THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

INTELLIGENT EXTRACTION OF INTELLECTUAL CAPITAL FOR VALUE CREATION IN KNOWLEDGE-INTENSIVE ORGANIZATIONS

CAI LINLIN

M.Phil

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University

Department of Industrial and Systems Engineering

Intelligent Extraction of Intellectual Capital for Value Creation in Knowledge-intensive Organizations

Cai Linlin

A thesis submitted in partial fulfillment of the requirements for the

Degree of Master of Philosophy

June, 2015

CERTIFICATION OF ORGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Cai Linlin

Abstract

Since the world's economy has rapidly changed from industrial-based to knowledge based, it is pressing for organizations to identify Intellectual Capital (IC) in time and disclose as well as track this kind of important knowledge-based capital timely and accurately, which by doing so not only can improve decision-making processes internally, but also can demonstrate the organizational competitiveness externally. Due to the dynamics and complexity of IC, it remains a significant challenge for organizations and researchers to extract IC from the multitude of materials such as annual reports, IPO prospectuses, sustainability reports, large volume of interview transcripts as well the space of social media, etc.

The conventional method of extracting IC-related information heavily relies on manual content analysis. However, because it is highly time-consuming, the manual method loses its popularity both in academia and in industry, automatic extraction methods that are assisted by computers offer great help to cope with the huge volume of data. However, these computer-assisted methods have the shortfalls of ignoring the context in practical utilization, the complexity of IC components, and the difficulty to replicate the process as well as low accuracy in terms of extraction which also compromises this method's effectiveness to be used in the public domain and on a large scale.

Redressing these problems, an intelligent methodology is designed by using the computational linguistics and artificial intelligence (AI) techniques to extract and analyze IC-related information automatically and intelligently. The demonstrated intelligence is manifested in several ways. Firstly, an IC knowledge repository is constructed based on the IC-related

keywords/phrases and keywords combination patterns which consist of the IC academic and practical repositories. The established IC knowledge repository offers sources of matching words to extract IC-related information. Based on the IC knowledge repository, an IC information extraction algorithm is designed to achieve the goal of extracting the IC information automatically. In this step, the knowledge-based intellectual capital extraction (KBICE) algorithm increases the efficiency of extracting IC-related information. The repository together with the algorithm helps to identify IC-related sentences and paragraphs more accurately and faster than the methods merely using the IC terms checklist. IC sentimental analysis determines the overall nature of the extracted IC information from a news-tenor perspective which can recognize news from negative and positive perspective.

Through setting up an IC knowledge repository, IC-related keywords/phrases improve the ability of extracting practical IC information used. The keywords extraction patterns express the inter-relationship of IC components and increase the accuracy and relevance of the extraction. The dynamic IC term checklist enables the same methodology to be used in various contexts. With this added knowledge, the IC sentimental analysis greatly increases the relevance of Intelligent Extraction of Intellectual Capital (IEIC) in IC research and application development in terms of determining whether the extracted IC information is indeed positive or negative.

After conducting the experiment by using company reports containing IC information, the two parts' results are very encouraging when compared with other existing methods. In testing the KBICE） algorithm, three standards including precision, recalls as well F-measure are adopted to measure the usability of KBICE. Compared to rule-based reasoning (RBR) and bag-of-words (BoW) models, KBICE has demonstrated better results. In testing the sentimental analysis, the processed results of testing IC sentimental analysis of IC news also exhibited high recall and precision accuracies.

Meanwhile, processing company annual reports and online news that contain IC has shown that IEIC can help to produce IC integrated reports. The reports, despite their decline in popularity, can be produced timely and contain more comprehensive information than the conventional manually compiled version.

This research has reveals additional challenges for intelligent extraction of IC information. Many aspects still need to be improved. More refinements need to be made to the automatic extract algorithm to reduce the subjectivity. Companies in more industries also need to be involved by offering more comprehensive data. In addition to IC news, other IC critical information such as employees' social media should also undergo the IC sentimental analysis.

List of publications arising from the study

Dumay, J., & Cai, L. (2015). Using content analysis as a research methodology for investigating intellectual capital disclosure: a critique, Journal of Intellectual Capital, 16(1).

Cai, L., Tsui, E., & Cheung, B. (2014). An exploratory study on an Intellectual Capital eco-system. . Paper presented at the conference of the International Forum on Knowledge Asset Dynamics, Metera, Italy.

Tsui, E., Wang, W. M., Cai, L., Cheung, C. F., & Lee, W. B. (2014). Knowledge-based extraction of intellectual capital-related information from unstructured data. Expert Systems with Applications, 41(4), 1315-1325

Cai, L., Tsui, E., & Cheung, C. F. (2013, May). A taxonomic approach to the identification of intellectual capital from company reports. Paper presented at Software Engineering and Service Science (ICSESS), 2013 4th IEEE International Conference on (pp. 338-341). IEEE.

Cai, L., Tsui, E., & C. F. Cheung. (2013). A Critical Analysis of Intellectual Capital Reports in Banking Industry from 1994 to 2011. Paper presented at the 10th International Conference on Intellectual Capital, Knowledge Management and Organizational Learning (ICICKM 2013). Washington, DC, USA.

Acknowledgement

It was really an exciting moment to write this section. It is really a great moment to express my gratitude to the people that are important to me.

First and foremost, I would like to express my heartfelt gratitude to my chief supervisor, Prof. Eric Tsui, for taking care of me in terms of my life and my research. His support exceeded by far what a research student can expect from a supervisor. It is he who enabled a small country girl from China to have the opportunity to communicate with all kinds of international scholars and travel from America to Europe.

Prof Benny Cheung, my co-supervisor, has also been a very important person in my IC learning journey to whom I would like to express my gratitude. His rigorous attitude to IC research always made my mind clear. He was available to discuss the research wherever I met him. I have been quite moved by his patience.

In addition to the above, special thanks go to Dr John Dumay and Prof. Leif Edvinsson. I am really a lucky student to have had the opportunity to discuss research with such prominent IC researchers in the world. I would also like to pay gratitude to Dr. Wang Wai Ming, a research associate of the KMIRC, without whose great guidance and help, this project could not have been finished so fast.

The work described in this thesis was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 529210). Many thanks are also due to Miss Carrie M.Y. Cheung and Mr. Matthew K.K. Cheung for their technical support of the research work.

My sisters and brothers in Christ are the people I cannot forget to say thank you. It is your love without any condition that has made me arrive at this point. Your encouragement, your prayers as well as your consideration always let me feel the sunshine and warmth in my life.

To my parents, I am full of tears writing thanks to you. Firstly, I am so proud that I have such excellent parents who always try their best to support me financially and spiritually. It is you who taught me that helping others never requires you to wait for you to become strong. It is you who offered me the happy and warm home waiting for me back no matter what situation I was in. Secondly, I am lucky that I have an excellent boyfriend. Thank you so much for taking me into the world of research. Thank you for your company when I am happy and not happy.

In the end, praise my lord for everything you have done. I love you!

Table of Contents

List of figures

List of tables

List of abbreviations

IC - Intellectual capital

HC - Human capital

SC - Structural capital

RC - Relational capital

IEIC - Intelligent Extraction of Intellectual Capital

KBICE - Knowledge-based intellectual capital extraction

.

Chapter 1. Introduction

This section mainly introduces the background of the study, problem statement, research objectives and organization of the thesis.

1.1 Background of the study

In the past two decades, the world's economy has rapidly changed from an industrial to a knowledge-based economy. Intellectual capital (IC) is a kind of non-financial capital having attracted a lot of interest (Edvinsson, 1997a; Roos, Roos, Dragonetti, & Edvinsson, 1997; Grindley&Teece,1997; T. Stewart & C. Ruckdeschel, 1998) in terms of value creation, especially in the context of knowledge-intensive organizations. IC research has entered the fourth stage. The first stage accomplished the important mission of raising the awareness of IC in the knowledge era (Petty & Guthrie, 2000). However, IC was misunderstood due to a lack of empirical studies at that stage. The second stage of IC research mainly focuses on the management, measurement and reporting of IC (Bernard Marr, 2004; Lev, 2004). However, the problem of this stage was that IC research mostly involved accounting issues and a plethora of such IC frameworks, which in turn may have weakened IC's relevance to management. The third stage of IC research began by examining the role of IC in a critical way (Guthrie, Ricceri, & Dumay, 2012). Undoubtedly, IC is the key driver of value creation in an organization. However, not all aspects of IC can create value.

In recent years, additional emphasis has been placed upon the importance of the products and services offered to customers (Dumay & Garanina, 2013). The disadvantage of the third stage of IC research was that intellectual capital (IC) is only seen as a tool to develop strong organizations but, as a result, the sustainable development of the organization is often ignored (Dumay, 2013). The legacy of the former three stages of development has revealed that IC is a complex concept which should be understood within a context. Meanwhile, IC should be studied more from some critical perspectives. Nowadays, IC has been applied in external reporting, auditing, accountability and governance, management control/strategy, and performance measurement (Guthrie, Ricceri, & Dumay, 2012). Among these applications, management control and external reporting are the two most popular areas.

In the fourth stage, the concept of the IC eco-system was proposed. IC research is attached to the social eco-system in which the good quality of human beings' lives and a healthy environment are also contributing factors to good economic development. However, the relevance of IC beyond the organizational dimension is still hidden due to a lack of empirical data. Therefore, a large volume of documents like internal reports, related interviews, surveys, guidelines, as well as the fast-changing social media are all waiting for extraction for both academic and practical application.

To research and manage IC, IC must be found, revealed and managed in the organization. The conventional method of extracting IC-related information is via manual content analysis, a method which has been used for over 14 years. However, manual content analysis of IC information contributes little knowledge to IC research. There are several reasons for this. Firstly, most research does not use content analysis (CA) in the right way which focuses on constructing an IC term checklist and identifying units of analysis. Secondly, the dynamics and complexity of IC are often ignored by the CA method. Thirdly, the news-tenor of IC information does not attract much attention (Dumay & Cai, 2015). Furthermore, manual content analysis is quite labour-intensive and time-consuming. Last but not least, the method's accuracy is greatly affected by personal bias in the coding process. Thus, it is very difficult to apply manual content analysis in the practical world. Therefore, computer-aided CA shows great potential in terms of analyzing IC at both academic and practical levels. However, the current utilization of this technique is that they are still heavily influenced by manual content analysis.

1.2 Problem statement

Based on background study of the development of IC research and the current difficulty encountered by various research techniques, three challenges have been identified

(1) The challenge of designing the appropriate technique for IC extraction and analysis based on the IC theory development and current practical needs.

(2) The need to handle large volumes of fast changing data in order to extract IC in a dynamic and complex context

(3) The accuracy of current techniques in extracting the IC information.

1.3 Research objectives

In order to address the three challenges mentioned and the shortcomings of the current computer-aided technique to extract IC information from large volumes of mostly unstructured data, this project aimed:

(1) To review the development of modern IC theory and research methods

(2) To identify the shortfalls of current techniques in terms of extracting IC information from a large amount of unstructured data

(3) To design a methodology to extract IC information based on the modern IC theory

(4) To validate the above methodology in one industry

1.4 Organization of the Thesis

This thesis is composed of six chapters. The first chapter is about the introduction of the research, including the illustration of the background of the study, the problem statement as well as the research objectives. The second chapter is a literature review. In this chapter, the development of IC research is examined which covers the development of IC-related definition and theory, evolution of major IC research methods to study IC, critical analysis of the manual content analysis as well as the computer-aided content analysis methods for coping with IC-related information. The third chapter presents the methodology, in which the process of designing are introduced including repository construction, knowledge-based intellectual capital extraction as well as IC sentiment analysis. The methodology and evaluation are introduced. The fourth chapter covers the implementation and experiment verification of the whole process. The fifth chapter presents the research discussion including significance and limitations. The last chapter concludes the significance of the research and makes suggestions for future work.

Chapter 2. Literature review

In this section, firstly, the four stages of IC development are examined critically. Secondly, the IC-related research and theories are reviewed. Thirdly, the evolution of IC research methods is summarized. Then the author reviews the development of manual content analysis and automatic content analysis. Finally, a research gap is proposed based on a critical analysis of the methods for automatics analysis of IC.

2.1 The development of IC research

Since intellectual capital (IC) was first proposed by some visionary practitioners (Bontis and Fitz-Enz, 2002; Edvinsson, 1997b; Edvinsson & Malone, 1997; Roos et al., 1997; Thomas Stewart & Clare Ruckdeschel, 1998; Sveiby, 1997), IC research has had four stages of development (see Table 2.1). The first stage accomplished the important mission of raising the awareness of IC in the knowledge era (Richard Petty & Guthrie, 2000). However, IC was misunderstood as there was a lack of empirical studies at that stage. The second stage of ICR mainly focused on the management, measurement and reporting of IC (Bernard Marr, 2004). However, the problem of this stage is that ICR mostly involves accounting issues and all kinds of IC frameworks, which in turn has weakened IC's relevance to management. The third stage of IC research began to examine the role of IC in a critical way (Chatzkel, 2004; Guthrie et al., 2012; O'Donnell, Henriksen, & Voelpel, 2006). Undoubtedly, IC is the key driver of value creation in an organization; however not all IC can create value. In recent IC research, more and more emphasis is given to the importance of the products and services offered to customers (Dumay & Garanina, 2013).

The disadvantage of the third stage of IC research is that intellectual capital (IC) is only seen as the tool to develop strong organizations but as a result the sustainable development of the organization is often ignored (Dumay, 2013). The advent of the fourth stage of IC research is described as IC eco-system attempts to redress this gap (Dumay, 2013; Edvinsson, 2013; Wasiluk, 2013). In this stage, the concept of an IC eco-system is proposed. IC research is attached to the social eco-system in which the good quality of human beings' lives and a healthy environment are added as contributing factors to good economic development. Human beings are treated as social citizens (Allee, 2000; Wasiluk, 2013) rather than simply organizational employees. However, the role of IC in the IC eco-system and the relevance of IC beyond the organizational dimension are still under-explored due to lack of research and practical applications. Thus, with the advent of the new stage, it is necessary to examine critically the development of IC-related theories and the development of an IC research methodology.

Table 2.1 Development of IC research

Stages	Leading researchers	Features	Shortfalls
1st stage	Sveiby; Edvinsson; Stewart; Kaplan and Norton; Malone; Roos	Raise awareness; top-down	Lack of empirical approach and results
2nd stage	Kaplan and Norton; Guthrie; Abeysekera; Petty; Mourisen	Management, measurement, reporting IC; top-down	Lack of relevance
3rd stage	Guthrie; Mourisen; Dumay; Donnell; Chatzkel	Critical examination of IC research; bottom-up	Time-consuming to collect data
4th stage	Edvinsson; Allee; Dumay; Wasiluk	Establish an IC eco-system; Concerns more human beings and environmental issues	Lack of empirical approach and results

2.2 The development of IC-related definitions and theories

In this section, three IC-related definitions and theories are examined which cover the development of IC definitions and IC term checklists, the complexity and importance of IC as well as understanding IC from a social perspective. The three aspects of IC lay the theoretical foundation for IEIC. The development of IC definitions and IC term checklists help to understand the nature of IC. The complexity and importance of IC offers the instruction of establishing IC repository. To understand IC from a social perspective offers the evidence to rationalize IC sentiment analysis.

2.2.1 IC definitions and IC term checklists

In terms of IC definition and IC taxonomy, IC definition and classification, IC sociological implication as well their shortfalls are specifically examined (see Table 2.2).

Table 2.2 IC definitions and IC taxonomy

Leading researchers	IC definition and IC classification	IC implication	Shortfalls
Petty and Guthrie (2000)	Organizational ("structural") capital; human capital	Economic, managerial, technological	The relevance of IC classification is questioned; no sociological implication
Bernard Marr (2004)	Not specific	Economic, managerial	The relevance of IC classification is questioned; no sociological implication
Guthrie et al. (2012)	Human competencies; structural capital relational capital	Economic, managerial, technological	The IC classification is identified; no sociological implication

Petty and Guthrie (2000b) proposed a definition from the Organization for Economic Co-operation and Development (OECD) which defined IC as "the economic value of two categories of intangible assets of a company". In this definition, there are two classifications, organizational ("structural") capital and human capital. Besides, Petty and Guthrie (2000b) made the distinction between "intangible assets" and "intellectual capital" and emphasized that IC is a subset of intangible assets (rather than the same as intangible asset). "Proprietary software systems, distribution networks, and suppliant chains" are specially mentioned. Also, they make vague concepts clear, e.g. "knowledge management is about the management of the intellectual capital controlled by a company".

Meanwhile, Petty and Guthrie (2000b) found that it is increasingly difficult to understand IC deeply from the traditional accounting perspective given the "increased complexity of classification" and the "different presentations of classification" in all the models. These two challenges are still being actively debated. Through this argument, the implications of IC in economic, managerial and technological aspects are shown. However, the sociological implication of IC development is still not obvious.

Bernard Marr (2004) argued that "we do not believe that we need any more frameworks or classifications"; rather what we need is a clear definition which is defined critically when it is used. IC taxonomy lacks relevance in terms of both academic and practice areas. On one hand, the various understandings in the different contexts of IC classification cause communication barriers for both researchers and practitioners even though the same terms are used. On the other hand, the measuring purposes of IC classification are still ambiguous and controversial. Obviously, Bernard Marr (2004) mainly focused on the accounting and managerial effects that IC has. However, the critical perspective they proposed to understand IC takes us into new terrain.

Guthrie et al. (2012) summarized that IC classifications tend to be identified. They argued that "these components may have different names", but "they basically refer to: human competencies, the knowledge embedded in people; structural capital, the knowledge embedded in the organization and its systems; and relational capital, the knowledge embedded in customers and other relationships external to the organization". The "knowledge" they specially mentioned among the definitions illustrate how strong the managerial and technological implications of IC are.

From these discussions, we can see that IC classification is ultimately identified as human capital, relational capital, and structural capital(Dumay & Garanina, 2013). Meanwhile, Petty and Guthrie (2000) expressed that "intellectual capital is implicated in recent economic, managerial, technological, and sociological developments in a manner previously unknown and largely unforeseen". With a development history of more than 10 years, Guthrie et al.'s (2012) definition began to show the implication of the sociological developments which not only pay attention to the knowledge, but also begin to put the attention on "people".

2.2.2 The Complexity and Importance of IC

In this section, the features of the complexity and importance of IC, how they create values as well as the shortfalls of such research are discussed (see Table 2.3).

Table 2.3 Complexity and Importance of IC

Leading researchers	Features	Value Creation	Shortfalls
Stewart (1997)	Everything & Everyone	Financial value	No sociological implication
Marr, Schiuma, and Neely (2004)	Direct and indirect interrelationship	Financial value	No sociological implication
Cuganesan (2005)	Pluralistic and fluid; narrative	Financial value	No sociological implication
Dumay and Cuganesan (2011)	Liner model; Narrative	Employees' feeling	Have little consideration about sociological implication

Stewart (1997) defines IC as "the sum of everything everybody in a company knows that gives it a competitive edge[...] Intellectual Capital is intellectual material, knowledge, experience, intellectual property, information that can be put to use to create wealth". Even though Rastogi (2003) argued that the former part of this definition is not illustrated clearly, Stewart (1997) did not lead researchers into the IC classification "trap"; rather he shows us richer aspects of IC. Meanwhile, Rastogi (2003) argued through an example that "an organization may have knowledge that can be put to use to generate revenue, but the organization may be incapable of doing so, or 'the intellectual material' may have become ossified or obsolete". From this perspective, it can be seen that the static materials cannot fully represent the dynamic characters of IC. Value in this perspective is the financial wealth of organizations.

Marr et al. (2004) showed the dynamic aspects of IC through drawing an IC value creation map. This map is innovative because it shows us how IC interacts to create value, which is different from the conventional concepts merely showing what IC is and how IC creates value. Besides, Marr et al. (2004) partially demonstrated the complex interrelationship of IC components through mapping the direct and indirect dependencies. The complexity of IC interrelationship compensates the balanced Scorecard's shortcoming that stresses the cause-and-effects of IC. However, Bernard Marr (2004) argued that "there are few people who would doubt that IC is critical for most organizations". Like some other researchers, Marr et al.'s (2004) map also ignored the negative effects that the negative IC information may bring. Besides, the value concerned here is still the financial issue of organizations.

Cuganesan (2005) proposed the method of virtualizing the IC interrelationship by using narratives which describe IC by using the related texts and stories . He argued that narrative is a way to show 'how IC resources transform each other". It is found that IC transforms "often in a pluralistic and fluid manner". Meanwhile, Cuganesan (2005) also found that IC transformation may bring both negative and positive effects on value creation. However, Cuganesan (2005) did not illustrate how to use IC narratives in practice. The value mentioned here is still mainly related to the monetary aspects of organizations.

Based on Cuganesan's (2005) research, Dumay and Cuganesan (2011) make the IC complexity less ambiguous based on much more IC narratives from the common employees. The linear models that Dumay and Cuganesan (2011) used for estimating the strength of association identify a way for IC narratives to be used in practice. The value here is widely investigated by Dumay and Cuganesan (2011). They consider the common employees' feelings through deploying critical analysis methods to collect IC narratives, which is very different from the former studies that merely focus on how to earn and save money for organizations. However, this model at the same time constrains the diversity of IC complexity by one pattern. Besides, the process of collecting IC narrative is also very time-consuming.

To sum up, IC is not merely a combination of intangible assets with classifications but rather it is quite complex and often beyond comprehension. What we can do is to explore the methods to make this complexity be visualized to a larger extent. The method of using narrative shows the potential to explore complexity. What is more, the relationship between IC and value creation is not always positive. Thus, IC should be understood and researched in a critical way (Chatzkel, 2004). However, the process of collecting IC narratives is also very time-consuming. This time-consuming process directly constrains the relevance of the method.

2.2.3 IC from a social perspective

In this section, the features of the IC social implication, how they create value as well as the shortfalls of these research methods are discussed (see Table 2.4)

Table 2.4 IC from a social perspective

Leading researchers	Features	Value Creation	Shortfalls
Allee (2000)	Social citizenship; environmental health	Good quality of people's life	Lack of empirical data
Chatzkel (2004)	IC society level	Healthy living environment	Have a gap between theory and practice
Edvinsson (2013)	IC national level and IC eco-system	Not specific	Lack of empirical data
Dumay (2013)	IC eco-system and economic, social and environmental eco-systems	Good quality of life and healthy living environment	Lack of empirical data

Based on the conventional IC classification (human capital, structural capital and relational capital), Allee (2000) added "social citizenship" and "environmental health" to the IC definition. Allee (2000) stressed that enterprises and organizations need to depend on social systems in which people are seen as social citizens. Thus, a sustainable enterprise depends on the good quality of the lives of its people. The value concerned here involves employees' good quality of life and healthy environment. However, this example shows only the tip of the iceberg of this complex ecological system. The data which Allee (2000) collected were only from the annual reports and hence cannot fully represent people's actual lives and requirements for good quality of life.

Chatzkel (2004) cited Edvinsson's insight about IC's concept at the national level that "we are now moving into the urban design of knowledge cities with a nourishing and healthy impact and attraction on the knowledge workers to settle there". Meanwhile, Chatzkel (2004) argued that "academics and practitioners need to seek out all relevant perspectives and take them into account”. IC here is still a concept which calls for more practice and empirical proof. However, he does not propose the actual methods.

Based on the proposed insight (Chatzkel, 2004), Edvinsson (2013) outlined a global level to research and apply IC is surely needed. At this wider level, the most important element is people and organization’s talent. Edvinsson (2013) expressed that “talent will be the connector in bridging new IC alliances, creating a strategic and wide-ranging intangible capacity, with impact and societal well-being”. He argues that “IC is not a zero sum system, but rather an exponential growth ecosystem. This is due to the IC multiplier effect whereby human capital is needed to leverage relational capital and structural capital”. Furthermore, Edvinsson (2013) even argued that “without human capital, neither can work, nor be utilized”. However, the lack of empirical data makes the ecosystem ambiguous, and meanwhile there are still no new techniques to visualize this complex system.

Dumay (2013) supported Edvinsson’s (2013) insight about the IC sociological implications. Then he asserted the stage of IC eco-system advents. Dumay (2013) presented what the IC eco-system concentrates on is to build strong “economic, social and environmental eco-systems, where healthy organizations can flourish” rather than to build merely strong organizations. Meanwhile, Dumay (2013) argued that IC researchers should stop IC preaching; rather the researchers should go into the field to develop IC practice. However, the empirical data are still insufficient.

To conclude, The IC eco-system opens a wider window to understand IC from the social perspective. But, how the IC eco-system makes economic development more sustainable which considers the good quality of people's life and a healthy living environment need further research (Edvinsson, 2013; Wasiluk, 2013). However, lacking empirical data make IC eco-system still remain at the concept development stage. Therefore, it is necessary to take a few steps backward to collect more empirical data to prevent IC research going back to the old "traps" (Dumay, 2012).

2.2.4 Summary

Based on discussing the development of IC-related definition and theory, it is clear to see that IC is never just the simple classification of HC, RC, and SC; rather it is the complex web which is weaved by them. Meanwhile, IC has its own quality which requires understanding value creation in a critical way. Besides, the implication of social aspects brings a new insight of IC relevance. The mission of studying IC development is important enough to get more empirical data to unveil the often ambiguous picture. Thus, it is an opportunity to examine the research methods to meet the new challenges of IC research and practical application.

2.3 Evolution of major IC research methods to study IC

This section mainly analyzes the evolution of major IC research methods in terms of theory issues, sample issues, research methods used as well as their shortfalls (see Table 2.5).

Table 2.5 Evolution of the major research methods to study IC

Leading researchers	Theory issues	Sample issues	Research methods used	Shortfalls
Petty and Guthrie (2000)	No consistent theory to understand IC	Large sample	Interview, case study, questionnaire, survey of annual reports, focus groups.	Practical application is ignored; challenge of large sample is ignored
Bernard Marr (2004)	Call for testing theories	Large and longitudinal sample	Quantitative and large sample	Challenge of large sample is ignored; lack of longitudinal in-depth case studies
Guthrie et al. (2012)	Call for study which links theory with empirics	Bottom-up	"Commentary/normative/policy", "survey/questionnaire and case field study/interviews", "content analysis of annual reports and historical analysis" as well as "a combination of interview, surveys and case studies	

One of the most important review papers which critically analyzes the development of IC research methods is Petty and Guthrie (2000b). They concluded that "there is no generally accepted theoretical model for understanding IC" and the empirical research purposes mainly "focused on intellectual capital statements, intellectual capital frameworks and measuring and reporting on intellectual capital". Meanwhile, they point out that the methods used in the IC research are various: interview, case study, questionnaire, survey of annual reports and focus groups. Among these methods, Petty and Guthrie (2000b) proposed that a strategy which combines "interviews" and "questionnaires" is most popularly used. Besides, larger sample sizes are involved in this combination. But Petty and Guthrie (2000b) did not offer critique and specific guidance, which makes it ambiguous regarding how the two techniques supplement each other to answer the research questions. Besides, Petty and Guthrie (2000b) merely paid attention to the academic issues; practical application was ignored.

Bernard Marr (2004) summarized that more empirical research of IC measurement and management is needed. Besides, Bernard Marr (2004) argued that what is more important is to test the theories rigorously by using appropriate research methods rather than illustrate the theoretical nature and build new theories. Bernard Marr (2004) pointed out that the classical combination of "quantitative and large sample" cannot allow us to "really understand some of the idiosyncrasies of IC". As for the appropriate research methods, Bernard Marr (2004) proposed two strategies. The first one is that "we need both large samples as well as longitudinal in-depth case studies". Secondly, the multidisciplinary and cross-functional knowledge exchange should be enhanced. However, the same as Petty and Guthrie (2000b), Bernard Marr (2004) did not explain how to derive a cross-disciplinary view of IC. Meanwhile, they do not realize that the coping with large and longitudinal samples is very time-consuming and labor-intensive.

Guthrie et al. (2012) found that the empirical work had undergone a steady increase. Even though this trend is encouraging, they complement this that "there is a danger of over-dependence on empirical studies unsupported by theoretical underpinning". Guthrie et al.

(2012) concluded that the most commonly used research methods include "commentary/normative/policy", "survey/questionnaire and case field study/interviews", "content analysis of annual reports and historical analysis" as well as "a combination of interview, surveys and case studies". Meanwhile, Guthrie et al. (2012) pointed out that these research methods result in "a growing number of theoretical studies published, whilst there are fewer articles on theoretical/empirical studies that link theory with empirics". Guthrie et al. (2012) explained that the failure to convert IC theory into practice is caused by "a concentration of top-down ostensive research instead of bottom-up performative research" (see Dumay, 2009b, 2009c; Mouritsen, 2006). Here, Guthrie et al. (2012) led a direction to use the "bottom-up performative research" to indicate the methods being used.

To sum up, nowadays, IC research requires a larger sample, longer period that the sample can cover as well as more bottom-up IC research. Thus the capability to cope with a large volume of IC empirical data is the new challenge. Meanwhile, identifying any practical application which can bridge theory and practice is another new challenge for them.

2.4 Manual content analysis in coping with IC-related information

One of the most popularly used methods in terms of coping with IC-related information is content analysis (Dumay & Cai, 2014a; Guthrie et al., 2012). The first and most widely cited and representative of these articles is Guthrie, Petty, Yongvanich, and Ricceri (2004) who are the most widely cited content analysis paper. They argue that CA "is a method in need of further refinement and development if advances are to be made in the field of ICD". Meanwhile, they also specially address specific research method issues such as the use of annual reports, unit of analysis, data capture, reliability and validity, quality of disclosure, and the limitations of CA. One legacy of Guthrie et al. (2004) is the argument that annual reports should be used as the main data source for CA research of IC disclosures based on previous social and environmental reporting research. However, while arguments are justified or using annual reports no critique is offered. One legacy of Guthrie et al. (2004, p. 286) is their framework of 18 IC elements for use as a CA coding instrument (see also Guthrie & Petty, 2000). However, there is no detailed discussion of how the final elements were eventually determined; rather only the original sources of the coding instrument are outlined. In particular, they argued for a consistent coding framework, the counting of paragraphs rather than sentences or words and argued against the coding of pictures as being too subjective.

In 10 years, based on Guthrie et al.'s (2004, p. 286) research, Dumay and Cai (2014b) tracked the development of content analysis by analyzing 110 papers which used CA as a research methodology for inquiring into IC reporting. These papers cover from 1999 to 2013 and all of them had an academic impact. However, Dumay and Cai (2014b) did not hold a positive view any more on using manual content analysis to examine ICR because "there was little knowledge about the pattern of IC disclosure in annual reports and other possible intellectual forms". In terms of research methods issues, Dumay and Cai (2014b) re-analyzed the issues of IC index, unit of analysis as well as the quality of IC reporting. The utilization of these content analysis methods did not follow Krippendorff's (2013) standard guidelines. This improper utilization resulted in the declining reliability of manual content analysis in identifying IC. Positively, Dumay and Cai (2014a) and Dumay and Cai (2014b) noted that computer-aided CA has great potential for coping with large volumes of data. However, they found that only eight out of the 110 articles "used computer-aided analysis technique and this appears to remain as an underutilised and underdeveloped research technique".

2.5 Computer-aided content analysis methods in extracting IC-related information

This section is divided into two parts. In the first part, the development of computer-aided content analysis is reviewed. In the second part, a critical analysis of these techniques is conducted.

2.5.1 The development of computer-aided content analysis

This section will critically analyze the eight articles which used computer-aided content analysis technique to examine IC reporting to ascertain how they are used and developed. The summarization is shown in Appendix A.

Bontis (2003) is the first research which applied computer-aided CA to examine IC reporting. Electronic searching aims to extract IC-keywords from 11,000 to examine the level of IC reporting in Canada. A list of IC terminologies (see Table 2.7) was used as the IC disclosure reference which was compiled based on the most popular literature as summarized by the researchers from the World Congress. Every term is searched individually in the database. However, it is surprising that seven out of 39 IC terms are searched. Besides, most IC terms searched are only reported once in each annual report.

Based on the low level of IC reporting, Bontis (2003) concluded that intellectual capital reporting does not raise the awareness in Canadian companies. In the end, Bontis (2003) suggested that more annual reports from other geographical locations and other years can be used. However, Bontis (2003) did not question if negative results are caused by an inappropriate IC terms checklist. Neither do they argue if the IC terms can offer rich enough contexts to identify IC-related information. Meanwhile, they do not mention the quality of IC reporting which is one of the most important aspects of reporting.

Table 2.6 List of IC terminologies of Bontis (2003)

Business Knowledge	Employee Productivity	Intellectual Property
Company Reputation	Employee Skill	Intellectual Resources
Competitive Intelligence	Employee Value	KM
Corporate Learning	Expert Teams	Knowledge Assets
Corporate University	Human Assets	Knowledge Sharing
Cultural Diversity	Human Capital	Knowledge Stock
Customer Capital	Human Value	Management Quality
Customer Knowledge	IC	Organizational Culture
Economic Value Added	Information Systems	Organizational Learning
Employee Expertise	Intellectual Assets	Relational Capital
Employee Know-how	Intellectual Capital	Structural Capital
Employee Knowledge	Intellectual Material	Supplier Knowledge

Vergauwen and Van Alem (2005) deployed the same electronic search as Bontis (2003) to investigate the IC reporting in the Netherlands, France and Germany; 178 annual reports from three years were searched and 23 out 38 IC terms were found in the annual reports. Based on the results of searching, Vergauwen and Van Alem (2005) found that IC reporting varies between different countries. Meanwhile, Vergauwen and Van Alem (2005) argued that the search result is affected by accounting regulations and the tension created by the IC information. Similarly, Vergauwen and Van Alem (2005) had the same problems using electronic searching.

An electronic search was conducted (Vergauwen, Bollen, & Oirbans, 2007) with 20 annual reports to measure the level of IC reporting in Sweden, Britain and Denmark: 103 terms were compiled based on the former literature (Bozzolan, Favotto, & Ricceri, 2003; Brennan, 2001; Goh & Lim, 2004; Vergauwen & Van Alem, 2005) (see Table2.8). In Vergauwen et al.'s (2007) research, human intervention is involved to identify the IC-related information as Vergauwen et al. (2007) found that not all the IC-keywords in the checklist are searched. Among these attributes, RC elements are most often reported and SC elements are the least reported. Finally, Vergauwen et al. (2007) presented that "non-traditional industries seem to report more on RC". Thus, in future research, more attention should be paid to firms from different industries. However, Vergauwen et al. (2007) do not illustrate how the people involved in the process identified the IC context; this causes difficulty for others to replicate the research. Like many others, Vergauwen et al. (2007) ignored the quality of IC reporting.

Table 2.7 List of IC terminologies of Vergauwen et al. (2007)

Internal capital	**External capital**	**Human capital**
Network	Customers	Employees
R&D/research and development	Joint venture	Knowledge
Telecommunication	Brands	Personnel
Patents	Market share	Expertise
Innovation	Partnership	Competence
Leadership	Customer satisfaction	Education
Methodologies	Supply chain	Specialist
Intellectual property	Distribution channels	Employee benefits
Trademarks	Customer loyalty	Know-how

Philosophy	Distribution networks	Employee satisfaction
Management processes	Quality standards	Motivation
Corporate culture	Brand development	Career development
Information systems	Customer knowledge	Empowerment
Knowledge sharing	Customer base	Human capital
Knowledge resources	Business collaboration	Intelligence
IC	Customer recognition	Employee expertise
Electronic data interchange	Supplies Knowledge	Employee skill
Trade secrets	Customer capital	Human value
Management focus	Competitive intelligence	Expert team
Corporate university	Company reputation	Employee value
Software systems	Customer retention	Flextime
Cultural diversity	Customer turnover rates	Brain power
Proprietary process	Favourable contracts	Human assets
Intellectual assets	Corporate image	Expert network
Business knowledge	Franchising agreement	Employee retention
Technological processes	Licensing agreement	Value added statements
Value added	Financial contacts	Union activity
Soft assets		Training programmes
Operating systems		Vocational qualifications
Operating software		Work-related competence
Organizational learning		Work-related knowledge
Organizational culture		
Management quality		
Knowledge stock		
Knowledge assets		
Intellectual resources		
Intellectual material		
Economic value added		
Corporate learning		
Product development cycle		
New product success rate		
Research projects		
Networking systems		
Infrastructural assets		
Copyrights		

Kamath (2008) also used electronic searching to study the extent of voluntary intellectual capital reporting in India’s emerging information, communication and technology sector.

Thirty annual reports were searched. The operation process and IC term list used are similar to that in Bontis (2003). Kamath (2008) found that the extent of IC reporting is not high because there were only 13 out of 39 terms searched in the annual reports of these firms. Besides, 10 out of 32 firms did not report any terms in the list. Finally, Kamath (2008) suggested that more annual reports can be involved to "study the comparative picture across sectors and countries". Still, Kamath (2008) ignored the same problems that the simple IC terms in the electronic searching failed to recognize the IC-related information which is not expressed by the specific IC terms. Again, the quality of IC was not mentioned as well.

WordStat, Version 5.0, developed by Provalis Research, Montreal, Canada is deployed to perform the content analysis (Sonnier, Carson, & Carson, 2008) of 141 US companies' annual reports of three consecutive years. The purpose of this research is to assess intellectual capital disclosure levels. According to the identification results, IC reporting increased from 2000 to 2004. However, this computer-aided method is also a simple IC terms electronic search. A list of 121 words and phrases (see part of IC terms list in Table 2.9) were chosen based on the prior literature. Sonnier et al. (2008) expressed that the reason for choosing this literature is that it concerns the different words used between the academic and business worlds. It is the first computer-aided CA paper that addressed the difference between academic and practical IC terms. However, Sonnier et al. (2008) did not list all the IC terms. On the other hand, there are also some flaws. Firstly, this approach cannot capture the IC context in a word or phrase. Secondly, the software cannot distinguish IC information which is just planned. Thirdly, the IC words or phrases will be missed if they are not listed in the 121 compiled IC index. Last but not least, the software cannot recognize the IC underlying meaning behind the key words and phrases. These are all very important missions that content analysis should have encompassed.

Table 2.8 List of IC terminologies of Sonnier et al. (2008)

Internal capital	**External capital**	**Human capital**
Intellectual property	Brands	Employee loyalty
Patents	Company reputation	Staff competencies
Copyrights	Customer loyalty	Staff turnover
Trade secrets	Customer relations	Teamwork
Management philosophy	Customer satisfaction	Training
Management style	Distribution channels	Workforce
Organization learning	Franchising agreements	Work-related knowledge
Organizational culture	Licensing agreements	Career development
Organizational structure	Market share	Employee attitude
Structural capital	Repeat business	Employee development
Work processes	Alliances	Employee expertise
Competitive intelligence	Business collaborations	Employee knowledge
Corporate mission	Favourable contracts	Employee productivity
Information systems	Partnerships	Expert teams
Knowledge management	Supplier relations	Human assets
		Human capital
		Employee innovations

Software called PDF-X Change Viewer is used to search 72 Portuguese companies' annual reports (Branco, Delgado, Sá, & Sousa, 2010). The purpose of content analysis is to measure the level of IC reporting, and then the result is used to "evaluate size, industry and time effects on disclosure as well as the effects of ICD on the growth of a company". Only the annual reports' sectors which report IC voluntarily are analyzed. In Branco et al.'s (2010) findings, HC is the most reported IC element rather than the common results which RC is reported the most. Meanwhile, the level of IC reporting fluctuates from 2004 to 2008.

Branco et al. (2010) inferred that differences of IC reporting results may be attributed to the different content analysis method used, different country context, different industry as well as different legislation. This means that annual reports may not report all actual IC used in the organizations. The list of terms originates from Petty and Guthrie (2000a) which contained 24 sub-categories (see Table2.10). The terms/expressions related to intellectual capital were searched (Branco et al., 2010).

Table 2.9 List of IC terminologies of Petty and Guthrie (2000a)

Internal capital	**External capital**	**Human capital**
Intellectual property	Brands	Human resources
Patents	Company reputation	Employee loyalty
Copyrights	Customer	Teamwork
Management philosophy	Customer loyalty	Employee expertise
Corporate culture	Distribution channels	Employee knowledge
Structural capital	Franchising agreements	Entrepreneurial spirit
Work processes	Licensing agreements	Human capital
Information systems	Business collaborations	
Knowledge management		

The software program "concordance" is deployed to investigate the level of IC disclosure in Spain (Oliveras, Gowthorpe, Kasperskaya, & Perramon, 2008). Annual reports from 12 Spanish companies over a three-year period are analyzed. The level of IC reporting is measured by the frequency which each attribute appeared in these sample annual reports. In the end, Oliveras et al. (2008) stressed that the level of IC reporting is not high but the level of IC reporting increased over the 3 years. External capital is the largest reporting element. More years' annual reports are suggested to be analyzed. Oliveras et al. (2008) used Guthrie and Petty's (2000) classification; the keywords and phrases are summarized in

Table 2.11. However, Oliveras et al. (2008) did not discuss the sources of these words and phrases.

Table 2.10 List of IC terminologies of Oliveras et al. (2008)

Acknowledgement of expertise	Human resources	Professionals	Studies
Knowledge	Loyalty	Specialists	Education
Abilities	Expertise	Collaboration with academics	Learning
Skills	Trained managers	Training programmers	Motivation
Talents	Employees	Costs	Teams
Experience	Experts	Projects	Staff
Workforce			

The development of intelligent content analysis methods are still at the early stage; however, besides the automatic function being attractive enough, the great potential in terms of analysis leads to a new terrain. Lock Lee and Guthrie (2010) utilized Factiva's intelligent taxonomy to extract IC-related information from 156 IT firms' business and analyst reports published by external reporters. Lock Lee and Guthrie (2010) found that the level of IC reporting is limited to "newsworthy" elements. Meanwhile, Lock Lee and Guthrie (2010) stressed that one piece of IC information may be attributed to different classifications. Factiva taxonomy is a kind of off-the-shelf tool that assists automatically in classifying IC-related information by using the fixed terms in its database. The practical IC-related information stored in the Factiva data show its great potential in terms of practical application. However, it is not a publicly used method. Besides, human intervention is constantly involved in terms of correcting gross errors and mapping the results with the IC triple models (human capital, structural capital and relational capital).

Thus, the accuracy is difficult to guarantee. The IC terms in this research are listed in Table 2.1.2

Table 2.11 List of IC terminology of Lock Lee and Guthrie (2010)

IC classification equivalence (Guthrie & Petty, 2000)	Factiva intelligent taxonomy terms
Human capital: employees, education, training, work-related knowledge, entrepreneurial spirit	Employee training/development
	Workers' pay
	Labour disputes
	Lay-offs
	Recruitment
	Directors' dealings
	Executive pay
	Management moves
	Intellectual property
	Best practice
Internal capital: intellectual property, management philosophy, corporate culture, management processes, information/networking systems, financial relations	Competitive intelligence
	Corporate process redesign
	Knowledge management
	Supply Chain
	Information technology
	Debt/bond markets
External capital: brands, customers, satisfaction, company names, distribution channels, business collaborations, licensing agreements.	Marketing
	Joint ventures
	Contracts/orders
	Profiles of companies
	Society/community/work

Table 2.12 Extra list of IC terminology of Lin, Huang, Du, and Lin (2012)

Advanced training	Employee behaviour	Work environment
Employee welfare	Ethics	
Retirement programme	Safety	

A computer-aided automatic keywords search was conducted by Lin et al. (2012) on 660 Taiwan public listed companies' annual reports to examine the level of HC reporting. The reporting level was measured by the searched keywords frequency. The results are used to examine the impact of HC reporting on "form performance" and "the effects of knowledge intensity and organizational size". The complete list of keywords consisted of 40 items from Vergauwen et al.'s (2007) IC term list and local authority (see Table 2.13). In the end, Lin et al. (2012) suggested that other techniques can also be applied to reduce the bias as not all annual reports are text searchable. Meanwhile, Lin et al. (2012) argued that the double-edged effect of HC reporting and time lag effect of HC reporting should be further researched.

To conclude, it is not difficult to see that the results produced by computer-aided content analysis also add little knowledge to IC research (see Dumay & Cai, 2014a). Meanwhile, computer-aided content analysis also shares the same difficulties as for manual content analysis in terms of data of sources, IC index, and unit of analysis as well as the quality of disclosures.

2.5.2 The critical analysis of the development of computer-aided content analysis

This section critically analyzes the main step of computer-aided content analysis including IC index, unit of analysis as well as the quality of IC analysis.

(1) IC index

Among the various IC indexes used, most of these indexes are compiled based on former literature reviews. However, Indra Abeysekera and Guthrie (2005) found that "not a single annual report has explicitly made reference to the term 'intellectual capital'". Also, Carrington and Tayles (2012) argued that actually IC in the practical terms is not used the same as in the academic world. Thus, the words are all academic words which bring trouble in identifying IC that is not expressed by academic words. Surprisingly, Laurence Lock Lee and Guthrie (2010) used practical words to audit the IC information extracted in the Factiva. However, this index also has a limitation that it is difficult to be replicated as Factiva is a commercial product, so its taxonomy cannot be accessed nor used in the public domain. There is no IC index combining both the academic and practical IC terms. As one of the most important components of automatic methods, the relevance of IC index has a big impact on the effectiveness of this method.

Three methods use Bontis' (2003) IC index. The other six methods deploy a different index. Here is a contradiction that the same IC index can help to compare different contexts. However, until now, except the HC, RC as well as SC are suitable for nearly all contexts (Dumay, 2013), no consistent IC index is that is suitable for every context. Meanwhile, it is obvious to see that a static IC index begins to lose its relevance. Meanwhile, every IC index is compiled according to individual classification of IC which is judged according to the context

(2) Unit of analysis

Nearly all the methods use the IC terms to search the texts. Even Laurence Lock Lee and Guthrie (2010) did not specify any units of measurements. There is the same problem that the IC terms that express IC may not be used in these actual reports. Besides, IC terms are too narrow/shallow in terms of providing sufficient contexts for one complete IC element. Therefore, there is no doubt that the unit of analysis has a great effect on the accuracy of extraction results. Oliveras et al. (2008) are the only to use “sentence” as the context unit to read texts. The context unit deployed enables the analyst to analyzee the sentence in a more accurate dimension. Based on keywords searching, Vergauwen et al. (2007) suggested that human interventions are involved in identifying the hints which are revealed by keywords and sentences. However, they did not discuss the specific context unit that they used.

(3) The news-tenor of IC information analysis

It is a great pity that, despite the fact that IC theory shows the great relevance of the news-tenor of IC negative and positive aspects of IC information, none of these computer-aided content analysis methods touch on the issues of the news-tenor. However, news-tenor is quite important for examining the relationship between the disclosure of IC and the cumulative abnormal return of a firm’s share price (Dumay & Tull, 2007). Obviously, this important point related to IC is ignored.

To conclude, computer-aided content analysis is quite attractive for coping with large amounts of IC data. However, the problem is that these methods can only perform the function of “extraction” rather than “analysis”. It is the same for IC research with Krippendorff’s (2013) argument that there are limitations to analyze text, particularly as computers analyze high volumes of texts literally rather than in context, making “the computer a natural but also difficult ally of the content analyst”. Therefore, it is very necessary to clarify the ability of the method when it is used. Meanwhile, the methods of computer-aided techniques still have a lot of room for improvement.

2.6 Summary

To summarize, computer-aided extraction enables IC information to be extracted from a large volume of data. This method is done by matching the academic and practical IC terms, which can greatly save time and labour. These new sources of data achieve a wider dimension which formerly mainly focused on annual reports. However, the limitation of these methods is that many processes still need much human intervention which not only cost time and money, but also inject much personal biases. One of the biggest problems is that all of them fail to extract the inter-relationship of IC components. Meanwhile, the process of recognizing the news-tenor of IC is not sustainable and clearly used. Thus, the current method to extract IC-related information is losing its popularity.

Based on these gaps, the following research question is presented;

How to proceed to discuss ways to increase the relevance of IC extraction for value creation in knowledge-intensive organizations?

Chapter 3. Methodology

The use of computer-aided content analysis sheds light on the potential of automatic extraction of IC-related information. However, the current extraction techniques remain at the level of keywords searching (Bontis, 2003; Uergauwen and Van Alem, 2005; Uergauwen et al.,2007; Kamath, 2008; Sonnier et al. 2008; oliveras et al. 2008; Branco et al., 2010). They all have some shortfalls that should be improved. Firstly, the IC index in keywords searching consists of academic IC terms. The current keywords searching methods fail to search the IC-related information which is expressed by practical IC terms and words which are not included in the checklist. Secondly, the unit of analysis is confined to one fixed type. Therefore, the current methods fail to extract IC-related information which is expressed in a complex way. Thirdly, the current methods only have the ability to extract IC-related information; these techniques fail to analyze the news-tenor of IC information which is expressed in a critical way.

To overcome these difficulties, a method for intelligent extraction of intellectual capital (IEIC) (see Figure 3.1) is proposed based on the current IC development theory which has already gained empirical valuation. There are four tiers in representing the architecture of IEIC. The first tier is the external data tier. The sources of external data are from banks' annual reports, IC reports, IC news as well IC journals. Based on this tier, the IC knowledge repository contains IC-keywords and phrases as well as keyword patterns. Using the data from IC journals, an IC academic knowledge repository is constructed. The data from IC companies' reports and IC news enable an IC practical knowledge repository.

In the third tier, IC extraction and IC sentiment analysis are applied. By using the IC-related keywords and patterns, IC extraction is used to extract the IC-related information from company reports, media space, etc. After the IC-related information is extracted, IC sentiment analysis is conducted to analyze the IC new-tenor. The positive and negative aspects of IC increase the relevance of extracted IC information. The final tier is the output tier. There are two parts. One is the summary of IC news which is analyzed through the sentiment analysis. The second part is the IC-related information. This information will be stored into the IC repository for extracting new IC-related information.

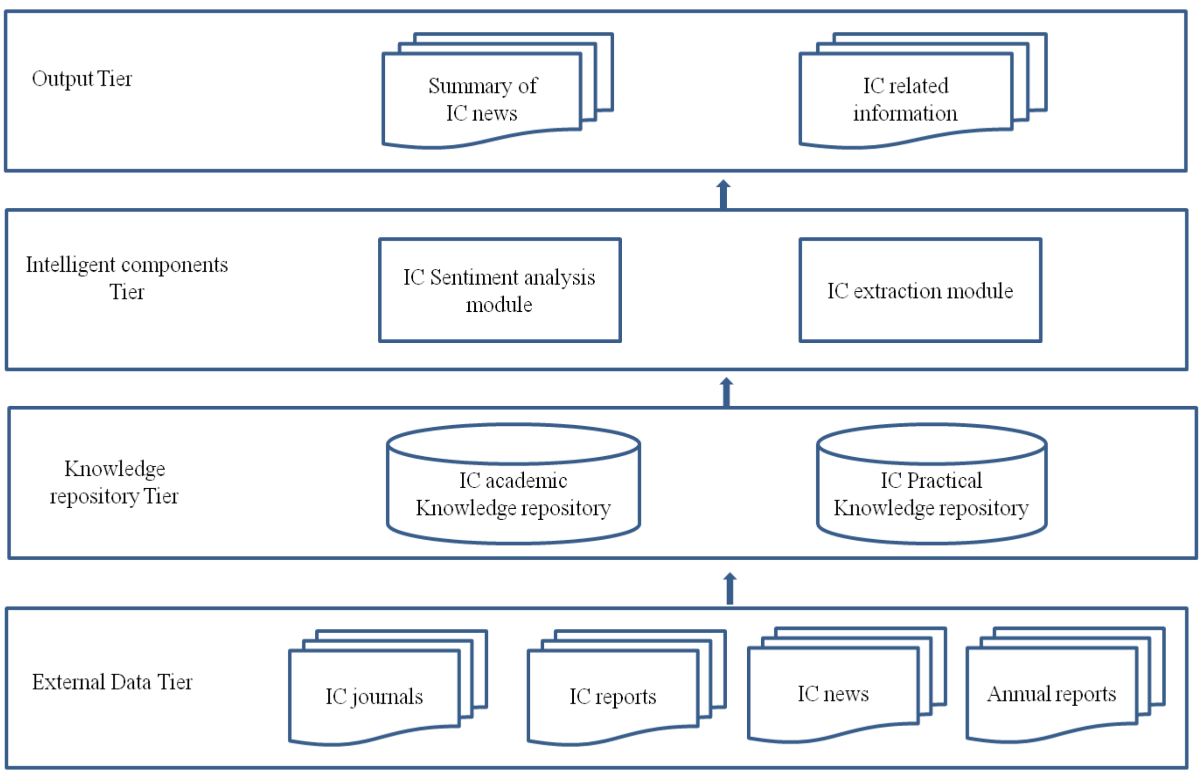

Figure 3.1: Architecture of IEIC

3.1 The Process of Repository Construction

This section mainly introduces the steps of constructing the IC academic knowledge and IC practical knowledge repository. To cope with the problems of failing to extract practical IC-related information, a practical IC repository is used to supplement the academic IC repository. The combined repository greatly helps to extract IC-related information which is described in practical and/or academic terms.

In the process of implementing IC knowledge repositories, with the help of IC extraction methods (KBICE), the IC knowledge repository is constructed and enlarged based on by parsing the same extracted IC elements and keywords manually.

3.1.1 The process of constructing an IC academic knowledge repository

Firstly, an IC academic knowledge repository is constructed (the whole process is shown in Figure 3.2). IC analysts identify IC elements by reviewing academic papers which mainly focus on using CA as a research methodology for understanding IC reporting. The author uses keywords “IC” and “CA” to search Google Scholar. The reason for using Google Scholar is that papers can be selected according to the citation number (Dumay & Cai, 2014b; Dumay & Cai, 2015). Thus papers selected have already had academic impact. Then the IC analysts review and analyze all IC indexes which contain the IC terms (see Table 3.1). Petty and Guthrie (2000a) provided the classical IC indexes. Based on a consistent IC framework (Dumay & Garanina, 2013) (see Table 3.2), an IC checklist is then compiled according to all the IC indexes that are mentioned. Then the IC analysts conduct a thorough analysis of the database to extract the keywords and keyword patterns which express and describe the IC elements. The reason why they choose these keywords is because keywords are the smallest unit (Dumay & Cai, 2014a) to help construct the context of IC elements and these contexts help one to understand IC in the context (see Table 3.2). Meanwhile, in this research, the classifications of IC defined are not used because in practical unitization, the definition of the classification is very much by context in which they are used.

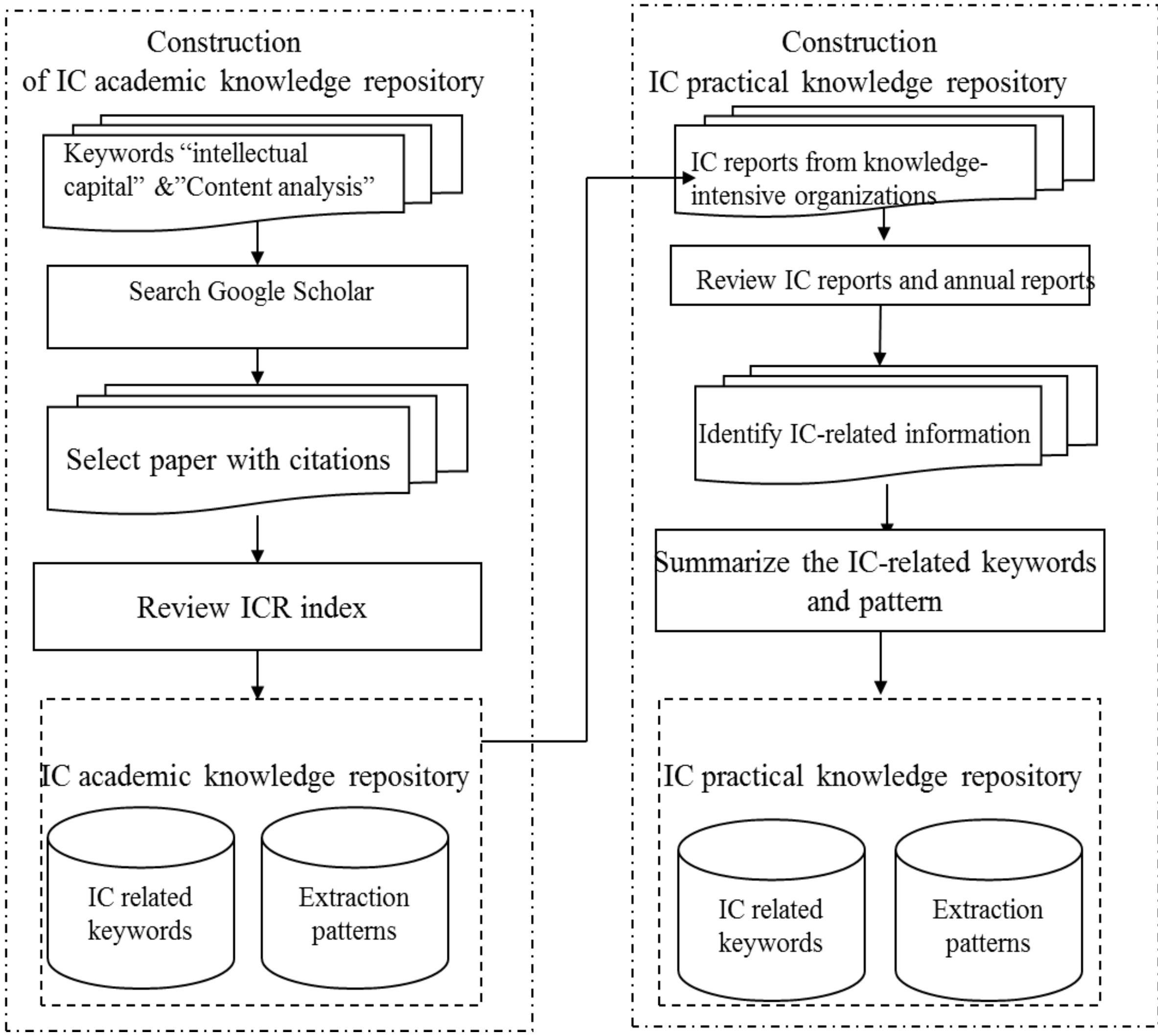

Figure 3.2: Construction of the IC academic and practical knowledge repositories

Table 3.1 List of IC terminologies identified in Petty and Guthrie (2000a)

Structural capital	**Relational capital**	**Human capital**
Intellectual property	Brands	Human resources
Patents	Company reputation	Employee loyalty
Copyrights	Customer	Teamwork
Management philosophy	Customer loyalty	Employee expertise
Corporate culture	Distribution channels	Employee knowledge
Structural capital	Franchising agreements	Entrepreneurial spirit
Work processes	Licensing agreements	Human capital
Information systems	Business collaborations	
Knowledge management		

Table 3.2 IC classifications and their definitions (Guthrie et al. ,2012)

HC	The knowledge embedded in people
SC	The knowledge embedded in the organization and its system
RC	The knowledge embedded in customers and other relationships external to the organization

3.1.2 The process of constructing an IC practical knowledge repository

An IC practical knowledge repository needs to be constructed (see Figure 3.2) considering the gap between how an IC element is expressed in an academic expression and mentioned in a practical application. An analyst who finished her final year project on IC reporting helped to construct an IC practical repository based on the actual IC reports from the knowledge-intensive organizations. Then I further refined the repository based on my 2 years of research experience. The main extraction work is based on the IC academic knowledge repository. However, it is merely used as a reference and offers a context to extract the actual utilization words. The IC academic knowledge repository offers the IC term index to search the IC reports. After collecting the IC reports from one industry, the

IC analyst began to review the actual reports and corresponding annual reports (see Figure 3.3). Then the analyst explored the IC models used in the reports, and the ICR index is also reviewed thoroughly (see Figure 3.4). Meanwhile, the analyst also reviewed the IC information reported in the annual reports to find the keywords and phrase patterns. The review context varies from sentence to paragraph. The work of extracting the pattern forms the basis of the inter-relationship of the IC components (Dumay & Cuganesan, 2011). As a result, a piece of IC may not belong entirely to a single classification but rather patterns cover different IC components. For example, there is one paragraph from the report of the ATP bank stated "Special contribution terms apply to two thirds of all public-sector employees, corresponding to 18-20% of all contribution-paying members. Of this group, 61% are women". "Employees" and "women" are the IC-related keywords. Meanwhile, "employees" and "women" can be used as the patterns to extract IC information.

Figure 3.3: OeNB IC reports

Indicators of Investment in Knowledge-Based Capital

	Unit	2006	2007	2008	Strategic target
Staff structure					
Full-time equivalent staff (year-end)	number	931.7	917.5	968.2	↘
Fluctuation rate	%	0.6	1.2	1.2	→
University graduates	%	34.3	35.6	41.3	↗
Gender management					
Ratio of women to men in staff	%	38 : 62	39 : 61	39 : 61	↗
Ratio of women to men in the specialist career track	%	23 : 77	24 : 76	32 : 68	↗
Ratio of women to men in management positions	%	20 : 80	17 : 83	18 : 82	↗
Flexible working arrangements					
Part-time employees	%	7,1	7,8	7,8	→
Staff in teleworking scheme	%	4,5	4,8	4,2	→
Staff on sabbatical	number	4	5	7	→
Knowledge acquisition					
Training days per employee (annual average)	days	4.2	4.1	4.2	→
Training participation rate	%	57.2	61.3	64.5	→
Cost of training and education per employee	EUR	2,332	2,330	2,280	↗
Staff who completed the development center program	number	–	24	24	→
Internal job rotations	number	41	44	37	↗
Completed and certified training courses (in-service)	number	13	9	20	→
Working visits to national and international organizations (external job rotations)	number	29	26	37	↗
Interns from universities of applied sciences	number	28	26	30	→
Average Intranet visits (daily average)	number	–	–	4,426	↗
E-learning	number	–	–	265	↗
Innovations					
Staff working on innovative projects (core business areas + IT)	%	6.2	7.5	6.0	→
Staff suggestions for improvements[3]	number	48	182	92	→

Figure 3.4: Part of OeNB's IC report's content

3.2. The process of IC extraction （KBICE）

The knowledge-based intellectual capital extraction method is used to increase the efficiency and accuracy of extracting IC-related information that utilizes IC-keywords/phrases and combination patterns. On the one hand, this method accurately extracts the IC-related information which is expressed in a complex way. On the other hand, this method also helps to supplement the IC knowledge repository with new found terms. The second module is the repository of IC news which is searched with the help of the IC knowledge repository. The third module is the IC extraction and sentimental analysis.

In IC extraction, the knowledge-based intellectual capital extraction （KBICE）algorithm is designed to extract IC-related information. To conduct this extraction, firstly, IC-related key phrases and their pattern-based rules are pre-coded into the knowledge repository to extract IC-related information manually by applying the initial setting rules. Meanwhile, the repository is further expanded as new cases are being reviewed. With the help of IC extraction, IC sentimental analysis is conducted to analyze the news-tenor. The final module is the module of output which consists of two parts including the IC-related information and the quality of the IC information.

After constructing the IC knowledge repository (see Appendix B), the knowledge-based intellectual capital extraction (KBICE) algorithm is used to carry out IC extraction; this algorithm is based on a combination of the technologies of computational linguistics and artificial intelligence (AI) for automatic extraction of IC-related information from large volume of documents.

Furthermore, Case-based reasoning (CBR) approach and Rule-based reasoning (RBR) are also utilized to enhance the learning capability of the automatic extraction of IC-related information. As shown in Figure 3.5, the knowledge repository is supplemented constantly by the IC-related key phrases and pattern-based rules. These key phrases and keywords are extracted according to the initial setting which is summarized by the IC analysts who reviewed the training data when constructing the knowledge repository.

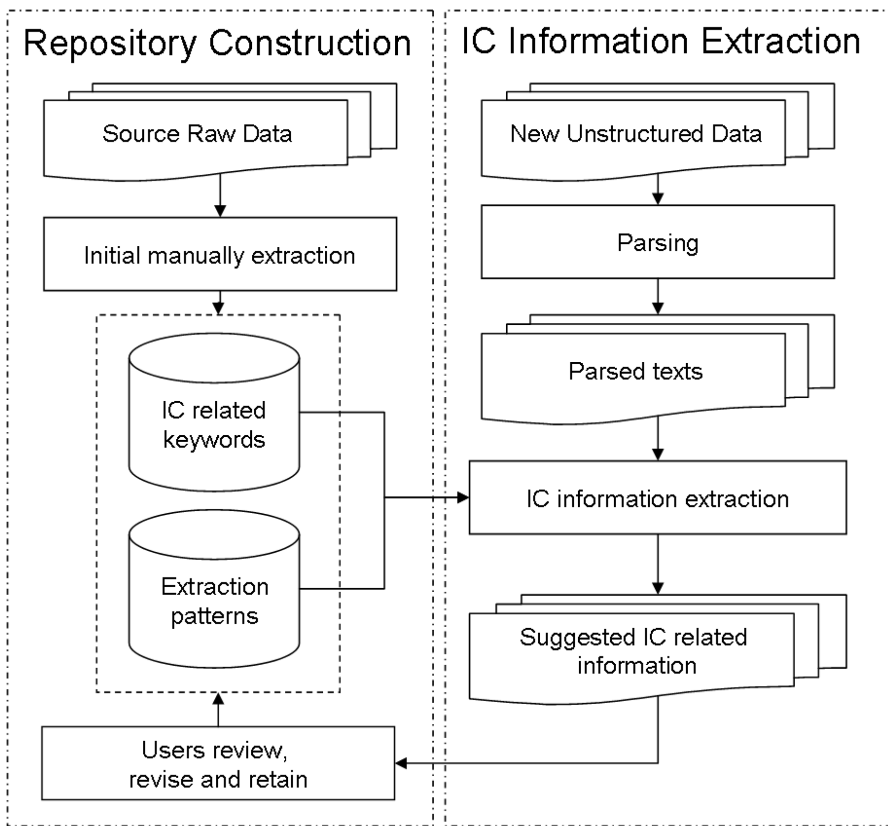

Figure 3.5: Schematic diagram of Knowledge-based Intellectual Capital Extraction

In the process of IC information extraction, a new document is firstly preprocessed through an ad hoc sentence boundary detection algorithm. This algorithm is according to the regular expressions (for instance a new line character, full stop and question mark) so as to identify the paragraphs and sentences. The sentence is further divided into tokens by regular expressions such as space, comma, parentheses, etc. The tokens are then tagged with their parts of speech (POS) using a POS tagger developed by Schmid (1994). Meanwhile, the heuristics rules and certain specific data types are assigned to the tokens (e.g. date, year and numbers). The flow of the parsing process is shown in Figure 3.6

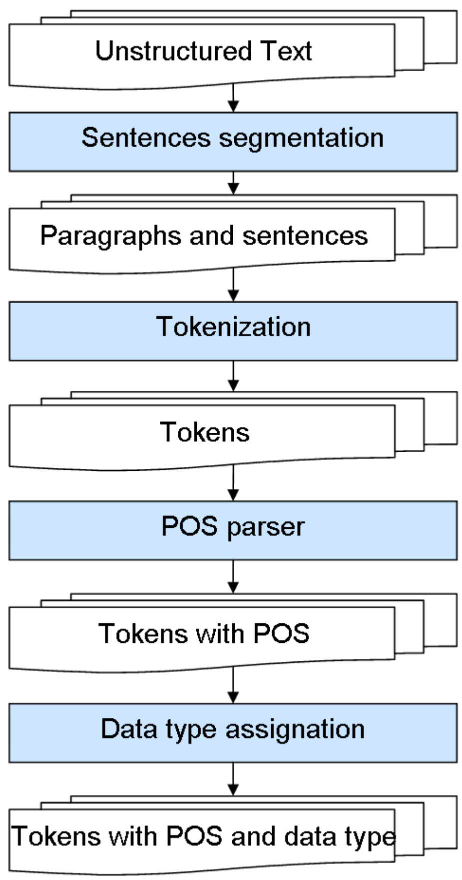

Figure 3.6: Process flow of the parsing process

The parsed text is then delivered to the IC information extraction. The pattern-based rules with IC-related key phrases in the knowledge repository are applied to the parsed texts. IC-related keywords and patterns constitute a rule. The IC information extraction algorithm is presented in Figures 3.7 and 3.8. Two examples of rules are shown as follows:

1. **Rule No: 1**
   Pattern: a total of [NUMBER]
   Key Phrase: Press conference

2. **Rule No: 2**
   Pattern: with a number of [NUMBERS]
   Key Phrases: members participating

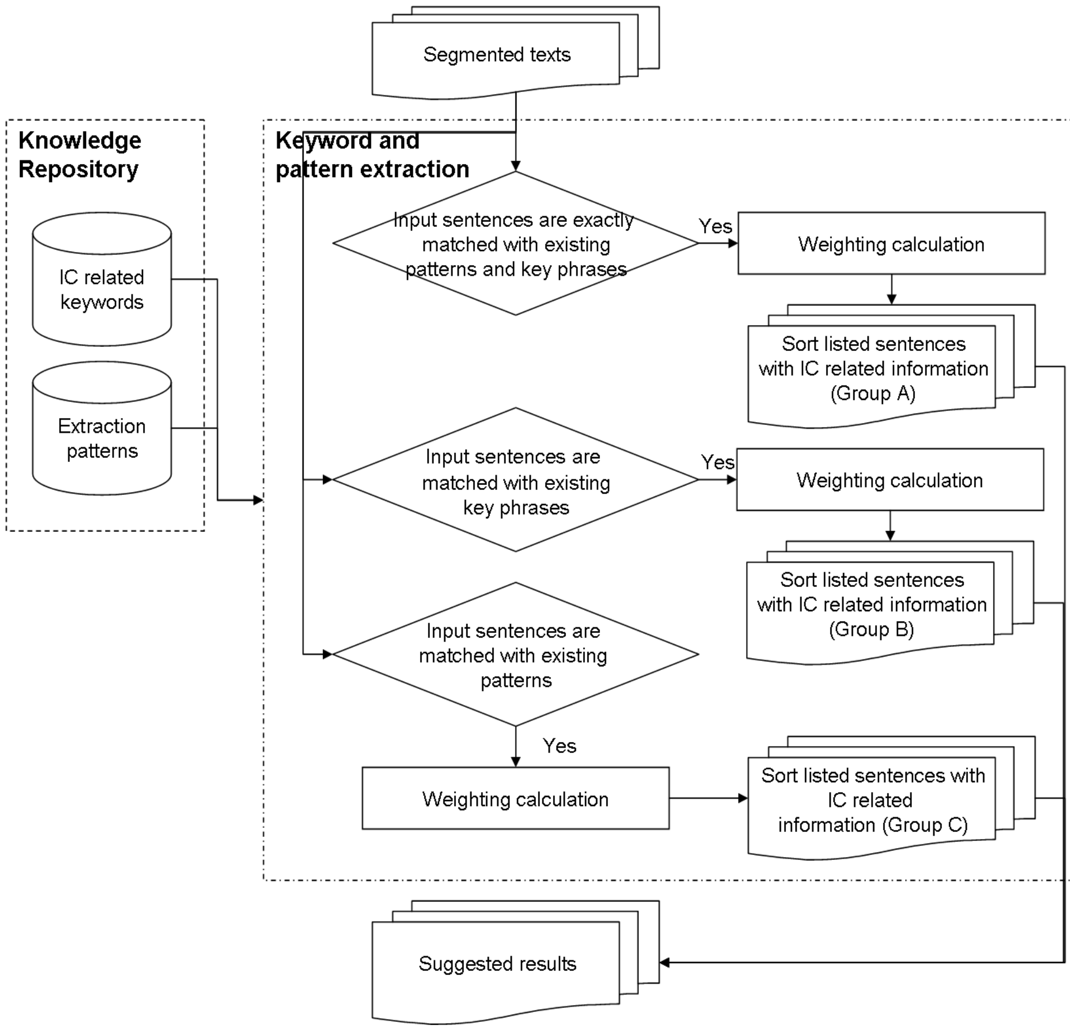

Figure 3.7: Schematic diagram of the IC information extraction algorithm

```
1.      Divide the input text into paragraphs
2.      Divide the paragraphs into sentences and assign to SENTENCE_LIST
3.      Divide the sentences into tokens and assign to SENTENCE_LIST
4.      Tag the tokens with POS and assign to SENTENCE_LIST
5.      Tag the tokens with Data type and assign to SENTENCE_LIST
6.      Get the rules from database and assign to RULE_LIST
7.      For I From 1 To NumOf(SENTENCE_LIST)
7.1         For J From I To NumOf(RULE_LIST)
7.2             If SENTECT_LIST[I] exactly matched with RULE_LIST[J] Then
7.3                 Add SENTECT_LIST[I] To RESULT_A
7.4             End If
7.5             If SENTECT_LIST[I] matched with RULE_LIST[J].KEYPHRASE Then
7.6                 Add SENTECT_LIST[I] To RESULT_B
7.7             End If
7.8             If SENTECT_LIST[I] matched with RULE_LIST[J].PATTERN Then
7.9                 Add SENTECT_LIST[I] To RESULT_C
7.10            End If
7.11        End For Loop J
7.12    End For Loop I
8.      Export RESULT_A, RESULT_B, RESULT_C
```

Figure 3.8: An IC information extraction algorithm

As shown in Figures 3.7 and 3.8, three groups are defined. They are Group A, Group B and Group C respectively. Once the extracted sentences are exactly matched with one or more rules, it will be defined as Group A. The number of rules determines the weights given to the sentences. They are shown in Equations (3.1) and (3.2) and sorted by weights in descending order. Group A are the most relevant IC-related information.

$$Weight_i^A = \sum_j^n Score(Sentence_i, Rule_j) \tag{3.1}$$

$$Score(Sentence_i, Rule_j) = \begin{cases} 1 & \text{if } Sentence_i \text{ is exactly matched with } Rule_j \\ 0 & \text{if } Sentence_i \text{ is not exactly matched with } Rule_j \end{cases} \tag{3.2}$$

where n is the total number of rules.

For Group B and Group C, sentences are extracted if the sentences are partially matched with one or more of the existing rules. For Group B, sentences are extracted if the sentences are matched with at least one of the existing rules' key phrases. For Group C, sentences are extracted if the sentences are matched with at least one of the existing rules' patterns. The weight of each of the extracted sentences is determined by the number of matched rules and the similarity among the extracted sentences. It is argued that the more rules that exist the more accurate the results will be. The calculation of weight of the *i*-th paragraph is shown in Equations (3.3) to (3.5).

$$Weight_i^C = \sum_j^n Score(Sentence_i, Rule_j) \tag{3.3}$$

$$Score(Sentence_i, Rule_j) = \sum_k^m Similarity(Sentence_i, Sentence_k, Rule_j) \tag{3.4}$$

$$Similarity(Sentence_i, Sentence_k, Rule_j) = \begin{cases} \dfrac{|\mathbf{v}_i \cap \mathbf{v}_k|}{|\mathbf{v}_i \cup \mathbf{v}_k|} & \text{if } Sentence_i \text{ and } Sentence_k \text{ is partially matched with } Rule_j \\ 0 & \text{if } Sentence_i \text{ or } Sentence_k \text{ is not matched with } Rule_j \end{cases} \tag{3.5}$$

where n is the total number of rules, m is the total number of sentences, $\mathbf{v}_i = (v_{i1}, v_{i2}, \ldots)$ are the set of frequencies of words of the *i*-th sentence, $|\mathbf{v}_i \cap \mathbf{v}_k|$ is the size of the intersection between the sets $\mathbf{v}_i$ and $\mathbf{v}_k$, and $|\mathbf{v}_i \cup \mathbf{v}_k|$ is the size of the union between the sets $\mathbf{v}_i$ and $\mathbf{v}_k$.

Hence, the sentences of Group B and Group C are also sorted in descending order by the weights of the extracted sentences. Then these sentences will be used as suggestions for users to review. The new IC-related key words/phrases and patterns are put into the knowledge repository for reuse. Two examples are shown below:

1. **Original Content:** *"Apart from restructuring a number of communication tools — such as the school information kit or the quarterly publication Focus on Austria, which is to be reissued as Monetary Policy & the Economy — the OeNB published more than 160 press releases on various institutional, economics and statistics topics in 2003. Moreover, the OeNB held a total of 26 press conferences, attracting about 700 representatives of both the Austrian and international media."*

   Matched IC-related keywords: *press conference*

   Matched Pattern: Press releases; press conference

   Extracted Sentences: *"the OeNB published more than 160 press releases on various institutional, economics and statistics topics in 2003. Moreover, the OeNB held a total of 26 press conferences, attracting about 700 representatives of both the Austrian and international media."*

   Potential Key Phrase: representatives; *international* media

2. **Original Content:** *"Business Centres. At year end, Bankinter had a network of 102 Business Centres, distributed among the 13 regional headquarters of which the Bank is composed, with a total headcount of 240 employees. These Centres are mainly located in industrial estates where there is a high density of medium-sized enterprises, to whom*

they give specialised and close-at-hand commercial attention."

Matched IC-related Key Phrase: Business centre

Matched pattern: Business centre; headquarters

Extracted Sentences: *"At year end, Bankinter had a network of 102 Business Centres, distributed among the 13 regional headquarters of which the Bank is composed."*

Potential Pattern: a network of[NUMBER]

3. **Original Content:** *"Sound Expert Knowledge — A Major Success Factor The OeNBs staff members make a decisive contribution to fulfilling Austria's demanding tasks within the ESCB/Euro system and to coping with a business environment that is changing at an ever greater pace. Adjusted for employees on secondment or leave (such as maternity and parental leave), the average number of staff working in the OeNBs core business areas (expressed in full-time equivalents) came to 957.3 in 2004. 37.3% of OeNB staff held university degrees, which represented a year-on year rise by five percentage points."*

   Matched IC-related Key Phrase: business areas

   Matched pattern: employees; staff; business areas

   Extracted Sentences: *"Adjusted for employees on secondment or leave (such as maternity and parental leave), the average number of staff working in the OeNBs core business areas (expressed in full-time equivalents) came to 957.3 in 2004."*

   Potential Pattern: average number of staff working

3.3 The process of IC sentiment analysis

The last module is the sentimental analysis of the news-tenor of IC structural data. This part is proposed according to Dumay and Tull's (2007) theory of good news increasing share price and bad news decreasing share price. Meanwhile, the sentimental analysis applied also helps to analyze the feelings of human beings, which helps to establish the IC-eco-system.

The process of sentiment analysis is operated mainly based on Hu and Liu's (2004) work. The first reason is that there are already 2,224 (on 8/18/14) citations of their work. This means that their work has been applied to more applications in larger domains than other methods. Secondly, Hu and Liu (2004) offer a clear and general framework to conduct the sentiment analysis, which is appropriate for the IC sentiment analysis that is developing at the very early stage with little research focusing on how to classify IC news-tenor. There are four steps involved (see Figure 3.9): 1) Selecting IC news which has opinion orientations as the experiment sources. 2) Mining IC-related sentences by matching the IC-keywords with the repository. 3) Identifying opinion sentences and deciding whether each opinion sentence is positive or negative. 4) Summarizing the results. In the process of the third step, firstly, the opinion word which is near to the IC-keywords is extracted and then the opinion orientation is identified. Thirdly, the opinion words are stored to enhance the learning ability. In the following, the combination of the IC-keywords or keyword pattern, and the opinion is used to identify an opinion sentence. Then the orientation of the opinion sentence is predicted (see Figure 3.9). The method of predicting the presentation of an opinion is to use the dominant orientation. The dominant orientation means if there are more positive opinion keywords, the sentence is regarded as positive; if there are more negative opinion keywords, the sentence is regarded as negative. If the number of opinion keywords is the same, the sentence will be recorded as neutral. In the end, a brief summary is conducted. For example (see Figure 3.10), there is one piece of IC news: “UK consumers are showing an increasing preference to deal with their finances online as they believe it to be quicker, more convenient and cheaper.” In this sentence, the IC-keywords are “consumers” and “online”. The IC-keyword pattern is “finances online”. The opinion words which are near to the IC-keywords and keyword pattern are “increasing”, “quicker”, “convenient” as well as “cheaper”, respectively. To summarize, these opinion words are positive, so the orientation of the example sentence is positive.

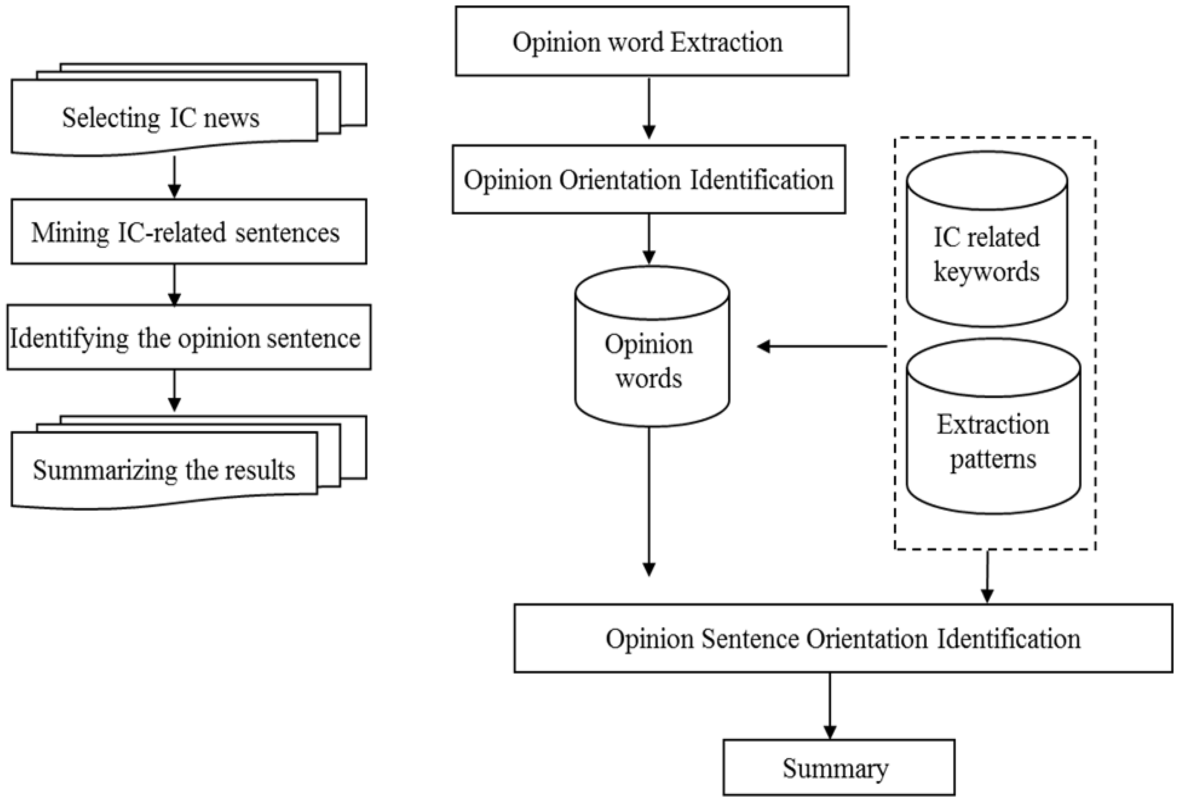

Figure 3.9: Process of IC sentimental analysis

Example:

IC news extraction: UK consumers are showing an increasing preference to deal with their finances online as they believe it to be quicker, more convenient and cheaper.

IC keywords: consumers; online

IC keywords pattern: finances online

Opinion sentence: an increasing preference; be quicker, more convenient and cheaper

Opinion words: increasing; quicker; convenient; cheaper

Opinion orientation: Increasing----Positive

Quicker----Positive

Convenient----Positive

Cheaper----Positive

Opinion sentence orientation: an increasing preference----Positive

be quicker, more convenient and cheaper----Positive

Summary: Positive

Figure 3.10: Example of IC news sentimental analyses

3.4 Summary

To summarize, there are three modules making up IEIC. In the first module, the process of constructing an IC knowledge repository is introduced. Based on the IC knowledge repository, knowledge-based IC extraction assisted by an algorithm is designed to increase the efficiency of extracting IC-related information. Meanwhile, the IC knowledge repository is enlarged by the application of KBICE. After IC-related information is extracted, IC sentiment analysis helps to judge the news-tenor of the information.

Chapter 4. Implementation and Experimental Verification of the IEIC

This section contains two main parts. One is the implementation of IEIC which mainly shows the results of relying on the IC academic knowledge repository and IC practical repository to support IC extraction. For the IC practical repository, the banking industry is chosen as it is the industry offering the largest number of IC reports. The other part of this chapter is the evaluation of KBICE algorithm and sentiment analysis. In this part, there are two sections: one is the process of evaluation which mainly introduces how the evaluation is conducted. The other section is the results and discussion.

4.1 To construct the IC-related repository

This section mainly shows the content of the IC knowledge repository. Based on the methods described in Chapter 3, an IC academic knowledge repository and IC practical knowledge repository are constructed by the two IC analysts through manual reading of reports.

4.1.1. IC academic knowledge repository

One hundred and ten papers (Dumay & Cai, 2014b) are extracted which mainly focus on using CA as a research methodology for understanding IC reporting. After finishing the review of all papers, 112 IC term checklists are analyzed (see Table 4.1).

There are 45 IC indexes that are miscellaneous or not explicitly mentioned. Among these IC indexes, there are two special IC indexes involving IC practical terms. Oliveira, Rodrigues, and Craig (2006) deployed an index based on analyzing the Management

Report and Chairman's Letter in annual reports. This kind of ICD index is constructed based on the practical use of specific texts rather than just analyzing entire annual reports. Another innovative example is Jindal and Kumar's (2012) ICD index, developed from the IC literature, whereby they filtered the index based on the Indian company Infosys' annual report.

Table 4.1 shows that ICD indexes used in ICD CA research are varied and derived from many different sources with the most common being those based on the work of Guthrie, Petty, Ferrier, and Wells (1999) (25), Sveiby (1997) (12), Guthrie and Petty (2000) (9), Bontis (2003) (7), Guthrie et al. (2004) (5) and Bukh et al. (2005) (4). As Petty and Guthrie (2000a, p. 245) outline, they "followed the contemporary classification scheme for intangibles derived from Sveiby (1997) intellectual capital framework: internal structures (organizational capital); external structures (customer/relational capital); and employee competence (human capital)". Thus, most articles use Sveiby's (1997) framework in some form or another, and there have been no substantially new ICD indexes that have taken over from what we initially presented in Table 3.2, as this ICD index appears to have been modified rather than replaced.

MERITUM's (2002) model "added a further 15 extra items (not contained in previous indices) after further consideration for the Australian socio-political and economic environment, and healthcare system". There are three IC indexes (Johanson, Koga, Skoog, & Henningsson, 2006; Pedrini, 2007) which are from accounting guidelines and initiatives.

Table 4.1 IC academic term checklist

IC academic term checklist	**Number of papers**
Miscellaneous or not explicitly mentioned	45
Guthrie et al. (1999)	25
K. E. Sveiby (1997)	12
Guthrie and Petty (2000)	9
Bontis (2003)	7
Guthrie et al. (2004)	5
Bukh et al. (2005)	4
MERITUM (2002) model	2
Japanese IC Guideline	1
Indicators from the Global Reporting Initiative (GRI)	1
Schmalenbach Gesellschaft Work Group on Financial Accounting (2002)	1

Based on these IC Indexes, an IC academic elements checklist is established. As the amount of elements in the IC academic repository is huge, only part of the checklist is shown in Tables 4.2 and 4.3

Table 4.2 IC-related phrases

Assortments	**Attitude**	**Access lines**
Business systems	Age of employees	Brand recognition
Business knowledge	Competitive Intelligence	Brand development
Brand name	Corporate Learning	Company reputation
Corporate culture	Career development	Customer relationships
Corporate know-how	Competencies	Citizens' satisfaction with the project
Corporate governance quotient	Community involvement	Distribution channels
Certifications	Development of personnel	External networking systems
Company names	Employee expertise	Franchising agreements
Corporate university	Employees' health and work safety	Initiatives regarding priorities planning activity
Corporate learning	Entrepreneurial skills	Marketing campaigns
Development stage of R&D	Executive compensation plan	Market penetration
Education/training	Expert Teams	Promotion
External information and knowledge	Formal training	Promotional events
holidays	Incitement programmers	Relationship with local community
Information systems	Pension plan	Supply contracts
IT projects	Teamwork	Supporters clubs
IT systems	Training programmes	Research collaborations
Internal process	Union activity	Shop localization
Work processes	Vocational qualification	Website hits
……	……	……

Table 4.3 IC academic keywords

Training			
Education Training	Education		
Employee training	Employee		
Training programmes	Programmes		
Training and in-house activities	Activities		
Training and development	Development		
Training and internal education	Education	Internal	
Training cost		Cost	
Formal training		Formal	
Customer education training	Education		Customer
…	…	…	…
System			
Inventory in systems	Inventory		
IT systems	IT		
Internal communication system	Internal communication		
Internal networking systems	Internal	Networking	
Information systems	Information		
Operating systems	Operating		
Software systems	Software		
Streamlining internal control systems	Internal	Control	Streamlining
Transparent systems	Transparent		
Business systems	Business		
System of knowledge sharing and team working	Knowledge sharing	Team working	
External networking systems	Networking		System
…	…	…	…

4.1.2 IC practical knowledge repository

After a significant amount of searching work, 22 IC reports with their corresponding annual reports were found published by four European banks (ATP, BBVA, Bankinter, OeNB) from 1994 to 2011 (see Table 4.4). BBVA and BankInter only have the IC metrics and IC models which mainly express IC by single IC classification. ATP and OeNB also use narratives to express IC. Meanwhile, the overall situation of these IC reports is listed in Table 4.4. However, it is shown that different companies have their own IC model.

Table 4.4 IC reports of banks

Company	**IC reports components**	**IC Model**	
ATP	IC metrics; Narrative; IC model	Balanced Scorecard	Clients
			Staff
			Business procedures
			Finance focus
BBVA	IC metrics; IC model	Intellectual capital measurement model	Human capital (HC)
			Structural capital (SC)
			Relational capital (RC)
BankInter	IC metrics; IC model	Intellectual capital measurement model	Human capital (HC)
			Structural capital (SC)
			Relational capital (RC)
OeNB	IC metrics; Narrative; IC model	Process-oriented model	Human capital (HC)
			Structural capital (SC)
			Relational capital (RC)
			Innovation capital (InC)

Based on these IC metrics in the reports and the IC-related key phrases in the annual reports, an IC practical elements checklist is established. As the amount of elements in the IC practical repository is huge, only part of the checklist is shown in Table 4.5.

Table 4.5 IC-related key phrases

Corporate laptops/PC	Gender distribution	Awards or public recognitions received
Network capability	Employees acting as mentors	Press releases
Process time	Contributions to knowledge communities	Customer service system
Suggestion box	Certified additional qualification courses	Press conferences
Development and maintenance of intranet/database content	Staff who completed the development centre programme	Customer knowledge
Internal communication channels	Employees who have received training	Customer loyalty
Employee accessing the internet daily	Training participation	Community support
Internal job rotation	Individual mentoring	Projects to promote social development and accessibility for handicapped staff
Strategic plans	Absence due to occupational hazards	Service Awards
New business system	Leadership	Enquiries and incidents reported to Telephone Banking handled by e-mail/ Telephone Banking staff
Development projects	Motivation	Lectures organized/delivered
Easy-to-understand language	College graduates	Alliances and collaboration projects with academic and research institutions
Prize-winning quality projects and initiatives	Permanent employees	Experts participating in the forums organized by the firm
Management information	Employees with advanced English language skills	Persons attending cash authentication courses
…	…	…

After reviewing the 104 IC-related sentences and paragraphs identified by the IC academic repository and part of IC practical repository, IC extraction patterns are formulated and stored. The sources of these patterns are the IC keywords in the IC knowledge repository. The principle of formulating IC extraction patterns is to check how these IC keywords may appear within multiple sentences. Some of these extraction patterns are shown in Table 4.6.

Table 4.6 IC extraction patterns

IC-related key phrases	**Key phrases in the actual reports**	**Patterns**
Employee Diversity	Women %	Employee, women
Customer communication (e.g. Telephone Platform, Internet)	Member receiving a pension	Receive a pension
Customer Relationships	Number of attendees in membership meetings	Held open meetings, meetings, number of participants
Customer Relationships	ATP-Pensioners receiving a current pension	Pensioners, receiving a current pension
Customer communication (e.g. Telephone Platform, Internet)	Number of attendees in membership meetings	Member, participated, membership meetings
Customers, Image and Stakeholders	ATP-Pensioners receiving a current pension	Member paid contributions
Customer communication (e.g. Telephone Platform, Internet)	Number of attendees in membership meetings	Members attended, information meetings
Quality and customer satisfaction	Proportion of satisfied/very satisfied attendees in membership meetings	Attendees, satisfied
Quality of Pension summaries	Customers find the summary to be written in easy-to-understand	Members find the pension

	language	summaries, written in easy-to-understand language
Staff Turnover	Departed employees (staff turnover)	Staff reduction
Employee flexibility	Number of employees	People, working
Employee flexibility	Number of employees	Originators of ideas
Employee flexibility	Number of employees	Total, staff
Corporation and networks	Number of SME Management Centres	SME segment, number of these branches
Corporation and networks	Number of SME Management Centres	New centre specializing in SMEs, total, such centre
Training (Competence development)	Total number of courses taught	Workforce, receive training, total, courses
Training (Competence development)	Average number of training hours per employee trained	Training hours, per employee trained
Corporation and networks	Number of SME Management Centres	Total number of SME branches
RC ungrouped	Transactions through channels other than Branch Network as % of total Bank transactions	%, transaction, remote channel
Quality Management and Improvements	Number of quality projects and initiatives carried out	Quality, improvement project, started

4.2 Experimental evaluation for the KBICE algorithm

In the section of experimental evaluation, the process of KBICE evaluation and sentimental analysis are described in detail. Then the results of the evaluation are discussed.

4.2.1 The process of KBICE evaluation

Based on the IC-related keywords, phrases as well-formed patterns in the IC knowledge repository, 937 paragraphs are used as the experimental data which contain both IC-related information and non-IC-related information. These paragraphs are extracted from 110 pages of annual reports with 41,706 words. An IC analyst helps to identify the right answer and the training set of rules. As shown in Figure 4.1, a leave-one-out method is chosen as the validation method for testing the learning algorithm's accuracy of predicting data which were not trained on. Leave-one-out is a method of not wasting data. Two kinds of data comprise the whole sample. They are validation data and training data. The validation data are just one single case. The training data are the data left (see Figure 4.1). The process does not stop until each case is used once as the validation data. The detailed process of the leave-one-out method is shown in Figure 4.2.

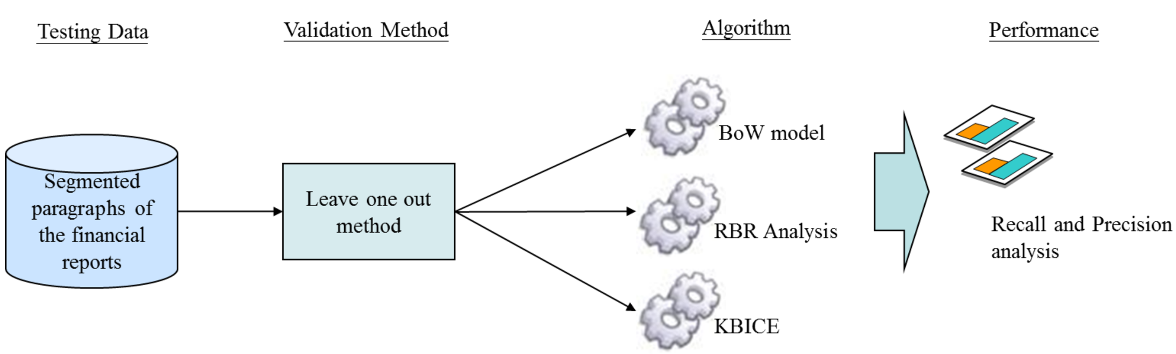

Figure 4.1: Experimental setup for measuring the performance of the KBICE algorithm

For i = 1 to N (where N is the number of training cases)

Temporarily remove the i-th case from the training set

Train the learning algorithm on the remaining N - 1 points

Test the removed case and note the precision and recall

End For

Calculate the overall precision, recall, and F-measure over all N cases

Figure 4.2: Algorithm of leave-one-out method for the evaluation of the KBICE algorithm

To test the scalability of KBICE, various numbers of training cases are conducted as experiments. A paragraph is chosen as the unit of analysis to identify a single case. Case numbers vary from 100 to 900 cases with a 100 case increment. The reason for an incremental number of cases is used to test the effect of increasing number of training cases to the results. There are a number of rules used in the KBICE analysis, we used equal weighting for each rules due to simplicity. The results of testing the KBICE are also compared with RBR analysis and the bag-of-words (BoW) model. For RBR analysis, paragraphs are extracted once the paragraphs are exactly matched with at least one of the rules as described in section 3.3.

The BoW model is a commonly used method in document classification. Here are two simple examples:

Mary likes to eat apples. Linda likes to too.

Linda also likes to cook apple pies.

Based on these two texts, a dictionary is constructed as:

{"Mary": 1, "likes": 2, "to": 3, "eat": 4, "apples": 5, "Linda: 6, "too": 7, "also": 8, "cook": 9, "pies ": 10}

It has 10 different words. And using the indexes of the dictionary, each document is represented by a 10-entry vector:

[1, 2, 1, 1, 1, 1, 1, 0, 0, 0]

[1, 1, 1, 0, 1, 0, 0, 1, 1, 1]

in which each entry of the vectors refers to the count of the corresponding entry in the dictionary. There are already some successful applications of testing this method, for instance, email filtering (Sivic & Zisserman, 2009). No classification is needed, which decreases the subjectivity greatly in this process. During the process of this evaluation, all the paragraphs are represented according to the BoW model and two categories are chosen (i.e. paragraphs with IC-related information and paragraphs without IC-related information). Tested paragraphs are compared with the training paragraphs by using cosine similarity. The cosine similarity is calculated as Equation (4.1).

$$\text{Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}||\mathbf{B}|} \tag{4.1}$$

where A and B are the term frequency vectors of the paragraphs. A tested paragraph is classified by comparison with each of the training paragraphs. The most relevant training paragraphs will be chosen as the testing paragraph. Precision, Recall and F-measure are used for the performance measurement through comparing the IC-related paragraphs suggested by the RBR, BoW and KBICE with the IC-related paragraphs selected by the IC analyst. The Precision, Recall and F-measure are defined as Equations (4.2), (4.3) and (4.4), respectively.

$$\text{Precision} = \frac{|\{relevant\ paragraphs\} \cap \{retrieved\ paragraphs\}|}{|\{retrieved\ paragraphs\}|} \tag{4.2}$$

$$\text{Recall} = \frac{|\{relevant\ paragraphs\} \cap \{retrieved\ paragraphs\}|}{|\{all\ retrieved\ paragraphs\}|} \tag{4.3}$$

$$\text{F} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.4}$$

4.2.2 Results and discussions of the evaluation

The results of the evaluations of different case numbers are shown in Figures 4.3 to 4.5 and Table 4.7. The results show that KBICE is a more effective way than RBR and Bow in terms of recall, precision as well as F-measure. KBICE's recall increased when the amount of training data increases, When comparing with RBR and BoW, KBICE has a higher recall, and at the same time, the precision rate remain similar to RBR and BoW.

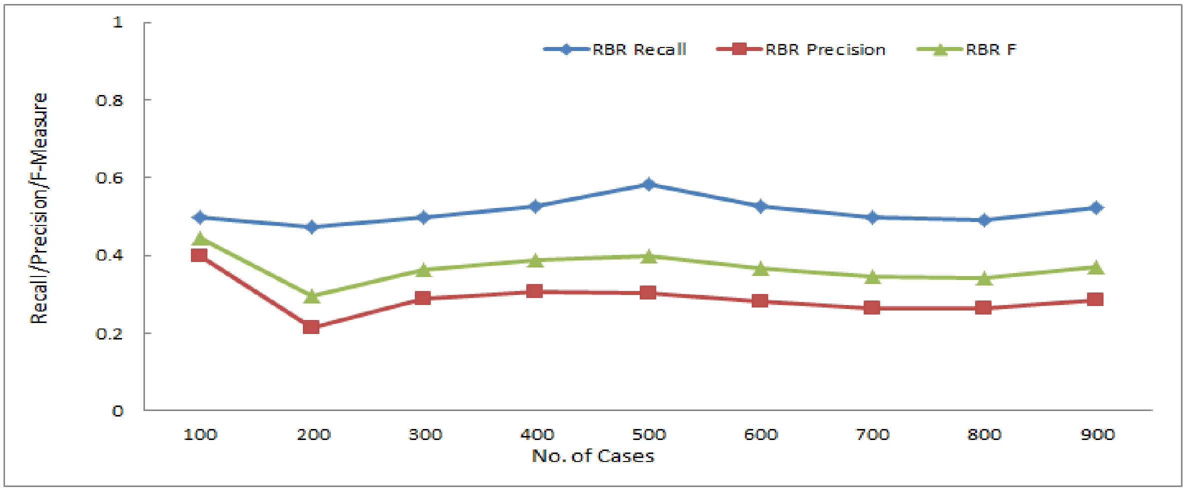

Figure 4.3: Evaluation result of RBR

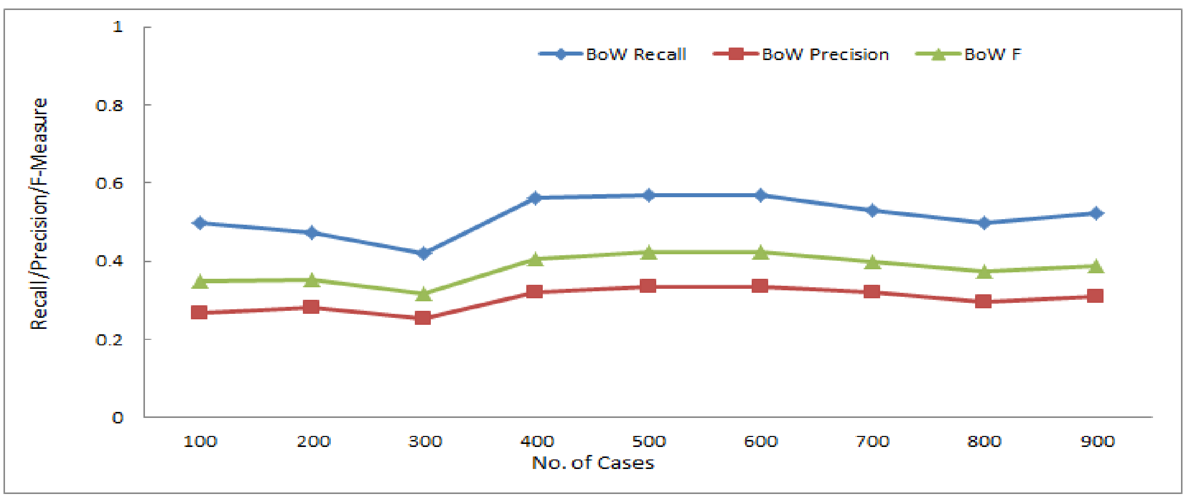

Figure 4.4: Evaluation result of the BoW model

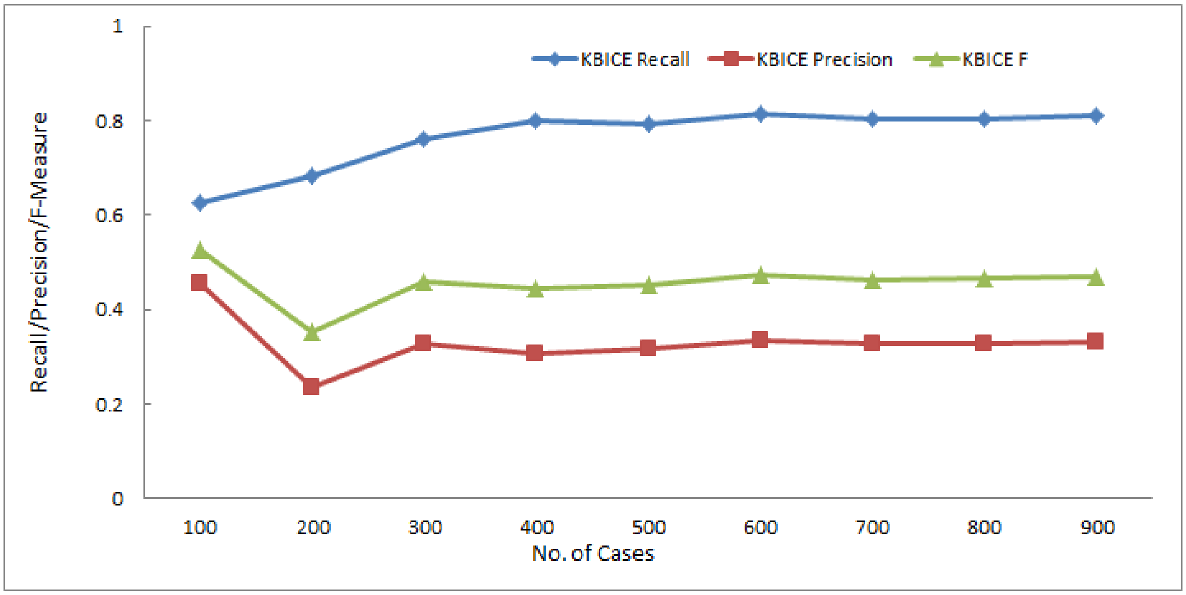

Figure 4.5: Evaluation result of KBICE

Table 4.7: Evaluation result

	No. of Cases								
	100	**200**	**300**	**400**	**500**	**600**	**700**	**800**	**900**
RBR Recall	0.500	0.474	0.500	0.527	0.583	0.529	0.500	0.491	0.524
RBR Precision	0.400	0.214	0.288	0.309	0.304	0.280	0.266	0.263	0.286
RBR F	0.444	0.295	0.365	0.389	0.400	0.367	0.347	0.343	0.370
BoW Recall	0.500	0.474	0.421	0.564	0.569	0.569	0.529	0.500	0.524
BoW Precision	0.267	0.281	0.254	0.320	0.336	0.336	0.320	0.298	0.310
BoW F	0.348	0.353	0.317	0.408	0.423	0.423	0.399	0.373	0.389
KBICE Recall	0.625	0.684	0.763	0.800	0.792	0.816	0.804	0.804	0.810
KBICE Precision	0.455	0.236	0.330	0.308	0.318	0.335	0.327	0.327	0.330
KBICE F	0.526	0.351	0.460	0.444	0.454	0.475	0.465	0.465	0.469

A summary of evaluation results of all approaches with 937 cases is provided in Table 4.8. From the results, we can see that the proposed method KBICE outperforms the other three methodologies in all the three measures. High recall rate always accompany with low precision rate. It is interesting to note that KBICE has a higher recall rate (around 0.8) against RBR and BoW (around 0.5), but KBICE continues to maintain a higher precision rate (around 0.33) against RBR (around 0.29) and BoW (around 0.3).

Table 4.8: Evaluation results of all approaches with 937 cases

Approaches	**Recall**	**Precision**	**F-measure**
RBR	0.531	0.289	0.374
BoW	0.515	0.306	0.384
KBICE	0.815	0.330	0.470

Different statistical analysis and a series of student's t-tests are conducted to compare the recall, precision and F-measure of different case numbers (i.e. 100, 200, 300, 400, 500, 600, 700, 800 and 900) in RBR and BoW with those in the proposed method, respectively. The results are shown in Table 4.9. Based on the results, one can see that nearly all the measures of the proposed approach are significantly better than those of the other approaches ($p < 0.05$). The exception is the comparison of precision in BoW with the proposed approach. The result shows that the precision of the proposed approach is not significantly better than that of BoW ($p = 0.136$). However, the mean of precision of the proposed approach is better than that of BoW.

Table 4.9: Statistical analysis among different approaches

Measure type	Analysis	RBR	BoW	KBICE
Recall	Mean	0.514211873	0.516720376	0.766349317
	Variance	0.001004415	0.002441958	0.004455682
	Standard Deviation	0.031692507	0.04941617	0.066750898
	Standard Error	0.010564169	0.016472057	0.022250299
	p (t-test)	7.60379E-07	1.10145E-06	
Precision	Mean	0.290000884	0.30231798	0.329505701
	Variance	0.002476143	0.000875173	0.003118895
	Standard Deviation	0.049760859	0.029583322	0.055847067
	Standard Error	0.016586953	0.009861107	0.018615689
	p (t-test)	0.000401184	0.135743416	
F-measure	Mean	0.368896938	0.381343595	0.456688859
	Variance	0.001715798	0.001329782	0.002091208
	Standard Deviation	0.041422187	0.03646618	0.045729732
	Standard Error	0.013807396	0.012155393	0.015243244
	p (t-test)	5.39629E-06	0.001972817	

In conclusion, compared with the other two methods, KBICE has better performance in terms of extracting IC-related information. KBICE is superior in that it uses IC-related inference rather than the pure lexico-synatactic patterns to search the information. Meanwhile, the proposed methods can also suggest the new patterns and IC-keywords/phrases thereby maintaining the dynamics of the IC knowledge repository.

4.3 Experimental evaluation of sentimental analysis

4.3.1 The process of evaluation

The new source of data originates from the online IC news of HSBC banks within one month. By parsing the IC-keywords and keyword patterns with the IC knowledge repository, 38 pieces of news from 2014.05.6 to 2014.05.30 are selected and summarized.

Compared to the traditional data sources, news is one of the external data sources which have clear opinion orientation (Lock Lee & Guthrie, 2010). By using the opinion keywords attached to IC-keywords and keyword pattern rules in the repository, the IC sentimental content is extracted. Precision and recall are used for the performance measurement of IC sentimental content extraction. The precision and recall are defined in Eqs. 4.2-4.4, respectively. Then the accuracy of opinion prediction of opinion sentences is evaluated.

4.3.2 Results and discussions of the evaluation

Table 4.10 shows the evaluation results of the other two procedures: opinion sentence extraction and sentence orientation prediction in the news and product feature review. The recall of opinion sentence extraction with the IC news is nearly 95.2% (see Table 4.10), which is higher than Hu and Liu's (2004) average recall of 69.3% (see Table 4.11). The precision of opinion sentence extraction is 68.9%, which is also higher than Hu and Liu's (2004) average recall of 64.2%. Compared to Hu and Liu's (2004) sentence orientation accuracy, the sentence orientation accuracy is higher, at 84.2% and 95.7%, respectively. This results show that Hu and Liu's (2004) method can be applied to the IC sentiment analysis.

Table 4.10: Results of opinion sentence extraction and sentence orientation prediction

Text name	Opinion sentence extraction		Sentence orientation accuracy
	Recall	Precision	
IC news	95.2%	68.9%	95.7%

Table 4.11 Hu and Liu's (2004) average results of opinion sentence extraction and sentence orientation prediction

Text name	Opinion sentence extraction		Sentence orientation accuracy
	Recall	Precision	
Product feature review	69.3%	64.2%	84.2%

To conclude, this evaluation has to choose IC news as the sample, as there are few negative aspects published in the annual reports and IC reports. It is also a very important finding of this research. The comparison shows that IC sentimental analysis also can use Hu and Liu's (2004) opinion extraction and prediction method. Meanwhile, apart from adjectives that can help to analyze the sentiment of IC information, some verbs also can predict the opinion orientation to some extent. For example, "facilitate" and "increase" can predict positive IC information. "Disagree" and "reduce" can help to predict negative IC information. To enrich the repository and enhance the algorithms for accuracy, verbs which can predict the opinion orientation can be selected manually. However, the reliability of this method should be tested before it can be put to practical usage.

4.4 Summary

This chapter mainly implements the method proposed by Chapter 3. Firstly, the IC knowledge repository is constructed based on 110 research papers and 22 IC reports with their corresponding annual reports. Then IC-related keywords and keyword patterns are stored in the repository. Then an experimental evaluation of KBICE and sentiment analysis is conducted by using the IC knowledge repository. Both the experiments of KBICE and sentiment analysis show very positive results.

Chapter 5. Discussion

The chapter covers two aspects: the advantages and limitations of IEIC. In terms of advantage of IEIC, three aspects are presented including the advantages of the IC knowledge repository, the advantages of IC extraction as well as the advantages of IC sentimental analysis. In terms of limitations of IEIC, there are also three aspects: too much human intervention involved, samples are from a single industry as well as the inadequacy of data sources for experimentation.

5.1 Aadvantages of IEIC

(1) The process of IC knowledge repository construction

In the IC practical world, there are large volumes of extraction patterns analyzed and stored, which reduce the gap of lacking of academic extraction pattern. These extraction pattern increases the accuracy of extracting IC information. Meanwhile, the IC keywords in both academic and practical words help to extract the paragraph to some extent.

(2) The process of IC extraction

Compared to the research extracting IC only using the IC academic index, the supplement of the IC practical repository has a stronger ability to extract IC information which is not expressed by the IC academic terms. Especially the keyword pattern greatly increases the extraction accuracy and relevance of the results. Besides, compared to conventional content analysis by using the IC academic terms to extract IC-related information, the method identifies IC by using bottom-up keywords matching, which greatly helps to recognize the place where IC exists, especially the fast changing ones, like websites, social media space, etc. The IC knowledge repository is not a static one; rather it is a dynamic IC-related keywords register. Meanwhile, the intelligent ability of learning can recognize the different keywords used in different contexts. Thus the dynamic capability can relieve the contradiction of replication and comparison among the various countries, industries, years, etc.

(3) The process of sentimental analysis

The sentimental analysis method applied helps analyze the quality of IC information which has received little attention in IC research but has great value in regard to the IC information extracted. Meanwhile, this research offers a clear process of analysis, which can be replicated by other researchers. The good and bad IC-related information analyzed also facilitates the development of an IC eco-system by focusing on more human beings' perceptions and opinions. Thus, the value in this project is extended to caring the value of human beings' good life.

5.2 Limitations of IEIC

(1) Too much human intervention

Firstly, it is a huge task to construct the knowledge repository. Although the learning mechanism of KBICE can improve the performance of the extraction process, it still requires intermittent human interventions to review the suggestions of the system and provide feedback to the system. The more human interventions, the more subjective the results will be. Meanwhile, the standard result of IC sentimental analysis is only set by one IC analyst; this also results in subjectivity to some extent.

(2) Single industry

Secondly, the practical repository is constructed based on the IC reports of a single industry. However, the factor of industry is still not an identified factor that affects the IC information being used (Branco et al., 2010; Brüggen, Vergauwen, & Dao, 2009; Bukh, Nielsen, Gormsen, & Mouritsen, 2005; Flöstrand, 2006; Oliveira et al., 2006). Thus the IC knowledge repository may be ineffective for other industries. Meanwhile, the banking industry is a special industry which produces the most IC reports; however, for other industries which do not publish IC reports one may have to use annual reports instead.

(3) Lack of experimental materials

Thirdly, the IC sentimental analysis bridges the gap of recognizing the quality of IC research which has already received empirical validation. The results of extraction also show the effectiveness of the method. However, the experiment of this thesis was just done with annual reports, which cannot guarantee the same or even higher accuracy of other reports, for example, the internal and external social media which are written in various styles.

Chapter 6. Conclusion and suggestions for Future Work

This chapter concludes the significance of the research work of the whole thesis and suggests the direction of future development. Regarding the significance of the research, the contribution to the relevant literature and how the methods of this research reach the objective of the thesis are highlighted.

6.1 Significance of the research

This research conducts a thoroughly critical review of the IC development of IC research in regard to the theory and methods. It is concluded that now IC research and application have already entered the fourth stage of the IC eco-system. In this stage, this research finds that one of the biggest challenges which constrain the development of the IC eco-system is the lack of methods to cope with data where IC exists. The current methods fail to extract relevant IC information in a reliable and highly efficient way from the data which is huge in volume, unstructured as well as the fast-changing environment. On one hand the methods are not used according to the current IC research theory development needs, but on the other hand, the dynamic and complex environments pose new challenges for the data where IC exists.

After critical analysis of the challenges and deficiencies of content analysis, it is found that manual content analysis is a very time-consuming and labour-intensive method in terms of extracting data. Meanwhile, automatic content analysis encounters many difficulties without much human intervention as IC is a concept that should be understood in the context and manual content analysis has weakness in identifying practical IC-related information. Besides, both the manual and automatic modes ignore the complexity of IC and pay little attention to the specific process of analyzing the quality of IC information.

Based on these problems, an intelligent technique is designed which contains an IC knowledge repository, a dynamic IC repository as well as IC sentimental analysis. The IC knowledge repository consists of an IC academic and practical knowledge repository. In the repository, the IC-related keywords/phrases and their combination patterns are stored. For the IC academic knowledge repository, the sources originate from the 110 academic papers in which there are many IC academic terms that are used to extract IC-related information. For the IC practical repository, the content is from IC reports and corresponding annual reports. In the process of constructing the IC knowledge repository, the IC academic knowledge repository offers the indication to construct the IC practical knowledge repository and, meanwhile, is supplemented with practical words by the practical knowledge repository. Thus, the accuracy of the IC knowledge repository in extracting IC is greatly increased. Especially the capability is improved for extracting IC-related information which is not expressed by the IC academic words. Meanwhile, the IC-keywords/phrases combination patterns also help to virtualize the inter-relationship of different components, which reveals the complexity of IC application extraction. Then the accuracy and relevance of the intelligent technique are both improved.

IC information extraction which deploys the computational linguistics and artificial intelligence (AI) technique has the ability to learn new cases. Therefore, a dynamic IC checklist is compiled to help the method to be deployed in different contexts. Besides, the bottom-up keyword matching has value in conducting more 'bottom up' analyses for what IC may be used 'in practice' to complement the academic top-down proposals. The sentiment analysis used in this research describes the very specific process of how the IC information is analyzed from the perspective of news-tenor. This ability greatly helps to analyze the quality of IC, which helps to classify IC information into good, bad and neutral. Meanwhile, the IC information can be extracted which is related to the opinion of people also.

In the end, the methodology is successfully validated by the content analysis papers and data from the banking industry. From the process of IC practical knowledge repository construction and the validation of IC information extraction and sentimental analysis, it is shown that the banking industry is the pioneer industry with regard to publishing IC reports. Meanwhile, IC news tends to be more critical that is produced by third parties than the IC information produced their own organizations.

6.2 Suggestions for future research

Even though this research has made many improvements, there are still many aspects that can be improved further. Firstly, more automatic processes should be developed as there is still a large amount of human interventions involved in the process of building and maintaining the IC knowledge repository. Secondly, the IC practical knowledge repository built is based only on the banking industry. In the future, more industries should be involved to enrich the repository. Thirdly, more techniques to conduct IC sentimental analysis data should be used to test the technique. Then the technique can be improved to find more data which can be used to support the fourth stage of IC development and application.

Appendix A: Computer-aided content analysis methods in coping with IC-related information

Electronic Search	Purpose	Text	Time period	ICR index	No.	Recording unit	Methods-related findings	Method Implication
Bontis (2003)	To examine the level of IC reporting in Canada	11000 annual reports	1 year	Most popular literature which is summarized by researchers from the World Congress	38	Keywords	The level of IC reporting is low	More annual reports from other geographical locations and other years can be involved.
Vergauwen and Van Alem (2005)	To investigate the IC reporting in the Netherlands, France and Germany.	180 annual reports	3 years	Bontis (2003)	38	Keywords	IC reporting varies across different countries.	Searching result is affected by accounting regulations due to the tension created by the IC information
Vergauwen et al. (2007)	To measure the level of IC reporting in Sweden, Britain and Denmark	20 annual reports	Not specifi c	The former literature (Bozzolan et al., 2003; Brennan, 2001; Goh & Lim, 2004; Vergauwen & Van Alem,	103	Keywords /Manual identification of hits	Not all the IC-keywords in the checklist are searched; RC elements are most often reported; IC reporting varies from	Subjective as people are involved; More attention should be paid on the firm from different industries.

				2005)			different industries.	
Kamath (2008)	To study the extent of voluntary intellectual capital reporting in India's emerging information, communication and technology sector	30 annual reports	1 year	Bontis (2003)	38	Keywords s	The extent of IC reporting is not high	More annual reports can be involved to "study the comparative picture across sectors and countries."
Sonnier et al. (2008)	To measure the level of IC reporting in the US	423 annual reports	1 year	Prior literature	121	Keywords and phrases	IC reporting increased from 2000 to 2004.	This approach cannot capture the IC context in a word or phrase; The software cannot tell the IC information which is just planned; The software cannot recognize the IC

								underlying the meaning behind the key words and phrases.
Oliveras et al. (2008)	To measure the level of IC reporting in Spain	36 annual reports	3 years	Main classifications are the same as Petty and Guthrie (2000a) but the IC elements are different	25	Keywords; context unit: sentence	The level of IC reporting is increasing over the 3 years. External capital is the largest reporting element. The level of IC reporting is not high	An extension of this longitudinal study to cover a longer time span; Qualitative studies relating to development of IC reporting e.g. interviews with corporate managers.
Branco et al. (2010)	To measure the level of IC reporting in Portuguese companies	72 annual reports	3 years	Petty and Guthrie (2000a)	72	Keywords	HC is the most reported IC element; The level of IC reporting fluctuates from 2004 to 2008.	Differences of IC reporting results may be attributed to the different content analysis method used, different country context, different industry as well as different legislation; Annual reports

								may not report all actual IC used in the organizations.
Lock Lee and Guthrie (2010)	To identify IC-related information from 156 IT firms' business and analyst reports published by external reporters	624 observations	3 years	Guthrie and Petty (2000); the sub elements are from the factiva database	21	Not specific	The level of IC reporting is limited to "newsworthy" elements; one IC information may be attributed to different classifications.	Different research methods should be compared
Lin et al. (2012)	To examine the level of HC reporting	660 annual reports	1 year	Vergauwen et al. (2007) + Table extra	40	Keywords	No IC reporting level is mentioned.	Not all annual reports are text searchable; Double-edged effect of HC reporting and time lag effect of HC reporting should be further researched

Appendix B: Examples of Rule Set

IC Category	IC Element	Key Phrases	Patterns
HC	Employee Diversity	Women %	Employee, women
RC	Customer communication (e.g. Telephone Platform, Internet)	Member receiving a pension	Receive a pension
RC	Customer Relationships	Number of attendees in membership meetings	Held open meetings, meetings, number of participants
RC	Customer Relationships	ATP-Pensioners receiving a current pension	Pensioners, received current pension
RC	Customer communication (e.g. Telephone Platform, Internet)	Number of attendees in membership meetings	Member, participated, membership meetings
RC	Customers, Image and Stakeholders	ATP-Pensioners receiving a current pension	Member paid contributions
RC	Customer communication (e.g. Telephone Platform, Internet)	Number of attendees in membership meetings	Members attended, information meetings
RC	Quality and customer satisfaction	Proportion of satisfied/very satisfied attendees in membership meetings	Attendees ,satisfied
SC	Quality of Pension summaries	Customers find the summary to be written in easy-to-understand language	Members find the pension summaries, written in easy-to-understand language.
HC	Staff Turnover	Departed employees (staff turnover)	Staff reduction
HC	Employee flexibility	Number of employees	People, working
HC	Employee flexibility	Number of employees	Originators of ideas
HC	Employee flexibility	Number of employees	Total, staff
RC	Corporation and networks	Number of SME Management Centres	SME segment, number of these branches
RC	Corporation and networks	Number of SME Management Centres	New centre specializing in SMEs, total, such centre

HC	Training (Competence development)	Total number of courses taught	Workforce, receive training, total, courses
HC	Training (Competence development)	Average number of training hours per employee trained	Training hours, per employee trained
RC	Corporation and networks	Number of SME Management Centres	Total number of SME branches
RC	RC ungrouped	Transactions through channels other than Branch Network as % of total Bank transactions	%, transaction, remote channel
SC	Quality Management and Improvements	Number of quality projects and initiatives carried out	Quality, improvement project, started
SC	De-layering and transparency	Number of virtual work forums in operation	Virtual , channel, work forums
SC	Flexibility	Internal job rotation	People rotated

References

Allee, V. (2000). The value evolution: addressing larger implications of an intellectual capital and intangibles perspective. *Journal of intellectual capital, 1*(1), 17-32.

Bernard Marr, J. C. (2004). Intellectual capital at the crossroads: managing, measuring, and reporting of IC. *Journal of business research, 5*(2), 224-229.

Bontis, N. (2003). Intellectual capital disclosures in Canadian corporations. *Journal of Human Resource Costing and Accounting, 7*(1), 9-20.

Bontis, N., & Fitz-Enz, J. (2002). Intellectual capital ROI: a causal map of human capital antecedents and consequents. *journal of Intellectual Capital, 3*(3), 223-247.

Bozzolan, S., Favotto, F., & Ricceri, F. (2003). Italian annual intellectual capital disclosure: an empirical analysis. *journal of Intellectual Capital, 4*(4), 543-558.

Branco, M. C., Delgado, C., Sá, M., & Sousa, C. (2010). An analysis of intellectual capital disclosure by Portuguese companies. *EuroMed Journal of Business, 5*(3), 258-278.

Brennan, N. (2001). Reporting intellectual capital in annual reports: evidence from Ireland. *Accounting, Auditing & Accountability Journal, 14*(4), 423-436.

Brüggen, A., Vergauwen, P., & Dao, M. (2009). Determinants of intellectual capital disclosure: Evidence from Australia. *Management Decision, 47*(2), 233-245.

Bukh, P. N., Nielsen, C., Gormsen, P., & Mouritsen, J. (2005). Disclosure of information on intellectual capital in Danish IPO prospectuses. *Accounting, Auditing & Accountability Journal, 18*(6), 713-732. doi: 10.1108/09513570510627685

Carrington, D., & Tayles, M. (2012). Intellectual Capital in the Caribbean Hospitality Industry: Two Case Studies. *Electronic Journal of Knowledge Management, 10*(3).

Chatzkel, J. (2004). Moving through the crossroads. *Journal of Intellectual Capital, 5*(2), 337-339.

Cuganesan, S. (2005). Intellectual capital-in-action and value creation. *Journal of Intellectual Capital, 6*(3), 357-373.

Dumay, & Cuganesan. (2011). Making sense of intellectual capital complexity: Measuring through narrative *Journal of Human Resource Costing & Accounting, 15*(1), 24-49.

Dumay, J. (2013). The third stage of IC: towards a new IC future and beyond. *Journal of intellectual capital, 14*(1), 5-9.

Dumay, J., & Cai, L. (2014a). A review and critique of content analysis as a methodology for inquiring into IC disclosure. *Journal of Intellectual Capital, 15*(2), 264-290.

Dumay, J., & Cai, L. (2015). Using content analysis as a research methodology for investigating intellectual capital disclosure: A critique. Journal of Intellectual Capital, 16(1), 121-155.

Dumay, J., & Garanina, T. (2013). Intellectual capital research: a critical examination of the third stage. *Journal of Intellectual Capital, 14*(1), 10-25.

Dumay, J., & Tull, J. (2007). Intellectual capital disclosure and price sensitive Australian stock exchange announcements. *Journal of Intellectual Capital, 8*(2), 236-255.

Dumay, J. C. (2012). Grand theories as barriers to using IC concepts. *Journal of Intellectual capital 13*(1), 4-15.

Edvinsson, L. (1997a). Developing intellectual capital at Skandia. *Long range planning, 30*(3), 366-373.

Edvinsson, L. (1997b). Developing intellectual capital skandia. *Long Range Planning 30*(3), 366-373.

Edvinsson, L. (2013). IC 21: reflections from 21 years of IC practice and theory. *Journal of intellectual capital, 14*(1), 163-172.

Edvinsson, L., & Malone, M. S. (1997). Intellectual Capital: Realizing Your Company\'s True Value by Finding Its Hidden Brainpower.

Flöstrand, P. (2006). The sell side – observations on intellectual capital indicators. [DOI: 10.1108/14691930610709112]. *Journal of Intellectual Capital, 7*(4), 457-473.

Goh, P. C., & Lim, K. P. (2004). Disclosing intellectual capital in company annual reports: evidence from Malaysia. *journal of Intellectual Capital, 5*(3), 500-510.

Guthrie, J., & Petty, R. (2000). Intellectual capital: Australian annual reporting practices. *Journal of Intellectual Capital, 1*(3), 241-251.

Grindley, P. C., & Teece, D. J. (1997). Managing intellectual capital: licensing and cross-licensing in semiconductors and electronics. California Management Review, 39(2), 8-41.

Guthrie, J., Petty, R., Ferrier, F., & Wells, R. (1999, June, 1999). *There Is no accounting for intellectual capital in Australia: A review of annual reporting practices and the internal measurement of intangibles within Australian organisations.* Paper presented at the International Symposium on Measuring and Reporting of Intellectual Capital: Experiences, Issues and Prospects, Amsterdam.

Guthrie, J., Petty, R., Yongvanich, K., & Ricceri, F. (2004). Using content analysis as a research method to inquire into intellectual capital reporting. [Literature Review; Survey; Comparative/evaluators]. *Journal of Intellectual Capital, 5*(2), 282-293.

Guthrie, J., Ricceri, F., & Dumay, J. (2012). Reflections and projections: a decade of intellectual capital accounting research. *The British Accounting Review, 44*(2), 68-82.

Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews.* Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

Indra Abeysekera, & Guthrie, J. (2005). An empirical investigation of annual reporting trends of intellectual capital in Sri Lanka. *Critical Perspectives on Accounting 16*(3), 151-163.

Jindal, S., & Kumar, M. (2012). The determinants of HC disclosures of Indian firms. *Journal of Intellectual Capital, 13*(2), 221-247.

Johanson, U., Koga, C., Skoog, M., & Henningsson, J. (2006). The Japanese government's intellectual capital reporting guideline: What are the challenges for firms and capital market agents? *Journal of Intellectual Capital, 7*(4), 474-491.

Kamath, B. (2008). Intellectual capital disclosure in India: content analysis of "TecK" firms. *Journal of Human Resource Costing & Accounting, 12*(3), 213-224.

Krippendorff, K. (2013). *content analysis: An introduction to it methodology* (3 ed.).

Laurence Lock Lee, & Guthrie, J. (2010). Visualising and measuring intellectual capital in capital markets: a research method. *Jouranl of Intellectual Capital, 11*(1), 4-22.

Lev, B. (2004). Sharpening the intangibles edge. Harvard business review, 6, 109-116.

Lin, L.-S., Huang, C., Du, P.-L., & Lin, T.-F. (2012). Human capital disclosure and organizational performance: The moderating effects of knowledge intensity and organizational size. *Management Decision, 50*(10), 1790-1799.

Lock Lee, L., & Guthrie, J. (2010). Visualising and measuring intellectual capital in capital markets: A research method. [DOI: 10.1108/14691931011013307]. *Journal of Intellectual Capital, 11*(1), 4-22.

Marr, Schiuma, & Neely. (2004). The dynamics of valus creation: mapping your intellectual performance drivers *Journal of Intellectual Capital, 5*(2), 312-325.

MERITUM. (2002). Guidelines for Managing and Reporting on Intangibles (Intellectual Capital Statements). In L. Cañibano, P. Sánchez, M. García-Ayuso & C. Chaminade (Eds.). Madrid: Vodafone Foundation.

O'Donnell, D., Henriksen, L. B., & Voelpel, S. C. (2006). Guest editorial: becoming critical on intellectual capital. *Journal of Intellectual Capital, 7*(1), 5-11.

Oliveira, L., Rodrigues, L. L., & Craig, R. (2006). Firm-specific determinants of intangibles reporting: evidence from the Portuguese stock market. *Journal of Human Resource Costing & Accounting, 10*(1), 11-33.

Oliveras, E., Gowthorpe, C., Kasperskaya, Y., & Perramon, J. (2008). Reporting intellectual capital in Spain. *Corporate Communications: An International Journal, 13*(2), 168-181. doi: 10.1108/13563280810869596

Pedrini, M. (2007). Human capital convergences in intellectual capital and sustainability reports. *Journal of Intellectual Capital, 8*(2), 346-366.

Petty, R., & Guthrie, J. (2000a). Intellectual capital literature review: Measurement, reporting and management. *Journal of Intellectual Capital, 1*(2), 155-176.

Petty, R., & Guthrie, J. (2000b). Intellectual capital literature review: measurement, reporting and management. *Journal of Intellectual Capital, 1*(2), 155-176.

Rastogi, P. (2003). The nature and role of IC: Rethinking the process of value creation and sustained enterprise growth. *Journal of Intellectual Capital, 4*(2), 227-248.

Richard Petty, & Guthrie, J. (2000). Intellectual captial literature review: Measurement, reproting and management. *Journal of Intellectual Capital, 1*(2), 155-176.

Roos, J., Roos, G., Dragonetti, N. C., & Edvinsson, L. (1997). *Intellectual capital*: Macmillan.

Schmid, H. (1994). *Probabilistic part-of-speech tagging using decision trees.* Paper presented at the Proceedings of the international conference on new methods in language processing.

Sivic, J., & Zisserman, A. (2009). Efficient visual search of videos cast as text retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31*(4), 591-606.

Sonnier, B. M., Carson, K. D., & Carson, P. P. (2008). Intellectual capital disclosure by traditional US companies: a longitudinal assessment. *Journal of Accounting & Organizational Change, 4*(1), 67-80.

Stewart, T., & Ruckdeschel, C. (1998). Intellectual capital: The new wealth of organizations: Wiley Online Library.

Stewart, T., & Ruckdeschel, C. (1998). Intellectual capital: The new wealth of organizations. *Performance Improvement, 37*(7), 56-59.

Stewart, T. A. (1997). *Intellectual Capital: The New Wealth of Organisations*: Doubleday/Currency, London.

Sveiby. (1997). *The new organizational wealth: Managing & measuring knowledge-based assets#*: Berrett-Koehler Publishers.

Sveiby, K. E. (1997). The new organisational Wealth: Managing and Measuring Knowledge Based Assets. *Berrett Koehler, SanFrancisco, CA.*

Vergauwen, P., Bollen, L., & Oirbans, E. (2007). Intellectual capital disclosure and intangible value drivers: an empirical study. *Management Decision, 45*(7), 1163-1180.

Vergauwen, P. G., & Van Alem, F. J. (2005). Annual report IC disclosures in the Netherlands, France and Germany. *journal of Intellectual Capital, 6*(1), 89-104.

Wasiluk, K. L. (2013). Beyond eco-efficiency: understanding CS through the IC practice lens. *Journal of intellectual capital, 14*(1), 102-126.