{d : d \ni t_i\}|}$$

*tf* term frequency

- *idf* inverse document frequency
- |D| total number of document in corpus
- $|\{d: d \ni t_i\}|$  number of documents where the term  $t_1$  appears

$$tfidf = tf_1 \cdot idf_1$$

Thus, by using the tf-idf procedures in the process of creating the map it is possible to identify how important the word or term is to the collection of group of documents. So the map actually presents the collection of most dominant topics in the whole group of files.

Once the frequencies and inverse frequencies are calculated by the software, there are files created automatically where all the data for every single article is being stored. And these are exactly the files that are being analyzed for deciding on the key topics for every single article mentioned on the map. In case for one article being placed in one cell, the topic that was on the top of the frequency list – on the first or second position – was considered the key topic of that given article. In case of clusters (multiple articles placed in one cell) lists for each article were analyzed and the topic that was common for every article from the top five positions on the frequency lists, was considered the key theme for the cluster. In some articles the chosen topic was on the first position on that list, while for others it could have been placed on second or third position. Very rarely it could have been

the topic placed on fourth or fifth position. Unfortunately the process of identifying the topics was not automatic and needed to be performed manually.

What needs to be mentioned is that the terms given as the most frequent were not always according to the key words provided by the authors. There might be several explanations why it happened. In some cases authors claimed that their article was about innovation, while in fact most frequent term found was 'family business'. That of course does not mean that the article was not about the innovation, however text-mining tool works with words and is unable to catch the authors' intentions or underlying thoughts. It other words, by using text-mining methods the richness of data can be lost.

Another issue can be that authors give a list of key topics to convince reviewers and readers that their article fits into 'typical journal contents'. For that reason most of articles published in Family Business Review have the 'family business' as their key topic, or term 'innovation' for publications in "Technovation' journal.

Once the core topics are identified, the topics are placed, again manually, on the corresponding cell giving a visual representation which topics are present on the map. The map gives a clear representation of the dominant themes, however once the topics are placed on the map, it becomes impossible to judge which topics are the most frequent. To avoid the confusion, a separate list is created where all identified topics are being placed, but with respect to their frequency. In other words, if the topic 'product manufacturing' is the key topic in 12 articles, it is placed 12 times on the list. Next, all the terms are incorporated into a 'word clouds' (see Attachments I – V), where all dominant themes are presented but with respect to their occurrence, therefore the most frequent topics will be visually bigger and more dominant.

Accordingly, a map of clusters of frequent terms is created. Some of the topics form clusters – either on the cell level, or on the map level. 'Cell level cluster' is when multiple articles are placed into one cell, while 'map level cluster' is several cells contain articles about the same subjects. The minimum number of articles in the cluster is 2. These clusters are placed on the map for a better overview of the field structure and easier interpretation of the map.

These maps, word clouds, and lists are compared to the year dictionaries, that were developed in the early stage of the process of map creation – one dictionary for each studied year. Although it might seem that all the analyses and comparisons are doubled and being repeated over and over, they were in fact necessary for deciding on the core topics. To put it simply – all findings needed to be consistent and show the same or very similar results, only then it could be considered reliable and could be trusted.

The results are presented in detail in the following Chapter 5.

## 5. Results

In this chapter the results for the text-analysis are presented. Most dominant topics published in a given year are presented below in the form of graphics together with the description of main observed trends. First results for every year are presented, then the general analysis for four levels of topics (i.e. person, venture, team and environment) is presented.

#### 5.1. Most dominant topics published in 2006

The Map 1 below represents the most dominant topics present in the analyzed journals in the 2006. Each color on the map represents one journal, and the multiple numbers in the cells represent several articles dealing with the same or similar topic.



Map 1. Kohonen map for most dominant topics published in the entrepreneurship literature in 2006.

The Map 2 below represents the encoded map where each cell shows one topic. The cells with a little blue circle represent the cluster – topics that were present in more than one article. All topics presented in the map can be considered the key terms for 2006, although some of these topics are more dominant than others.



Map 2. Most dominant topics published in the entrepreneurship literature in 2006.

The words that best describe the entrepreneurship literature in the year 2006 are 'innovation' (appeared 5278 times in 2006), ' technology' (4167), 'product' (3753), 'network' (2218), and 'manufacturing process' (1865).

Other dominant topics published in 2006 in the entrepreneurship literature were: 'management' (appeared 4818), 'family business', 'entrepreneur' (12094 however this number includes word like 'entrepreneurs', 'entrepreneurial', and 'entrepreneurship'), 'venture capital' (2391), 'information' (4163), ' entrepreneurial planning' (1298), 'social capital' and 'social entrepreneurship' (total number of appearance 3392), 'alliance' (1915), 'business success' (3990), 'business planning' (1298) , and 'organization' (4909, includes terms like 'organizations', 'organizational', etc.).

Content published in 2006 shows an interest in different categories of topics relevant to the entrepreneurship. Researchers have studied topics related to the creation of a new business, innovation and new product, pursuit of opportunities, risk, growth and profits, management, strategy and finance. There is clearly an emergence of a social side of entrepreneurship with topics like 'trust' (1899), 'commitment' (853), and 'social entrepreneurship' (65).

There is little publications on organizational design and almost no publications on entrepreneurshipresearch related topics. The field of research remains broad with several dominant themes.

Map 3 below shows all the clusters present in the entrepreneurship literature in 2006. Because map can only show topics without any numerical data on it, Figure 9 below presents the clusters together with the number of articles in the given cluster. This way one can make conclusions about size and

importance of the clusters. The Figure 9 includes only clusters with minimum 3 articles, because when smaller clusters are also included, the graph is no longer readable.

Number of articles in the cluster does not give an exact measure of how many articles in the entire year dealt with a given topic. It needs to be noted that these numbers are solely based on the clusters derived from the self-organizing maps, and they only give an estimation of a trend, and cannot be treated as an exact indication.



Map 3. Clusters identified in the entrepreneurship literature in 2006.

Clusters with 10 or more articles include key terms like: 'small firm', 'employment', 'manufacturing', 'networks', 'family firm', 'product innovation', and 'technology'. Clusters with at least 5 articles, but less than 10, include topics like 'venture capital', 'family business', 'innovation', 'start', 'women entrepreneur', 'firm', 'innovation process', 'investment', 'knowledge', 'smes', 'credit', 'growth', 'internet information', 'project management' and 'succession planning'. Clusters with less than 5 articles involve topics like 'family', 'firm resources', 'industry', 'innovation networks', 'market', 'organization', 'owner', 'partners', 'product development', 'firm alliance', 'information', 'process', 'product', 'product management', 'supply and demand', and 'trust'.

Although every topic that is placed on the map can be considered the dominant topics, it is pretty apparent that topics forming the clusters are more dominant than others. What can be said for the entrepreneurship literature in 2006 it is that the topics are broad, diverse and focused more on practical aspects of entrepreneurship.



Figure 9. Key topics in the clusters in 2006.

### 5.2. Most dominant topics published in 2007

Year 2007 could be described as the year of an 'entrepreneur'. This term was the most frequent and most dominant in the literature. Researchers were clearly interested in the personal side of the entrepreneurship – entrepreneurial interests, motivations, behavior, background, gender, etc.



Map 4. Kohonen map for most dominant topics published in the entrepreneurship literature in 2007.

The figure below presents the encoded map of topics published in 2007. Although topics like 'innovation' (appeared 4493 times in the literature in 2007) and 'technology' (3815) are still present and quite frequent in fact, researchers start to give some more attention to other topics as well.



Map 5. Most dominant topics published in the entrepreneurship literature in 2007.

As the Map 4 above represents, in 2007 entrepreneurship scholars were interested in topics related to finance: 'bank' (2373), 'loan' (1388), 'investment' (2792), 'venture capital' (1178), and 'financial capital' (118). Topics like 'organizational form' (42), as well as socially oriented topics like 'social capital' (1883), 'social venture' (94) received some attention as well.

Although two themes are clearly dominant – person and finance – the topics are still broad and covering many aspects of entrepreneurship. There is less interest in the creation process, but growing attention is given to market-related topics, including subjects like 'marketing' (821), 'market orientation' (258) and 'international expansion' (42).

For a better overview of the topics, a map of clusters is developed. Below presented Map 6 gives a clear overview of the most dominant groups of topics, and Figure 10 shows number articles present in every cluster. Clusters that are formed out of minimum 10 articles, involve topics like: 'entrepreneur', 'men entrepreneur', 'family firm', 'men firm', 'men', 'organizational form', 'family business', and 'innovation'.

Smaller clusters (3 to 9 articles per cluster) deal with the subject like: 'men business', 'network', 'technology', 'men business', 'work', 'network', 'technology', 'entrepreneurial opportunities', 'smes', 'bank', 'cluster', 'knowledge', 'loan', 'nascent entrepreneur', 'opportunity', 'social capital', and 'venture form'.



Map 6. Clusters identified in the entrepreneurship literature in 2007.

Similarly to the previous year, the research is still not focused, instead it becomes broader with the new topics becoming more dominant. Researchers are interested in entrepreneur, team and venture. There is very little interest in environment or entrepreneurship research itself.



Figure 10. Key topics in the clusters in 2007.

## 5.3. Most dominant topics published in 2008.

Entrepreneurship literature in 2008 did not really show any dominant theme. Instead, researchers demonstrated an interest in the wide variety of topics related to *person, organization, team* and *venture*.



Map 7. Kohonen map for most dominant topics published in the entrepreneurship literature in 2008.

Map 6 on the page 34 presents the map of key themes published in 2008 in the entrepreneurship literature. Terms like 'innovation' (appeared 6062 times in the literature in 2008), 'technology' (4317), 'networks' (1905), 'knowledge' (6499), 'organizational learning' (1706), and 'clusters' (1832) are becoming dominant. Some of the terms were dominant in the previous years, however some terms like 'network ties' (97), 'science park' (218) and 'cluster' appear on the map for the first time.

Another theme dominant in 2008 is finance, with topics like 'ipo' (1162 times), 'capital' (4873), 'performance' (4933), 'investments' (2328), 'venture capital' (1680), and 'financial assets' (877). Topics related to the organization itself emerge with terms like 'organizational change' (48), 'organizational process' (128), 'human resources' (128), and 'venture creation' (279). There is still huge interest in the 'entrepreneur' (15083, however the number includes all terms with the core 'entrepreneur'), and some interest in the environment related topics.



Map 8. Most dominant topics published in the entrepreneurship literature in 2008.

Themes based on articles published in the literature in 2008 demonstrate completeness in the entrepreneurship research. The interests of researchers are broad and cover many aspects of entrepreneurship. It is not possible to identify one or just two most dominant topics – almost every important part of the entrepreneurship phenomenon is being studied in this year. Researchers become more aware of the environment and regions what shows in the interest in 'environment' (2457), and various aspects of international business and entrepreneurship (word 'international' appeared 2758 times in 2008)

When looking at clusters of the most dominant topics published in the entrepreneurship literature in 2008, the field seems to be more focused on entrepreneurship, innovation, technology, networks and knowledge. Although the number of clusters is exact the same as in the year 2007 (both years have 56 clusters), year 2008 shows more clusters with at least 3 articles in.

Most dominant clusters in 2008, formed with minimum 10 articles include following topic: 'entrepreneur', 'ties', 'firm', 'firm innovation', 'knowledge', 'networks', 'technology', 'entrepreneurship', and 'family firm'. Clusters with more than 5, but less 10 articles, include topics like: 'venture capital', 'firm product', 'family business', 'resource', 'self-employment', 'family', 'firm technology', 'ipo', 'performance', 'business region', 'change', 'cluster', 'firm survival', and 'product development'.



Map 9. Clusters identified in the entrepreneurship literature in 2008.

Research in 2008 is still broad and involves wide range of topics. Again there is more focus on practical aspects of entrepreneurship, that on the entrepreneurship as a research field.



Figure 11. Key topics in the clusters in 2008.

### 5.4. Most dominant topics published in 2009.

Entrepreneurship literature in 2009 is somehow more focused on several aspects of entrepreneurship. Some key topics can be definitively identified, for instance 'family firm' (1593 times in all articles in 2009), 'entrepreneur' (15462), 'entrepreneurship', 'technology' (4572), 'innovation' (8374), 'networks' (1420), and 'knowledge' (6447). There is more interest in topics like 'strategy' (1875), 'business growth' (323), 'opportunity' (1565), and 'risk management' (95), similar level of interest in terms like 'investments' (2940), but somehow lower interest in 'venture capital' (only 201 times in the literature in 2009) and term 'loan' that was not even on the top 150 single-word terms.

There is still an interest in the 'entrepreneur' itself (15462), 'business owner' (3699) and 'family firm'. Actually the word 'family' is one of the most frequent terms in this year (10381).



Map 10. Kohonen map for most dominant topics published in the entrepreneurship literature in 2009.

Some new topics start to be dominant, for instance 'trust' (1434), entrepreneurial, organizational and social change (word 'change' alone appeared 3630 times), and 'organizational learning' (213).



Map 11. Most dominant topics published in the entrepreneurship literature in 2009.

When looking at the clusters map of the entrepreneurship literature in 2009, one can get the impression that there are less clusters and that they are somehow smaller in terms of numbers of articles constituting the cluster. Most dominant clusters (minimum 10 papers about the topic) include subjects like: 'firm', 'entrepreneur', 'product', 'innovation', 'service', 'technology', 'family firm', 'family', 'family business', 'business', 'employment', 'investments', and 'knowledge'. Somewhat smaller clusters (between 5 and 9 papers) include following topics:' entrepreneurship', 'business family', 'network', 'work', 'firm innovation', 'learning', and 'venture'.

When compared to previous years, in 2009 there were 48 clusters made of 345 papers. Number of clusters is a bit smaller than in the previous years, however the research is not more focused. The interest areas of researchers remain very broad.



Map 12. Clusters identified in the entrepreneurship literature in 2009.



Figure 12. Key topics in the clusters in 2009.

### 5.5. Most dominant topics published in 2010.

Analysis of the content published in the entrepreneurship literature in the year 2010 has shown some interesting results. Topics related to finance were very dominant. Researchers were interested in terms like 'venture capital' (4248), 'investment' (3558), and 'loan' (700), but for the first time terms like 'financial debt', 'bank debt' and 'firm debt' were actually placed on the map (word 'debt' appeared 1198 in the literature in the 2010).

'Family business' (3909) is still very dominant topic, but some other terms has become popular as well. For instance 'self-employment' was the term present in previous years, but in 2010 it was somehow more dominant (925). Another key terms are 'technology' (3812), 'networks' (2120), 'network ties' (105), 'firm growth' (489), 'social entrepreneurship' (714), 'employment' (2011), and 'new venture' (1303).



Map 13. Kohonen map for most dominant topics published in the entrepreneurship literature in 2010.

Map 10 below presents the key topics published in 2010. Two themes can be considered most dominant – 'finance' and 'venture' related topics.



Map 14. Most dominant topics published in the entrepreneurship literature in 2010.

When it comes to the clusters, entrepreneurship literature seems to be finally a bit more focused than in the previous years. In the map created for 2010 there are 47 clusters, and some of them can be considered quite dominant. Clusters with at least 10 papers seems to be focused on topics like: 'firm', 'form', 'family', 'entrepreneur', 'venture', 'business', 'technology', 'entrepreneurship', 'network', 'social entrepreneurship', 'growth', 'age', 'capital', 'ties', and 'innovation'.

There is big interest in entrepreneur, venture, capital and growth, and innovation. Financial topics are a bit more frequent than in the previous years, what can be perhaps explained by the financial crisis hitting entrepreneurs in 2009. Researchers in 2010 are interested in four different levels of entrepreneurship – *person, team, venture and environment.* The research field is far from focused, but it does show some consistency in the topics of interest.



Map 15. Clusters identified in the entrepreneurship literature in 2010.



Figure 13. Key topics in the clusters in 2010.

#### 5.6. Articles vs. clusters



Figure 14. Number of articles in the clusters.

Figure 14 above presents number of articles, clusters and number of papers inside the cluster for every analyzed year. The number of papers analyzed differs by year, and the number of clusters is decreasing each year. However what is the most interesting to see is the number of articles forming the clusters compared to the total number of articles analyzed each year.

In 2006 there were 67 clusters formed out 270 papers. The total number of articles analyzed for that year was 368, what means that 73% of articles published this year were clustered. The trend is similar for every subsequent year. For 2007 there were 319 articles analyzed, out of which 241 formed 56 clusters (75.5%); for 2008 330 papers were analyzed, and 274 formed 56 clusters (75%); for 2009 out of 345 articles analyzed 297 papers formed 48 clusters (86.6%), and finally for the year 2010 there were 309 articles included in the analysis, and 262 papers formed 47 clusters (85%).

These number of clusters differ per year, but it does get continuously smaller. Nevertheless it is still too big to conclude that the research is narrowing and getting specialized in selected areas.

### 5.7. Convergence by Simpson's diversity index

Simpson's diversity index is used to measure the degree of concentration when samples are classified into types. The measure equals the probability that two entities taken at random from the dataset of interest represent the same type. The formula for Simpson diversity index is following:

$$D = \frac{\Sigma n(n-1)}{N(N-1)}$$

Where:

D= Simpson's index

n = the total number of organisms of a particular species

N = the total number of organisms of all species

The value of D ranges from 0 to 1, where 0 represents infinite diversity and 1, no diversity. This is neither intuitive nor logical, so D is often subtracted from 1 to give Simpson's diversity index (1-D), where the greater the value, the greater the sample diversity. This index represents the probability that two samples randomly selected from a given dataset will belong to different groups. (Simpson E. H., 1949)

Table 2 below represents the values and index for the entrepreneurship data for every analyzed year. The Simpson's index for each year is very high, what means that topics published in entrepreneurship literature are really diverse and two randomly selected topics most probably will not belong to the same category.

	2006	2007	2008	2009	2010
N(N-1)	124256	98910	135792	171810	199362
Σn(n-1)	1894	3122	2684	5100	9502
$D = N(N-1)/\Sigma n(n-1)$	0,015243	0,031564	0,019766	0,029684	0,047662
Simpson's Index of Diversity (1-D)	0,984757	0,968436	0,980234	0,970316	0,952338

 Table 2. Diversity index for every analyzed year.

These findings support what was already evident on the self-organized maps. Although there is a slight decrease in the index over years, it is still very high. This means that entrepreneurship research remains very fragmented, broad, and topics do not seem to be connected.

### 5.8. How did the key topics changed over time

Table 3 presented on p. 55 shows detailed data for most frequently published terms in entrepreneurship literature for every analyzed year. This data allows analysis on how the most dominant topics changed over time. Figures below present change trends for several selected terms:



Figure 15. Changes in topics related to innovation.

Figure 15 shows topics related to 'innovation'. According to the figure above, there are big differences in frequency between word 'innovation' and the rest of the terms. It needs to be noted that single words occur more frequently, than pairs of words, therefore one should look at trends, not values. Interest in the innovation related topics slightly changes in time, however it remains rather stable.



Figure 16. Changes in topics related to finance and growth.

Figure 16 represents changes in topics related to finance and growth. Term 'venture capital' is the most frequently appearing term, with an only exception in 2009 where term 'firm performance' was most frequent. The interest in 'venture capital' dropped rapidly in 2009, while interest in other financially related topics remained stable. In 2010, just after the financial crisis of 2009, the interest in 'venture capital' grew really quickly.



Figure 17. Changes in topics related to entrepreneurship.

According to Figure 18 interest in selected entrepreneurship related topics remained rather stable over analyzed years, however 'social entrepreneurship' became really dominant in 2010. Amount of publications on 'entrepreneurial firms increased quickly in 2010, when compared to 2009, and at the same time interest in corporate entrepreneurship decreased in 2010, when compared to one year before.

		2006			2007		2008		2009			2010				
	Term	Total number of words	Term freq.	%												
1	innovation	2769046	6155	0,22228	2582174	4493	0,17400	2668463	6062	0,22717	2854156	8374	0,29340	2764821	5868	0,21224
2	venture capital	2769046	2391	0,08635	2582174	1232	0,04771	2668463	1680	0,06296	2854156	380	0,01331	2764821	3159	0,11426
3	firm performance	2769046	695	0,02510	2582174	592	0,02293	2668463	855	0,03204	2854156	636	0,02228	2764821	833	0,03013
4	competitive advantage	2769046	646	0,02333	2582174	585	0,02266	2668463	637	0,02387	2854156	636	0,02228	2764821	449	0,01624
5	product development	2769046	778	0,02810	2582174	382	0,01479	2668463	365	0,01368	2854156	554	0,01941	2764821	303	0,01096
6	entrepreneurial activity	2769046	366	0,01322	2582174	408	0,01580	2668463	609	0,02282	2854156	446	0,01563	2764821	299	0,01081
7	product innovation	2769046	397	0,01434	2582174	217	0,00840	2668463	391	0,01465	2854156	405	0,01419	2764821	515	0,01863
8	economic growth	2769046	224	0,00809	2582174	117	0,00453	2668463	236	0,00884	2854156	389	0,01363	2764821	277	0,01002
9	technology transfer	2769046	290	0,01047	2582174	328	0,01270	2668463	365	0,01368	2854156	332	0,01163	2764821	353	0,01277
10	entrepreneurial firms	2769046	320	0,01156	2582174	299	0,01158	2668463	306	0,01147	2854156	213	0,00746	2764821	388	0,01403
11	venture creation	2769046	209	0,00755	2582174	302	0,01170	2668463	279	0,01046	2854156	395	0,01384	2764821	317	0,01147
12	risk taking	2769046	215	0,00776	2582174	221	0,00086	2668463	187	0,00701	2854156	390	0,01366	2764821	432	0,01562
13	nascent entrepreneurs	2769046	99	0,00358	2582174	97	0,00376	2668463	165	0,00618	2854156	300	0,00105	2764821	168	0,00608
14	social networks	2769046	211	0,00762	2582174	307	0,01189	2668463	118	0,00442	2854156	250	0,00876	2764821	393	0,01421
15	entrepreneurial orientation	2769046	145	0,00524	2582174	217	0,00840	2668463	221	0,00828	2854156	272	0,00953	2764821	308	0,01114
16	financial performance	2769046	250	0,00903	2582174	188	0,00728	2668463	180	0,00675	2854156	218	0,00764	2764821	255	0,00922
17	entrepreneurial opportunities	2769046	174	0,00628	2582174	350	0,01355	2668463	105	0,00393	2854156	245	0,00858	2764821	214	0,00774
18	new technology	2769046	283	0,01022	2582174	193	0,00747	2668463	177	0,00663	2854156	222	0,00778	2764821	178	0,00644
19	innovation process	2769046	410	0,01481	2582174	165	0,00639	2668463	216	0,00809	2854156	306	0,01072	2764821	225	0,00814
20	social entrepreneurship	2769046	65	0,00235	2582174	94	0,00364	2668463	75	0,00281	2854156	121	0,00424	2764821	714	0,02582
21	corporate entrepreneurship	2769046	209	0,00755	2582174	185	0,00716	2668463	206	0,00772	2854156	254	0,00890	2764821	136	0,00492

**Table 3.** Changes in key entrepreneurship topics.

### 5.9. Group analysis

The analyzing software provides very reach data on every single word or phrase that was used for creating the map. Although word-based analysis for every single word, or even topic, could be potentially interesting, it is outside the scope of this paper. However, the data collected for every word is later grouped and used for words group analysis on four different levels: person, team, venture and environment.

The values collected for every word are 1) absolute values, 2) relative frequencies, 3) values normalized against the maximal word count in the entire collection of documents, and 4) average values. Analysis based on each type of values show consistent results, therefore for the group analysis the absolute values are being used.

As it was already mentioned the words are being grouped into for levels:

- 1) *Person* with topics like entrepreneur, female entrepreneur, owner, founder, behavior, skills, experience, education, etc.
- 2) *Team* with topics like team, family, employees, etc.
- 3) *Venture* with topics like organizational form, performance, growth, production, innovation, strategy, etc.
- 4) *Environment* with topics like risk, opportunity, network, ties, clusters, market, etc.

Figures 10 - 14 presented on page 39 represent the shares of the particular groups within the total number of publications for every analyzed year.

The group analysis shows that the most dominant group of topics is related to the person with the share of 37% in 2006, 52% in 2007, 44% in 2008, 47% in 2009 and 42% in 2010. Even though the interest slightly changes over years, this group of topics does remain the most dominant.

The least dominant are topics related to environment with shares between 8% in 2008, 9% (2009), 11% (2006, 2010) and 12% in 2007. The interest in environment related subjects is constant and remains at the stable level.

Topics related to team and venture receive a comparable level of attention. Interest in venture related topics oscillates between minimal 22% in 2007 and 27% in 2010. The interest slightly grew and it will most likely continue to grow.

Topics related to team has made 26% of all published content in 2006, then it dropped to 14% in 2007, grew again in 2008 to 23%, and from 2009 onwards its declining again.

It seems that in the period between 2006 and 2010 the focus is on the person.











**Figure 18.** Shares of particular topics groups within the total number of publications for 2006.

**Figure 19.** Shares of particular topics groups within the total number of publications for 2007.

**Figure 20.** Shares of particular topics groups within the total number of publications for 2008.

**Figure 21.** Shares of particular topics groups within the total number of publications for 2009.

**Figure 22.** Shares of particular topics groups within the total number of publications for 2010.

# 6. Conclusions and discussion

Based on the text and SOM analysis, several conclusions about the entrepreneurship research and its structure arise. Main and most apparent conclusion is that the entrepreneurship research field remains very broad. It is possible to identify several key themes that can be considered as constant and fixed elements for entrepreneurship research, however the field resembles more of a patchwork than a solid, homogenous structure. These findings are obviously not unique, already Davidsson (2004) pointed out that fragmentation of the field is one of the factors that make it difficult to establish its structure.

Findings presented in this master thesis support the results of earlier studies. Morris, Lewis and Sexton (1994) have identified several most dominant themes in the entrepreneurship research as creation, innovation, pursuit of opportunities, risk management, uncertainty, pursuit of profit, personal advantages, new production methods, management, coordination of resources and value creation. Even though their research was published almost 20 years ago, their findings are still actual. However results of this research can add few positions to this list. These new positions would be:

- Finance,
- Social entrepreneurship,
- Environment awareness,
- Organization theory,
- Strategy, and
- Marketing.

Although one can argue that above mentioned themes are not entrepreneurship specific and some of them are even independent and well established fields of science, it is pretty apparent that these topics, together with those mentioned by Morris et al., are like jigsaw puzzle that together make up a whole. It is worth to notice that Kohonen maps actually offer excellent graphical representation of the proposed concept – one should really see the entrepreneurship field as a map consisting out of many smaller pieces.

These pieces, according to the SOM analysis are: 'innovation', 'venture capital', 'firm performance', 'competitive advantage', 'product development', 'entrepreneurial activity', 'product innovation', 'economic growth', 'technology transfer', 'entrepreneurial firms', 'venture creation', 'risk taking', 'nascent entrepreneurs', 'social networks', 'entrepreneurial orientation', 'financial performance', 'entrepreneurial opportunities', 'new technology', 'innovation process', social entrepreneurship', and 'corporate entrepreneurship'. These are the most dominant themes across all analyzed years, and even though there are slight frequency fluctuations, these are the key subjects in entrepreneurship research.

Self-organized maps do not only provide the list of most dominant topics, the method also allows to cluster similar data together. Clusters are based on number of articles, so that not only the topic can be identified, but also the number of articles in which the topics was published. This means that one can analyze cluster's size and its key subjects.

The sizes and topics of clusters differ per year, but terms like: 'firm', 'form', 'family', 'entrepreneur', 'venture', 'business', 'technology', 'entrepreneurship', 'network', 'social entrepreneurship', 'growth', 'age', 'capital', 'ties', and 'innovation' seem to be stable elements of clusters in every given year.

Clustering of data means that research is focused on selected areas of topics, however in case of an entrepreneurship research the number of clusters in every year is very big (67 clusters in 2010, 56 clusters in 2006 and 2007, 48 clusters in 2009, and 47 clusters in 2010). Even though the number of cluster is slightly decreasing with every year, it does not mean that the research is narrowing – the amount of clusters in every year is still too big meaning that entrepreneurship research is broad and remains fragmented.

Simpson's diversity index used to measure the degree of similarity in data collected from entrepreneurship research supported the findings reported on the maps. Simpson's index ranges from 0 to 1, where the greater the value, the greater the sample diversity. This index represents the probability that two samples randomly selected from a given dataset will belong to different groups. (Simpson E. H., 1949)

In case of terms present in entrepreneurship literature. the Simpson's index calculated for each year is very high, what means that topics published in entrepreneurship literature are really diverse and two randomly selected topics most probably will not belong to the same category. This proves that entrepreneurship research remains very fragmented, broad, and topics do not seem to be connected.

Further, a group analysis was performed for four levels of analysis – *person, team, venture*, and *environment*. Group *Person* included topics like entrepreneur, female entrepreneur, owner, founder, behavior, skills, experience, education, etc. Group *Team* involved topics like team, family, employees, *Venture* included topics like organizational form, performance, growth, production, innovation, strategy, and finally *Environment involved* topics like risk, opportunity, network, ties, clusters, market, etc. Group analysis has shown that the most dominant group of topics is related to the person with the share of 37% in 2006, 52% in 2007, 44% in 2008, 47% in 2009 and 42% in 2010. Even though the interest slightly changes over years, this group of topics does remain the most dominant.

The least dominant are topics related to environment with shares between 8% in 2008, 9% (2009), 11% (2006, 2010) and 12% in 2007. The interest in environment related subjects is constant and remains at the stable level.

Topics related to team and venture receive a comparable level of attention. Interest in venture related topics oscillates between minimal 22% in 2007 and 27% in 2010. The interest slightly grew and it will most likely continue to grow.

Topics related to team has made 26% of all published content in 2006, then it dropped to 14% in 2007, grew again in 2008 to 23%, and from 2009 onwards its declining again. All in all, it seems that in the period between 2006 and 2010 the focus is on the person.

Moreover, next to the ongoing discussions about defining entrepreneurship and its research domain, there are constant debates about convergence and maturity of the field. Convergence is the idea that when a research field matures, it becomes increasingly characterized by a set of codified theories, models, methods, and/or measures - which are to direct ongoing research (Grégoire et al, 2006). In case of entrepreneurship research convergence is difficult to see, because although the field has its own core topics, authors, journals and conferences, paradoxically there are no entrepreneurship theories. Cornelius et al. (2006) has mentioned in their article that according to their findings the field becomes self-reflective in a way that it researches itself , however results presented in this master thesis do not support this statement. Topics like 'convergence', 'field development', 'entrepreneurship research' were not placed on any single map, what means they were not dominant in any of the studied years. It is also noteworthy that papers of Gregiore et al. (2006) and Cornelius et al. (2006) that dealt with the convergence and development of the field, have never made it to the so-called main stream. Researchers are still more interested in the practical aspects of entrepreneurship, than in the field itself.

Furthermore, Cornelius et al believe that the entrepreneurship research is moving from emerging field to the mature field. Grégoire et al. have described given time periods, in which they clustered the main topics and main researchers of four different times period. According to their findings the period between 1981 and 1986 was characterized by the focus on 'person' with his/her personal traits, motivation, age, education, etc. Next period, 1987-1992, was the time were new topics emerged, like 'process view on entrepreneurship', 'new venture development', 'emerging organizations',' new ventures strategies', and the' performance of new ventures'. Period 1993-1998 was the period of 'strategy', and in the last described period, between 1999-2004, researchers have shown interest in topics like "*Opportunity identification and exploitation, organizational emergence, the relationships between social capital, inter-organizational learning and innovation, the implications of entrepreneurship as a firm-level dimension, and the syndication network of venture capitalists financing arrangements.*" Also noticeable is the emergence of new themes, namely the entrepreneurship research on the concepts of opportunity, organizational emergence, social capital, and cognitive psychology.

This paper brings the research further and completes the story by the most up-to-date results. According to the SOM outcomes, the period between 2006 and 2010 is best characterized by the word 'innovation'. The findings suggest that innovation related topics are very dominant in each studied year and it is pretty obvious that entrepreneurship and innovation are inseparable. However, researchers also became more aware of the gender, social and environmental topics. The interest in these themes is constantly growing and this trend will most likely continue.

All in all, this paper brings valuable results and enhances the body of knowledge. Obtained results have allowed to make some conclusions about the general structure of the field and most dominant themes in the entrepreneurship research. Based on the data it is not possible to make any predictions about the future trends, but it is very likely that the field will be evolving and real-time situations like financial-crisis, politics, and national regulations will influence the direction of the research. Entrepreneurship as a research field may be young and immature, but nobody can claim anymore that it is illegitimate or not independent.

### 7. References

- Acs, Z. J. / Audretsch D. / Evans D. (1994): "Why Does the Self-Employment Rate Vary Across Countries and Over Time?" in: Discussion Paper No. 871, Center for Economic Policy Research.
- Acs, Z.J. / Armington C. (2004): "Employment Growth and Entrepreneurial Activity in Cities", in: Regional Studies 38 (8): 911-927.
- Acs, Z. J. / Audretsch J. / Braunerhjelm P. / Carlsson B. (2004): "The Missing Link: The Knowledge Filter and Entrepreneurship in Endogenous Growth." Discussion Paper, Center for Economic Policy Research, London, UK.
- Acs Z.J. (2007): How is entrepreneurship good for economic growth? in: Innovations, 1(1), p. 97-107
- Audretsch, D.B. / M.A. Carree / Thurik A.R. (2001): Does entrepreneurship reduce unemployment? Discussion paper TI01-07413. Tinbergen Institute, Erasmus University, Rotterdam.
- Brush, C. G., Duhaime, I. M., Gartner, W. B., Stewart, A., Katz, J. A., Hitt, M. A., . . . Venkataraman, S. (2003). Doctoral education in the field of entrepreneurship. Journal of Management, 29(3), 309-331.
- Carree, M. / van Stel A. / Thurik R / Wennekers S. (2002): "Economic Development and Business Ownership: An Analysis Using Data of 23 OECD Countries in the Period 1976-1996.", in: Small Business Economics 19(3): 271-290.
- Carree M.A. / Thurik A.R. (2005): The Impact of Entrepreneurship on Economic Growth. In: Handbook of Entrepreneurship Research An Interdisciplinary Survey and Introduction, 437-471. International Handbook Series on Entrepreneurship. Edited by: Zoltan J. Acs Z.J. / Audretsch D.B. (2005)
- Carree M.A./Thurik A.R. (2006): Entrepreneurship and economic growth. Edward Elgar Publishing Limited
- Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. Academy of Management Review, 34(3), 511–532.
- Cornelius, B., Landstrom, H., & Persson, O. (2006). Entrepreneurial studies: The dynamic research front of a developing social science. Entrepreneurship Theory and Practice, 30(3), 375-398.
- Davidsson, P., Low, M. B., & Wright, M. (2001a). Editors' introduction: Low and Macmillan ten years on: Achievements and future directions for entrepreneurship research. Entrepreneurship Theory and Practice, 25(4), 5-15.
- Davidsson, P., Low, M. B., & Wright, M. (2001b). Editors' introduction: Low and Macmillan ten years on: Achievements and future directions for entrepreneurship research. Entrepreneurship Theory and Practice, 25(4), 5-15.
- Davidsson P. (2004): Researching Entrepreneurship. Springer Science + Business Media, Inc.
- Deboeck, G. J., & Kohonen, T. (1998). Visual explorations in finance : with self-organizing maps. New York: Springer.
- Embrechts, M., & Archiniegas, F. (2000). Neural Networks for Text-to-Speech Recognition. Proc. IEEE International Conference on Systems Man and Cybernetics (SMC'OO), Nashville, Tenessee, pp. 3582-3587.

- Embrechts, M. J., Linton, J., Bogaerts, W. F., Heyns, B., & Evangelista, P. (2007). Automated Text Categorization Based on Readability Fingerprints. ICANN'07 Proceedings of the 17th international conference on Artificial neural networks Springer-Verlag Berlin, Heidelberg.
- Gregoire, D. A., Noel, M. X., Dery, R., & Bechard, J. P. (2006). Is there conceptual convergence in entrepreneurship research? A co-citation analysis of Frontiers of Entrepreneurship Research, 1981-2004. Entrepreneurship Theory and Practice, 30(3), 333-373.
- Hansson, B. G. (1993). Vetenskapsfilosofi [Philosophy of science]. Lund: Filosofiska Institutionen.
- Kohonen, T. (1988). Self-organization and associative memory (2nd ed.). Berlin ; New York: Springer-Verlag.
- Kohonen, T. (1990). The Self-Organizing Map. Proceedings of the IEEE 78(9).
- Kohonen, T. (1998). The self-organizing map. Neurocomputing, 21, 1-6.
- Kotler, P. (1972). Generic Concept of Marketing. Journal of Marketing, 36(2), 46-54.
- Linton, J. D., Himel, M., & Embrechts, M. J. (2009). Mapping the Structure of Research: Business and Management as an Exemplar. Serials Review, 35(4), 218-227. doi: DOI 10.1016/j.serrev.2009.09.002
- Margolis, J. (1967). Citation indexing and evaluation of scientific papers. [Bibliography]. Science, 155(767), 1213-1219.
- May, K. O. (1967). Abuses of citation indexing. Science, 156(3777), 890-892. doi: 10.1126/science.156.3777.890-a
- McMullen, J., & Shepherd, D. (2006a). Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. Academy of Management Review, 31(1), 132–152.
- Morris, M. H., Lewis, P., & Sexton, D. (1994, Winter). Reconceptualizing entrepreneurship: An inputoutput perspective. Sam Advanced Management Journal. 59,21-31.
- Oja, E., & Kaski, S. (1999). Kohonen maps (1st ed.). Amsterdam ; New York: Elsevier.
- Ratnatunga, J., & Romano, C. (1997). A "Citation Classics" analysis of articles in contemporary small enterprise research. Journal of Business Venturing 12, 197-212.
- Reader, D., & Watkins, D. (2001). The intellectual structure of entrepreneurship: An author cocitation analysis. Paper presented at the XV RENT Conference, Turku, Finland.
- Rindova, V., Barry, D., & Ketchen, D. (2009). Entrepreneuring as emancipation. Academy of Management Review, 34(3), 477–491.
- Romer P. (1990): Endogenous technological change. Journal of Political Economy 98, 571-602
- Sandelands, E. (1996). Which journal? The politics of where to publish. Collection Building, 15(1), 27-37.
- Santana, O., & Araujo, A. (2010). A Self-Organizing Map for Controlling Artificial Locomotion Artificial Neural Networks-ICANN 2010, PT II, 6353, 420-425.
- Simpson, E. H. (1949) Measurement of diversity. Nature, 163, 688.
- Solow R. (1956): A Contribution to the Theory of Economic Growth, in: The Quarterly Journal of Economics, Vol. 70, No. 1, pp. 65-94
- Srivastava, G., Panda, S., & Mondal, P. (2010). Forecasting of rainfall using ocean-atmospheric indices with a fuzzy neural technique. Journal of Hydrology. 395(3-4), 190-198.
- Tellis, G. J., Chandy, R. K., & Ackerman, D. S. (1999). In search of diversity: The record of major marketing journals. Journal of Marketing Research, 36(1), 120-131.
- Theoharakis, V., & Hirst, A. (2002). Perceptual differences of marketing journals: A worldwide perspective. Marketing Letters, 13(4), 389-402.
- Weiss, S. M. (2005). Text mining : predictive methods for analyzing unstructured information. New York: Springer.
- Weiss, S. M., & Indurkhya, N. (1998). Predictive data mining : a practical guide. San Francisco: Morgan Kaufmann Publishers.
- Wennekers, A.R.M./Thurik A.R. (1999): Linking entrepreneurship and economic growth. Small Business Economics, 13, 27-55.
- Westhead, P., Wright, M., & McElwee, G. (2011).Entrepreneurship: Perspectives and cases. London: Pearson.
- Woodward, A. M., & Hensman, S. (1976). Citations to Review Serials. Journal of Documentation, 32(4), 290-293.
- Wong P.K./Ho Y. P./Autio E. (2005): Entrepreneurship, Innovation and Economic Growth: Evidence form GEM data, in: Small Business Economics 24: 335-350
- Yamamoto, T., Yoshino, H., & Kawamura, H. (2010). Simulation study of the inhomogeneous Olami-Feder-Christensen model of earthquakes. European Physical Journal B, 77(4), 559-564.
- Zahra, S. A. & Wright, M. (2011). Entepreneurship's Next Act. The Academy of Management Perspectives. 25(4), 67-83.

## 8. Appendix

Based on data retrieved from the articles published in the entrepreneurship literature between 2006 and 2010, word clouds represent key terms for each year were created. These word clouds are presented in this section.

- I. Most dominant topics found in the selected entrepreneurship literature in 2006.
- II. Most dominant topics found in the selected entrepreneurship literature in 2007.
- III. Most dominant topics found in the selected entrepreneurship literature in 2008.
- IV. Most dominant topics found in the selected entrepreneurship literature in 2009.
- V. Most dominant topics found in the selected entrepreneurship literature in 2010.

I. Most dominant topics found in the selected entrepreneurship literature in 2006.



II. Most dominant topics found in the selected entrepreneurship literature in 2007.



III. Most dominant topics found in the selected entrepreneurship literature in 2008.



IV. Most dominant topics found in the selected entrepreneurship literature in 2009.



V. Most dominant topics found in the selected entrepreneurship literature in 2010.

