CHALMERS

Dental Implant Technology Clustering and Technology Life-span Analysis Using Ontology-based Patent Intelligence

Master of Science Thesis in the Master Degree Programme, Biotechnology

SEAN LONG HOANG

Department of Chemical and Biological Engineering
Division of Applied Surface Chemistry
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2012

Title: Dental Implant Technology Clustering and Technology Life-span Analysis Using Ontology-based Patent Intelligence
Author: Sean Long Hoang

Department of Applied Surface Chemistry
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

Supervisor: Dr. Charles V. Trappey
Department of Management Science
National Chiao Tung University
300, Hsinchu
Taiwan

Examiner: Professor Krister Holmberg
Department of Chemical and Biological Engineering
Division of Applied Surface Chemistry
SE-412 96 Göteborg
Sweden

Göteborg, Sweden 2012

Dental Implant Technology Clustering and Technology Life-span Analysis Using Ontology-based Patent Intelligence

SEAN LONG HOANG

Department of Chemical and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden

Dental Implant Technology Clustering and Technology Life-span Analysis Using Ontology-based Patent Intelligence

Student: Sean Long Hoang

Advisor: Dr. Charles V. Trappey
National Chiao Tung University

Examiner: Prof. Krister Holmberg
Chalmers University of Technology

ABSTRACT

Rapid technology development shorter product life cycle, fierce competition in the marketplace, establishes patent analyses are an important strategic tool for R&D management. This thesis develops a technology clustering and life-span analysis framework based on data mining techniques to help companies effectively and rapidly gain domain-specific knowledge and technology insight. In addition, patent documents contain complex terminologies which require experts to perform patent analysis. This research applies patent analysis methodologies to create domain-specific ontologies. The advantage of using an ontology is that it contains specific domain concepts and helps researchers to understand the relationships between concepts. In addition, ontologies are used to effectively extract domain knowledge, cluster patents and create graphs for tend recognition. Life-span analysis of technology clusters helps companies to gain a quick snapshot of their own patent portfolio and identify potential technology clusters for investment.

This thesis proposes the process of knowledge extraction for domain-specific patents using patent analysis methodologies which improve domain-knowledge understanding. The methodologies proposed in this research include key phrase analysis, patent technology clustering, patent document clustering, domain-specific ontology, and life-span analysis. With these methodologies, companies quickly derive domain-specific ontologies to help R&D engineers relate data and increase understanding of a specific domain and the relationship between concepts. Life-span analysis helps companies' direct strategic R&D plans and evaluates the timing of investments using the methodologies proposed in this research. The validity and reliability of the methodology are tested by studying the application of a set of dental implant patents.

Keywords: Dental implant, ontology, key phrase analysis, clustering, life-span analysis

ACKNOWLEGEMENTS

I would like to thank my advisor Dr. Charles V. Trappey for guidance, advices, and helping me to excel to a higher level. I deeply appreciate his sharing of knowledge and experiences, and moreover his fun stories were always enjoyable. I would also like to thank a very kind person, Dr. Chun-Yi Wu, for his guidance, time, and effort throughout my time working on this thesis. Without the use of the IPDSS software, developed by Dr. Wu, this thesis would not have been possible to execute.

I would also like to thank my family and friends! My time in Taiwan has been wonderful thanks to your kindness and hospitality.

"We make a living by what we get, but we make a life by what we give"

By Sir Winston Churchill

Sean Long Hoang

Chalmers University of Technology

Hsinchu, Taiwan

December 2011

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

This chapter describes the value of patent analysis and the importance of ontologies. The background, motivations, procedures, and objectives of this research are also discussed.

1.1. Research background

The global dental equipment and product market is estimated to grow at a Compound Annual Growth Rate (CAGR) of 7% reaching US$27.6 billion by 2015 (Salisonline 2011). Some countries face challenges in the medical dental device industry, since they lack sufficient expertise and skills, public and private funding specially at early phases where innovation is developing at a high risk level, and also lack a national priority or strategy for the developing sector (University of Ottawa 2011). Other issues are litigation, government regulations and most importantly, Intellectual Property (IP) management of innovative technologies as well as existing core technologies (University of Ottawa 2011).

It is important to maintain a high quality patent portfolio to quickly create innovative products, advance innovative technology development and protect innovations as well as avoid litigations (Trappey et al. 2011). Traditional patent analysis requires time, effort and expertise to interpret the research results. Recent patent analysis techniques use data mining as a tool to extract data of large volumes from information which decrease the time and effort for analyzing patents (Lee et al. 2009). These techniques apply statistical analysis techniques to automatically perform key phrase analysis, document clustering, life cycle analysis, and so on. High technology companies strive to orient R&D and strategic plans with emerging technologies. Patent documents are available through databases and are rich sources of information which provide a foundation for technology trend analysis.

Patent documents in specific technology domains contain domain-specific terms which require domain experts and their experience to perform patent analysis (Jun & Uhm 2010). This limits the opportunity for researchers or R&D engineers to explore and understand patent information using data mining techniques. However, according to Wanner et al. (2008) using data mining techniques should include an ontology. Ontologies can be seen as an organized hierarchical structure with abstracted domain concepts and relations expressed in terms of domain terminologies and concepts. The advantage of using an ontology is that it contains a specific domain corpus to understand the meaning of these terminologies. The ontology helps researchers and R&D engineers relate the data and increase the understanding of a specific domain as well as understand the relationship between domain concepts.

The current life cycle stage of the technology influences companies when invest R&D capital on technologies (Haupt et al. 2007). A life cycle can be divided into four stages: introduction, growth, maturity and decline. The introduction stage of a new technology often includes technical problems or scientific fundamental problems and is associated with high risk of investments of R&D capital. A technology at the mature stage requires companies to evaluate the boundaries of patents to avoid infringement issues, trade IP, or license IP.

1.2. Research motivation

A company must meet the market demands for new products. Designs to maintain market position consistently require advancement in technology and innovation. Companies with a strong patent portfolio use intellectual property as leverage in the marketplace to gain competitive advantage. In addition, companies face increased competitiveness and budget constraints to effective allocate resources to specific technology areas for research and development. The rapid advancement in technology and shorter product life cycles results in time constraints for companies to strategically plan R&D activities. Moreover, patent documents contain complex terminologies that require hiring high-fee experts that require much time for analysis. Since R&D engineers often do not have the skills to perform patent analysis. The main advantage of using ontologies is that it contains domain-specific knowledge and concepts which enable R&D engineers to gain valuable insights of that domain to understand concept relationships. Companies need help to direct R&D plans and evaluate the timing for future technology investments, a life-span analysis can be used to gain a quick snapshot of the potential technology clusters for potential investment strategies. Patent analysis helps companies to target R&D plans towards recent technology trends and identify future R&D plans (Trappey et al. 2011). Companies with a strong patent portfolio and conduct strategic patenting activities are more successful than other companies that remain inactive in for example the field of mechanical engineering and biotechnology sector (Ernst 1995; Austin 1993).

Companies using patent analysis can observe technology development and identify potential competitors at the market place. Since a patent is granting the inventor exclusive rights for a limited time period to exclude anyone to produce or use this specific device, apparatus, process, or design (Grilliches 1998). In addition, a patent or patentability of a technology is also one of the preconditions of the commercial potential of a technology. Moreover, patents can be used to generate life cycle analysis to monitor technology development and help companies to identify potential R&D investment opportunity or strengthen their IP-position (Haupt et al. 2007). Patent analysis can be used to study a country's performance (Grilliches 1990), governments use it to allocate resources to specific technology areas (Yoon & Yongtae 2007), and monitor technology development of competitors (Jun & Uhm 2010). Some companies avoid filing patents and patent only the most successful innovation. Subramanian & Soh (2010) explain that most high technology firms are in a patent race and patents are considered to be linked to the firms' performance.

1.3. Research procedure

The research procedure is divided into 6 phases and listed accordingly:

- Phase 1: Motive and objectives
- Phase 2: Literature review
- Phase 3: Selection of research direction

  The direction of this research is selected according to predefined objectives and literature reviews. Selected development is: data mining, key phrase analysis, technology clustering, patent document clustering, and technology life-span analysis.

- Phase 4: Methodology development
- Phase 5: Case study
- Phase 6: Evaluation

The research procedure and framework is shown in Figure 1.

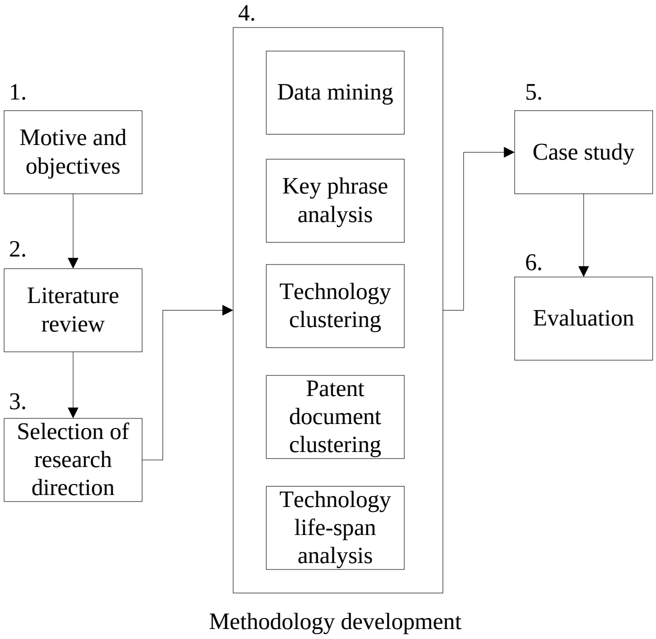


Figure 1. Research procedure and framework

1.4. Research objectives

The goal of this research is the development of a domain-specific ontology framework, which will be used for technology clustering and life-span analysis of clusters. This research intend to:

1. To synthesis methodologies to generate the useful information to construct a domain-specific patent ontology structure and procedure based on dental implants patents
2. To develop a patent domain-specific ontology framework based on patent classifications to gain domain-specific knowledge and relationships between concepts
3. To develop a procedure for ontological sub-domain technology clustering and technology life-span analysis

1.5. Research limits

This research is subjected to the following limitations. The first limitation points to patent data are only from the United States Patent and Trademark Office (USPTO). The ontologies are built based on patent data only from USPTO, the researchers experience and knowledge on the dental implant domain, and with help from dictionary (WordNet) to link the concepts in the ontology.

2. BACKGROUND

This chapter has eight sections, separately discusses dental implant, data mining, patent analysis, key phrase extraction, clustering, technology life cycle analysis and research framework. The background covers the foundation methodologies data mining, key phrase analysis, patent technology clustering, patent document clustering and technology life cycle analysis for developing the methodology of creating a domain-specific patent ontology.

2.1. Dental implant

The largest segment of the global market is restorative and preventive dentistry, but the fastest growing segment is dental implantology. Dental implant has a market share of 18% of the global dental device market (Markets and Markets 2010).The global dental equipment and product market is experiencing a steady growth due to a number of factors such as demand of dental implants, desire for aesthetics, and increased usage of dental preventive care (Brocair Partners 2010; Salisonline 2011). The popularity of cosmetic dental treatments and implants along with increasing demand for better dental care will drive the growth for new innovative dental implant technologies and products (Salisonline 2011).

Currently, millions of people benefit from having dental implant since this is the ideal option for people, who has lost a tooth or teeth due to gum disease, an injury, or some other reason. Dental implant is an artificial tooth root that is placed into the jaw that holds a replacement tooth. Dental implants are the only possible option of replacing missing teeth which closely resemble a natural tooth and that behaves exactly like real roots and bonds naturally to the jaw bone. The crown is then bonded to the top of the dental implant.

Paterson and Zamanian (2009) confirm that the dental implant industry will experience a strong growth from 2011 to 2015 for the global market and along with emerging technologies will improve the dental efficiency of dental procedures and reduce the time. For an example, Salisonline (2011) points out is 3D imaging techniques has improved patient diagnosis and procedure planning. Other emerging technologies in dental implant industry or biotechnology industry are dental biomaterials and tissue regenerative materials which offer a more natural and long-term solution. For the U.S and EU market, these trends are changing the customers need towards a shift to cosmetic dentistry and drive the dental implant market to high-end dental solutions and products (Salisonline 2011).

Dental implant is defined as an implant that replaces a natural tooth (WordNet Princeton University, 2011). The main components of a dental implant include a screw that is able to connect with a custom-made crown. Figure 2 describes the main components of a dental implant compared with a natural tooth.

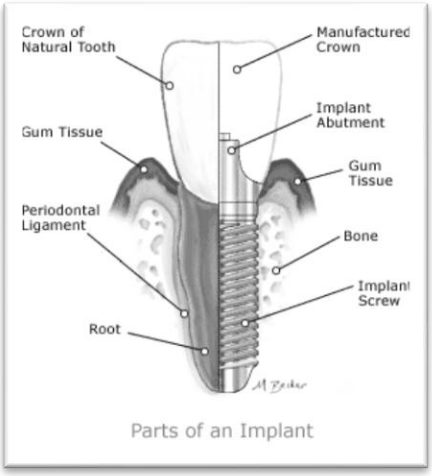


Figure 2. The construction of dental implants (Puja Dental Group 2011).

Various components of dental implants are: implant body, cover screw (prevent to access the bone), transmuscosal abutment (links the implant body to the mouth), healing abutment (temporarily placed on implant to maintain potency of the muscosal penetration), healing caps (temporary covers for abutments), crowns, bridges, gold cylinder (to fit an abutment and form part of prosthesis), laboratory analogue (a base metal replica of implant or abutment), etc. (AstraTech Dental 2011 and Free Dental Implant 2011).

2.2. Data mining

Recently, data mining is one popular alternative to use for extracting information from databases. Data mining applies machine learning and statistical analysis techniques to access and extract information from databases with large volumes of patent documents (Lee et al. 2009). It automatically discovers patterns in databases and is useful for mapping scientific and technical information for complex analysis of large volumes of information (Kim et al. 2008). Fayyad et al. (1996) states that it requires to develop methods and techniques to interpret the data for it to make sense for humans. The facts are that data volumes are growing rapidly in both objects and records in thousands of various fields, for example in the medical field it can easily be divided into hundreds of different fields (Fayyad et al. 1996).

Data are captured for different purposes, in business for gaining competitive advantage or environmental data to better understand the effects. It has been applied in marketing, investment, fraud detection, manufacturing, and telecommunications. For example in marketing, the primary application with data mining is to analyze the database to identify customer groups and forecast future buying patterns or behavior (Fayyad et al. 1996). Furthermore, data are a set of facts and pattern equals to the language that describes a model or finding a structure from the data. In general, by model or structure means the value of the pattern combined with validity, novelty, usefulness, and simplicity. All of these conditions does not define “knowledge” but rather define the framework of the pattern recognition of knowledge and it should also be taken into consideration that it is purely user oriented and domain specific as well as functions are determined by the user (Fayyad et al. 1996).

Fayyad et al. (1996) states that the term Knowledge Discovery in Databases (KDD) is the overall process of discovering knowledge from a database and that data mining is actually a step in that process. Data or text mining are the same, historically it has been given a variety of names but the concept of finding patterns in data is the same. The concept of data mining is to use algorithms to extract patterns in data, also known as information retrieval (Fayyad et al. 1996). KDD uses additional steps such as data preparation (storage and access), data selection, data cleaning, incorporation of appropriate prior knowledge, and logical interpretation of the results, are important and need to be understandable knowledge extracted from the data, statistics to provide the framework and language for discovery of patterns (Fayyad et al 1996). The overall data mining framework is shown in figure 3.

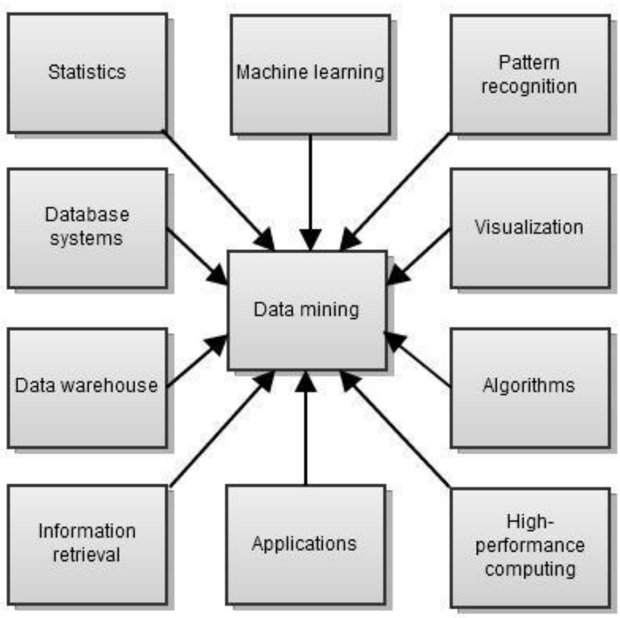


Figure 3. Overview of data mining, adapted from Han J. et al. (2011)

2.2.1. Text mining

Text mining is a technique developed from data mining to analyze textual data especially unstructured (free text, abstract, etc.) textual documents for example patent documents (Lee et al. 2009; Kostoff et al. 2009). Text mining utilizes a technique to put a label of each document and link them to specific words which allows the discovery be based on labels (Lee et al. 2009). A text document is unclear, and according to Nasukawa and Nagano (2001), it contains various types of information, richness in information and represents factual information. Most information stored in a database is in the form of text documents. Text mining is used for automatic discover knowledge and patterns in a database by applying statistical algorithms (Weiss et al. 2005). Text mining is a broad field that involves information retrieval, text analysis, information extraction, clustering, categorization, visualization, machine learning, and data mining (Tan 2011; Lee et al. 2005).

Patent documents contain detailed information in complex technical and legal terms only experts in specific field understand and the purpose is to make it difficult for non-specialist to read and analyze (Tseng et al. 2005). Patent documents are often lengthy and traditional patent analysis is inefficient, require long time and human effort to analyze the contents which also is highly expensive to maintain (Lee et al. 2009; Tseng et al. 2005).

In August 16, 2011 the United States Patent Trademark Office (USPTO) issued patent number 8 000 000 and according to USPTO statistics there are a couple hundred thousand patent applications pending for examination each year (USPTO 2011). The accumulation of patent documents at the USPTO has increased at a striking pace due to more patent applications and granted patents (USPTO 2011). Therefore, it exist a great demand for automated data mining techniques for extracting information from the rapid growing volume of data into a compact form that can be easier to absorb. Large text databases such as USPTO database potentially contain great amount of information (knowledge) if and only if it can be interpreted. The traditional method of turning raw data into knowledge require manual analysis and huge amount of reading and organizing the content, thus it requires huge amount of workload for analyzing only a tiny fraction of the database (Lee et al. 2009). Furthermore, it is expensive and very subjective analysis is provided by the analysts (Fayyad et al. 1996).

Recently, text mining has attracted researchers to apply it on patent analysis (Kim et al. 2008; Yoon and Park 2004). For example, Tseng et al. (2007) used text mining techniques to create a patent map for technology domain of carbon nano-tubes. Tseng et al. (2005) applied text mining techniques to automatically create important categorization features that might be good as human derived or even better. One of the advantages using text mining techniques in patent analysis is that it can handle large volumes of patent documents and extract useful information (Lee et al. 2009). Since patent documents are lengthy but contain significant technical information and automatic text mining will assist researchers, engineers or decision-makers in patent analysis (Lee et al. 2009). However, extracted data has to meet specific quality criteria to be comprehensible for humans and also represent the concept of the text or benefit for the user (Yoon and Park 2004). Furthermore, text mining techniques has been applied for summarization, term association, cluster generation, topic identification, mapping information, technology trend analysis, automatic patent classification, and so on (Lee et al. 2009; Yoon and Park 2004; Tseng et al. 2007).

2.2.2. Limitations in text mining

Although text mining seems to be a very promising technique for analyzing textual data there are some limitations such as areas that require accuracy. It is useful for providing supportive information for analysis (Smith 2002). Text mining techniques face a significant challenge in dealing with patent documents because algorithms cannot include compound words because of difficulties in determining them and cannot consider synonyms (Lee et al. 2009; Smith 2002). Furthermore, in terms of accuracy using text mining to make sure there is a distinction between documents it requires a large number of keywords (Smith 2002). Additionally, using text mining for unstructured data in patent documents, difficulties can be encountered in distinguishing texts or keywords that are describing “prior art” from texts which describe the invention (Smith 2002). This is important since the description will address the technical characteristics of the patent invention. However, despite these limitations, advances in computer science and better text mining algorithms are expected to strengthen text mining advantages making it more efficient and accurate.

2.3. Ontology

Huang et al. (2008) describes that the concepts of ontology is a model which contain the concepts, links and relationships in a specific domain that reflects the reality of the world. WordNet from Princeton University (2011) define ontology as: a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations. Ontology provides a unified knowledgebase expressed in the information domain that has integrated information from various sources (Taduri et al.

2011). For example, a company that needs to collect information of a specific technology and has only initial knowledge, domain-specific ontology can be used to collect relevant information much faster than existing systems (Taduri et al. 2011). The ontology contains domain concepts and relations which can be reused, modified, and shared among R&D engineers (Soo et al. 2006). Ontology provides an organized framework with a hierarchical structure and relationships of the domain which offers the possibility to understand relations between concepts (Rubin et al. 2007).

Ontology links the semantic data between concepts which makes it possible to perform pattern recognition, similarity, and clustering of patent documents with respect to its content (Wanner et al. 2008). A variety of methods has been proposed to create knowledge domains and one of the methods suggests a single ontology that integrates all knowledge domains (Taduri et al. 2011). The potential drawback of this method is that it creates a very large set of knowledge domains which depending on the application may be unnecessary and inefficient (Lau et al. 2011). Alternatives ontology architectures propose having separate ontologies that are domain-specific which are application specific (Noy & McGuiness 2001). Many methodologies have been proposed and are used to create ontologies for capturing domain knowledge of patents to enhance the information retrieval (Taduri et al. 2011).

Ontology-based patent intelligence

A significant effort has to be taken into consideration when gathering relevant patent information across different patent databases (Khelif et al. 2007). For example, a start-up company wanting to patent their technology in the field of dental implants wants to search patent databases, scientific publications, and perform patent analysis for infringement purposes and competitor analysis (Taduri et al 2011). They face challenges to thoroughly search for patent documents and the large volumes of patent documents makes it an almost impossible task (Lau et al. 2011). Patent documents in specific technology domains contain domain-specific terms which cannot be covered in common dictionaries, therefore an advantage of ontology is that it contain specific domain terminologies (Soo et al. 2006). Trappey et al. (2010) points out that ontology are useful to extract related concept of key phrases. Ontology serves as an organized structure for arranging or classifying a domain. In addition, ontology is a way of formally represent knowledge domains with concepts, their attributes and relations between terminologies expressed in some well defined logic (Rubin et al. 2007). According to Wanner et al. (2008), using text mining techniques such as key phrase extraction for representing the content of a patent document it should also contain important feature: ontology. The goal of using domain specific ontology is to reduce conceptual and terminological confusion among R&D engineers (Navigli and Velardi, 2004). In addition, sharing the domain-specific ontology can improve the communication, cooperation among people, better enterprise organization, and system engineering (reusability, reliability, and specification) (Navigli and Velardi, 2004).

Ontology approach for knowledge extraction has been applied in various fields. For example, Trappey et al. (2009) applied an ontology tree for automatic patent document summarization which extracts key information into a shorten abstract with the key concepts of the patent document. The goal was to use the ontology to create a knowledge base for their software program to improve its architecture and consistency in capturing knowledge of their information system. Figure 4 demonstrates the example RFID technology ontology tree from Trappey et al. (2010) that were used for patent document summarization. Rubin et al. (2007) propose how biomedical ontologies can help researchers to accelerate their research because biomedical information available is exploding. Biomedical ontologies help researchers to structure complex biological domains and relate the data, for example, the gene ontology have gained huge attention in the biomedical community (Rubin et al. 2007).

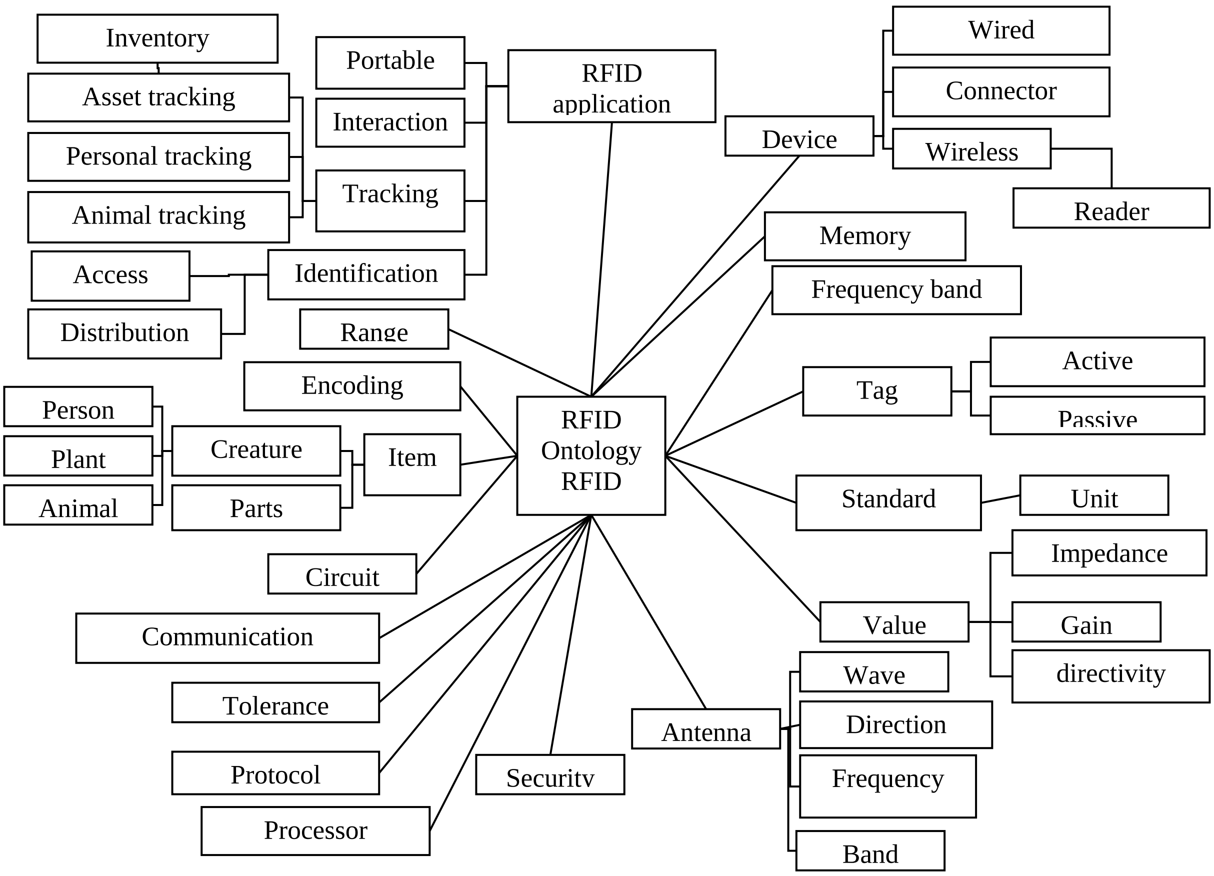


Figure 4. An example ontology tree: RFID technology (Trappey et al. 2010)

To create a domain-specific patent ontology requires phrases that describe the concepts of patent documents (Trappey et al. 2010). It requires identifying and defining relevant concepts and relating it to a given application (Navigli and Velardi, 2004). The challenge when dealing with specific technical domains such as telecommunications, biotechnology, biomedical, there is specific technical or domain specific terminology in patent documents (Taduri et al. 2011). The terms are often a challenge since it is presented in several forms such as synonyms and hyponyms etc. which makes the general language comparison in patent documents inefficient. Taduri et al. (2011) and Mukherjea & Bamba (2007) points out the advantages using ontology to capture the rich information available and allows the application to understand the semantics associations to avoid terminological inconsistencies. It also allows users to reason across the knowledge domain where some applications require small fragments of information which let users to choose to work with only information that is needed (Lau et al. 2011). For example, R&D engineers may only be interested in a technological sub-domain and ignore the other knowledge sub-domains.

2.4. Patent analysis

Patent documents contain rich detailed information about research results that are in complex technical and legal terms, it is valuable to the industry, business, law, and policy-making communities (Tseng et al. 2007; Choi et al. 2007). Thus, the detailed content in patent documents, if carefully analyzed, can reveal technology development, inspire novel technical solutions, show technical relations, or help investment policy (Tseng et al. 2005). Tseng et al. (2007) point out that patent analysis has become important even at government level at some Asian countries such as China, Japan, Korea, Singapore, and Taiwan. These countries have invested various resources to create visualized results of patent analysis (Lee et al. 2009). Patents are a useful vehicle for R&D and technology management research since it is a source of technical and commercial information which can be turned into knowledge (Choi et al 2007; Lee et al. 2009). Patent documents are often lengthy and require time, effort and expertise to interpret the research results into a technology development analysis. Tseng et al (2007) also emphasize that patent analysts also need a certain degree of expertise in information retrieval, domain-specific technologies, legal knowledge and some business intelligence. A typical patent analysis scenario is showed on Table 1, these multi-discipline area require hard to find analysts or costly to train and maintain. Thus, automated technologies assisting patent analysis are in great demand.

Patent analysis can be divided into two levels of analysis, macro level research of national or industrial analysis and micro level research of specific technology development or forecasting (Choi et al. 2007). Macro level analysis evaluates the major economical effect of technological innovations, technological development and competitiveness of countries (Grilliches 1990). At micro level, the focus is to identify technological development of specific areas/technologies, advantages and disadvantages of competitors, strategic planning of R&D activities, and patent data are analyzed to find the relation between companies and technologies (Haupt et al. 2006).

Many researchers have tried to identify indicators or determinants of patent value (Sapsalis et al. 2006). By using different types of data sets of patent data such as regional patent offices, particular sample, specific sector such as biotechnology, or particular company in a given country. Several researchers have studied patents and its relationship or effect on economy, technological innovations/development, or a country's competiveness (Grilliches 1990). It is important since on average only 1-3 patents out of 100 can generate significant financial returns. Although only a few patents have commercial success, most patents are developed by follow-up patenting into significantly important technologies (Ernst 1997). High value patents often have broad technical claims and a high citation index which increase the financial value of the company (Lerner 1995). Companies with a strong patent portfolio and conduct strategic patenting activities are more successful than other companies that remain inactive in the field of mechanical engineering and biotechnology sector (Ernst 1995, 2001; Austin 1993). Patent analysis can be effectively used for companies to gain competitive advantages at market place (Grilliches 1990). Moreover, patents are easily accessible throughout the world through databases in most countries (Lee et al. 2009).

Before issuing a patent at USPTO each patent document are given one or several patent classification based on invention, claims, and content (Tseng et al. 2007). These classifications are denoted as UPC (U.S. Classification) and IPC (International Patent Classification) these are given in most patent documents. According to Tseng et al. (2007), patent classifications are sometimes too broad or cannot meet the requirements for particular analysis. In this research, UPC and IPC are used for analysis and examples of UPC and IPC definitions for dental implant are shown in Table 2.

Table 1. Description of a typical patent analysis scenario adapted from Tseng et al. (2007).

A typical patent analysis scenario	
1. Task identification	Define the scope, concepts, and purposes for the analysis task
2. Searching	Iteratively search, filter, and download related patents
3. Segmentation	Segment, clean, normalize structured and unstructured parts
4. Abstracting	Analyze the patent content to summarize their claims, topics, functions, or technologies
5. Clustering	Group or classify analyzed patents based on some extracted attributes
6. Visualization	Create technology-effect matrices or topic maps
7. Interpretation	Predict technology or business trends and relations

Source: Tseng et al. (2007).

Table 2. Definition of patent classifications for dental implants

Patent classifications	
International Patent Classification (IPC)	
Class	**Dentistry; Apparatus or methods for dental hygiene**
A61C 8/00	Means to be fixed to the jaw-bone for consolidating natural teeth or for fixing dental prostheses thereon; Dental implants; Implanting tools
A61C 13/00	Dental prostheses; Making same (tooth crowns for capping teeth; dental implants)
U.S. Patent Classification (UPC)	
Class 433	**Dentistry**
Subclass 433/173	**By fastening to jawbone:** This subclass is indented under subclass 172. Subject matter wherein the denture is secured directly to the jawbone of the patient.
Subclass 433/174	**By screw:** This subclass is indented under subclass 173. Subject matter wherein the denture is secured to the jawbone by an elongated helically ribbed member

Source: USPTO and WIPO (2011)

2.4.1. Data in patent documents

A patent document contains items/details which can be divided into two groups, structured and unstructured data (Tseng et al. 2005). Structured data in patent documents are uniformthrough most patents such as patent number, filing date, inventor, and assignee. Unstructured data are defined such as free text of various length and content, claims, abstracts, or description of the invention. Patent analysis using structured information such as filing dates, assignees, and citations etc. have been in practice and literature for years (Ernst 1997; Lai & Wu 2005). The visualized results of structured data (patent number, filing date, etc.) is called patent graph and most use bibliometric data of patent documents to provide statistical results for patent analysis (Lee et al. 2009). Unstructured data (abstract, free text, etc.) are called patent maps. However, the general term patent maps can be used to describe both structured and unstructured data (Tseng et al. 2007). Patent maps are the visualization step in Table 1. Patent maps can be used for decision-making about future R&D directions (understanding patent relations and how patents are invented in the past), or predict technology/business trends (trend of major competitors in the same industry), and discover technological trends and opportunities as well as technological holes for future innovations (Tseng et al. 2007; Choi et al. 2007).

Bibliometrics is defined as the measurement of texts and information (Norton 2001). In general, most patent analysis utilize bilbiometric data (structured data) which explore, organize and analyze large amounts of data in order to identify patterns such as authors, technology field, citations, and so on (Daim et al. 2006). Although there are many items for analysis, one in particular has been employed more frequently, citation analysis. Patent citations or citation analysis are defined as the count of citations of a patent in subsequent patent, and thus citations per patent represents the relative importance of the patent (Lee et al. 2009). One possible reason can be as Sapsalis et al. (2006) point out is that citations analysis are closely associated with patent value (increase of financial value of a company). However, the scope of analysis using bibliometric data is easy to understand and to create but are subjected to limited access of the richness of information in patent documents since it only uses bibliometric fields (Lee et al. 2009). Text mining has been proposed as an alternative to analyze unstructured textual data in patent documents (Kim et al. 2008).

2.4.2. Limitations in patent analysis

There are certain limitations using patent analysis as indicator for forecasting technology development or business trends. First, not every company or organization patent their invention and Choi et al. (2007) mention that for example not all inventions meet the criteria made by patent offices and also some companies or industries rely on secrecy also it is a strategic decision not to patent an invention. Second, the results from patent analysis are interpreted differently across industries and companies, which results in inconsistent analysis. Third, patent laws changes over time which makes it difficult to analyze over time but recently companies are more inclined to file patents to mainly protect their invention from competitors (Choi et al. 2007).

2.5. Patent clustering

Recently, cluster analysis has become an important topic because of recent decade of advancement in data mining, increased computer power, and increased statistical software packages that include cluster analysis algorithms (Kettenring 2009). Given a set of documents, often there is a need to categorize documents into groups or clusters. For a small set of documents it can be done manually, on the other hand for a large set of documents the process will be time consuming and inefficient (Shahnaz et al. 2006). A patent document usually consists of a title, an abstract, claims, detailed description of the invention and bibliographic information. Moreover, all patent documents have manually assigned International Patent Classifications (IPCs) and if issued at USPTO it also consists of United States Patent Classification (UPC). Classification codes, IPC & UPC, are manually clustered by patent specialist or examiners. This type of classification is called supervised since it has predefined categories or topics for classification (Shahnaz et al. 2006). Unsupervised classification often deals with unstructured data. The goal is to organize and structure the unstructured data into groups or clusters based on the patterns of the collection itself (Dunham 2003). According to Trappey et al. (2010), patent documents with the same classification codes may be entirely different.

Clustering methodology is an important data analysis technique, which classify patterns of key phrases into categories based on the characteristics of relationship (Trappey et al. 2009). The main concept is to measure the similarity in data and categorize it to the most suitable cluster and maximize the similarity of specified variables within the same cluster, in other words, create a homogenous cluster. It is necessary that each patent document belonging to a cluster to be similar. The importance according to Almeida et al. (2007) is presence of high connectivity among these patent documents which is high association between objects.

Clustering methodology has been applied to numerous of different fields. For example, Taiwan Semiconductor Manufacturing Company, Ltd, use clustering analysis to detect errors in the manufacturing process, by isolating and separating failure symptoms and group suspicious process steps for evaluation by the process engineer (Kettenring 2009). It has also been used in predicting consumer behavior by creating shopping clusters of consumers' previous purchasing behavior or patterns, to forecast future shopping behavior (Kettenring 2009). A general clustering approach is shown in Figure 5.

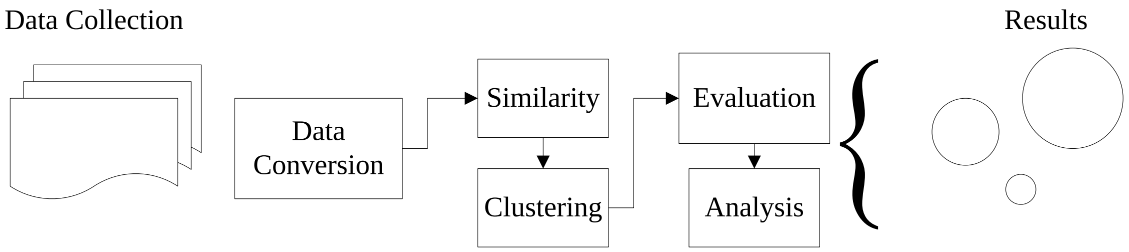


Figure 5. General process of clustering data. Adapted from Trappey et al. (2009)

<u>Patent technology clustering</u>

Patent technology clustering is a method to group similar or technology related patent documents into clusters rather than by UPC (Trappey et al. 2010). Patent technology clustering makes it possible to analyze the relationship between patent documents in specific domain technology and also the possibility to analyze patent or trends and development (Trappey et al. 2008). Patent technology clustering is derived from using key phrase correlation matrix as input and by applying the K-means algorithm (Trappey et al. 2010; Trappey et al. 2009). A more complete discussion on functions of K-means algorithm is provided by Han et al. (2011). Furthermore, the Root Mean Square Standard Deviation (RMSSTD) and R-Squared (RS) is used by Trappey et al. (2009) to find the optimal number of clusters in a set of data. RMSSTD is the standard deviation of all variables and represent the minimum variance in the same cluster therefore the value of RMSSTD should be as small as possible to gain optimal results. RS describe the maximum variance between different clusters and the value of RS should be as large as possible because RS is the sum of squares between different clusters divided by the total sum of squares for the set of data. A more detailed description of equations and functions of RMSSTD, RS and K-means are described by Trappey et al. (2009), Trappey et al. (2010) and Trappey et al. (2008).

Patent document clustering

Patent document clustering uses the correlation matrix generated from patent technology clustering as input in K-means algorithm (Trappey et al. 2010). Patent document clustering is a method that measures the internal relationship of the key points of the patent document and classifies patent documents based on the similarity of the technologies (Taghaboni-Dutta et al. 2009; Trappey et al. 2010). As a result it makes it easier for patent analyst to analyze the characteristics of patent documents in the clusters. This also solves the problem of patent classification systems (IPC and UPC) which may place the same code on patent documents which may be entirely different in technology (Taghaboni-Dutta et al. 2009). As shown in Table 3, the matrix is used as an input for patent document clustering.

Table 3. Key phrases and patent correlation matrix (Trappey et al. 2010)

	$Patent_1$	$Patent_2$	$Patent_3$	…	$Patent_n$
TC_1	$N_{1,1}$	$N_{1,2}$	$N_{1,3}$	..	N_{1N}
TC_2	$N_{2,1}$	$N_{2,2}$	…	…	N_{2N}
TC_3	$N_{3,1}$	…	…	…	N_{3N}
…	…	…	…	…	…
TC_n	…	…	…	…	N_{nm}

Source: Trappey et al. (2010)

2.6. Key phrase analysis

Key phrase extraction is useful for document or information retrieval, document clustering, summarization, text mining, and so on (Matsuo and Ishizuka 2003). Turney (2000) also point out a dozen useful applications with key phrase extraction for example, highlighting key phrases in text, document classification, text compression, or constructing human-readable text. Most information stored in databases is textual documents. Extracting key phrases makes it possible to determine which document is important and also identify the relation among several documents since it extracts relevant key phrases (Matsuo and Ishizuka

2003; Hammouda et al. 2005). According to Voorhees (1999), the majority uses statistical approaches for information retrieval (key phrase extraction) because of the assumption that two texts in the same topic use the same key phrases. Statistical approach measure the similarity of key phrases between textual documents. There are different approaches for key phrase extraction and the most commonly used are a lexical approach, natural language processing (NLP), or term frequency approach (Trappey et al. 2008). Hammouda et al. (2005) divide key phrases extraction algorithms into two categories: key phrases extraction that requires supervised learning and are applied for single documents, on the other hand, key phrase extraction on a set of documents are unsupervised and self-learning which discover rather than learning from examples, also known as knowledge discovery.

Research points towards that key phrases main goal are to represent the topics discussed in any text document (Turney 2000). Furthermore, Turney (2000) point out the relevance using key phrase extraction such as it enable the user quickly to determine if the key phrases are in the field of interest and it can be used for relevant indexing based on the key phrases. Key phrases extraction has been applied in many different fields, although mainly for summaries purposes (Turney 2000). For example Nenkova et al. (2006) studied the impact of automatic summarization systems based on key phrase extraction and its role in human summarization, the results showed that the key phrase frequency methodology used generated summaries comparable with state-of-the-art systems. Trappey et al. (2008) are using a hierarchy and semantic relationship concept to create a summarization system that uses key phrases to summarize any patent document based on the specific domain of the patent document.

2.6.1. Term frequency approach

The term frequency (TF) approach is based on the assumption that high frequent key phrases in a text document are more relevant to the concept of the content (Trappey et al. 2008). Robertson (2004) also points out that high frequency of a term represent a document better. Furthermore, in information retrieval of terms (key phrase), the most common terms are used in weighting schemes to represent text documents (Aizawa 2002). For example, Robertson and Sparck Jones (1976) study the relevance of weighting methods of key phrase using term frequency weighted with the inverse document frequency (TF-IDF). Trappey et al. (2008) uses a normalized TF-IDF to extract key phrases and phrases for clustering of patents.

The concept of TF-IDF is that it weight frequent key terms in a series of documents to determine its relevance. Therefore, frequent key terms in one document cannot represent a domain but frequent key terms in a series of document might represent the concept of the domain (Robertson and Sparck Jones, 1976).

The basic formula of IDF used by Robertson and Sparck Jones (1976) and Trappey et al. (2007) is expressed as:

$$idf_i = log_2\left(\frac{n}{df_i}\right) \tag{1}$$

where n is the total number of documents in the collection and df_i is the number of documents in the collection which containing term i. idf_i itself represent the inverse document frequency (IDF) of term i. Trappey et al. (2007) describe idf_i as a value of representation of term i and if idf_i becomes a significant high value, the term i can represent a specific document.

The weighting of key phrase using TF-IDF in text documents where TF are weighted in IDF is according to Trappey et al. (2007) expressed as:

$$w_{ik} = tf_{ik} \times idf_i \tag{2}$$

where w_{ik} is defined as weight of term i in document k of the collection, tf_{ik} is the number of term i that occurs in document k of the collection, and idf_i is the inverse document frequency of term i. Therefore, the highest value of w_{ik} equals the most frequent key phrase in a specific text document and are identified as the key phrase for any document k.

Furthermore, Trappey et al. (2008) normalize TF-IDF because of TF-IDF is a method that does not consider the difference of number of words in each document, therefore Trappey et al. (2010) applied a normalization of the weights frequency of key phrases by the number of words in each documents. According to Trappey et al. (2010), the normalized TF-IDF (NTF) can be expressed as following:

$$NTF = tf_{ik} \times \frac{\sum_{s=1}^{n} WN_s}{n} \times \frac{1}{WN_k} \tag{3}$$

where tf_{ik} is the number of term i that occurs in document k of the collection, WN_k is the words number of document k, and n is the total number of documents in the document collection.

2.6.2. Key phrase correlation matrix

The key phrase correlation matrix calculates the correlation of important key phrases (KP) in each patent document which is used to understand the logical link between concept and methodologies (Trappey et al. 2010). Trappey et al. (2010) describes the methodology of using TF-IDF and NTF to calculate the correlation between key phrases to create a key phrase correlation matrix using inner product of vectors expressed as:

$$\mathrm{Correlation}(KP_i, KP_j) = \frac{KP_i \cdot KP_j}{\|KP_i\|\|KP_j\|} = \frac{\sum_{k=1}^{n} w_{ik} \times aw \times w_{jk} \times aw}{\sqrt{\sum_{k=1}^{n} w_{ik}^2 \times aw^2 \times \sum_{k=1}^{n} w_{jk}^2 \times aw^2}} \quad (4)$$

where $KP_i = aw(w_{i1}, w_{i2}, \dots, w_{in})$ and $aw = \frac{\sum_{s=1}^{n} WN_s}{n \times WN_k}$ = average Word Number (WN). Trappey et al. (2010) use an algorithm of four stages. First, the algorithm transforms the patent document into a key phrases vector and analyzes the frequency of key phrases and phrases. Second, derive the key phrase vector by eliminating unnecessary key phrases and phrases. Third, the correlation values between key phrases are calculated using Equation (4). Fourth, the correlation coefficients are derived by the number of different key phrases occurring in each patent document. The key phrases correlation matrix is shown in Table 4.

Table 4. Key phrases correlation matrix

	KP_1	KP_2	KP_3	…	KP_n
KP_1	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	..	…
KP_2	$R_{2,1}$	$R_{2,2}$	…	…	…
KP_3	$R_{3,1}$	…	…	…	…
…	…	…	…	…	…
KP_m	…	…	…	…	…

Source: Trappey et al. (2010)

The key phrase correlation matrix is used as an input for patent technology clustering. Key phrase correlation matrix represents the technology in each patent document and thus it provide the internal relationship among patent documents instead of clustering patents according to classification codes such as UPC or IPC.

2.6.3. Key phrase and patent correlation matrix

In the key phrase and patent correlation matrix, the frequency (F_{nm}) of each key phrase (KP) appearing in each patent document is calculated as well as NTF, Rate (%) and NTFR. The Rate describes the percentage of KP_m occurring among $Patent_1$ to $Patent_n$. NTFR is the product of NTF and Rate which express the relevance of KP_m among the patent collection, shown in Equation (5). The key phrase, KP_m, is a representative phrase in the patent, $Patent_n$, if the frequency, F_{nm}, is large enough across $Patent_1$ to $Patent_n$, then KP_m is a representative phrase of $Patent_n$ (Trappey et al. 2010). The key phrase and patent correlation matrix is shown in Table 5.

$$NTFR = tf_{ik} \times \frac{\sum_{s=1}^{n} WN_s}{n} \times \frac{1}{WN_k} \times Rate \qquad (5)$$

$$Rate = \frac{\sum_{n=1}^{Patent_n} X_{nm}}{n}$$

If $F_{nm} = 0$; $X_{nm} = 0$

$F_{nm} > 0$; $X_{nm} = 1$

Table 5. Key phrases and patent correlation matrix

	$Patent_1$	$Patent_2$	$Patent_3$	…	$Patent_n$	NTF	Rate (%)	NTFR
KP_1	$F_{1,1}$	$F_{1,2}$	$F_{1,3}$	..	…	…	…	…
KP_2	$F_{2,1}$	$F_{2,2}$	…	…	…	…	…	…
KP_3	$F_{3,1}$	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
KP_m	…	…	…	…	F_{nm}	…	…	…

Source: Trappey et al. (2010)

2.6.4. Limitations in key phrases extraction of textual documents

Nasukawa and Nagano (2001) mentioned some issues using key phrases to represent a textual document. The problem is that textual documents are unclear because of natural language is ambiguity and same key phrase may have different meanings in the same textual document (Nasukawa and Nagano 2001). For example the word “watch” can represent a timepiece, to look, to observe or pay attention. Different words can also represent the same meaning, for example “laptop” and “notebook” or “cellular phone” and “mobile phone”.

2.7. Technology life cycle analysis

Life cycle analysis, as the name implies, is a straightforward methodology that assess all impact on a product or service, from initial extraction of raw material to the final output or disposal of the product (Ayres RU 1995). When companies invest R&D capital on technologies, it often depends on current life cycle stage of the technology (Haupt et al. 2007). According to Haupt et al. (2007) and Ernst (1997), patent documents inform us about technical development and the life cycle stage of an industry since patent documents contain core technology information. A patent or patentability of a technology is also one of the preconditions of the commercial potential of a technology. In addition to these information, patent document contain data about patent application date which inform us about the life cycle of different products, based on the technology, before it can start being commercialized (Haupt et al. 2007). The concept of technology life cycle is similar to product life cycles and can be divided into four stages: introduction, growth, maturity, and decline or saturation (Haupt et al. 2007; Trappey et al. 2010). Haupt et al. (2007) also point out that regardless of what reference factor is for technology life cycle or that the patent based life cycles starts earlier than product/sales based one, the principles can still be applied for technology life cycle as for product life cycle.

Several studies on technology life cycle based on patent document information show that an S-shaped curve can represent the technology life cycle. The S-shape curve include the four stages, introduction, growth, maturity and decline (Haupt et al. 2007). Andersen (1999) studied the S-curve with examples from the pharmaceutical industry. Trappey et al. (2010) studied the RFID technology in China and forecasted potential market and R&D opportunities. Another study by Trappey & Wu (2008) used S-curve analysis technique to evaluate short product life cycle products like electronics. The beginning of the life cycle, the introduction stage, of a new technology is the development of the scientific fundamental problems. These technical problems have to be solved in order to rapidly progress in technological advancement and during this period of time awaits radical innovations. At this stage, the patent applications are low but slowly increasing because during this period there is a lot of uncertainty and there are pioneer firms that are willingly to take the R&D risk (Haupt et al 2007; Trappey et al. 2010; Trappey & Wu 2008). During this stage the patent application per applicant is relatively high compared with other stages of the life cycle and this is because of the problems of new innovative technologies as well as the cost is too high for customers' acceptance or standardization of the product has not evolved yet. During the growth stage is when the fundamental technical problems have been solved and the market uncertainty has "vanished", many products is developed based on this technology, R&D risk decreases, and resulting in increase of patent applications (Haupt et al. 2007; Trappey et al. 2010). The growing number of patent application also decreases the patent application per applicant due to new competitors. The technology enters a mature stage when the number of patent applications is constant and there are now new features developed for this technology. Thereafter the technology enters the decline or saturation stage.

Patent activity is an important indicator of current technology life cycle and furthermore, Haupt et al. (2007) and Ernst (1997) have implemented this S-curve methodology on niche technologies such as pacemaker technology. Ernst (1997) proposed that all cumulative patent applications per year for a specific technology over a certain period of time can be plotted as S-curve and the different technology life cycle stages can be analyzed. An example of the principles of S-curve is shown in Figure 6.

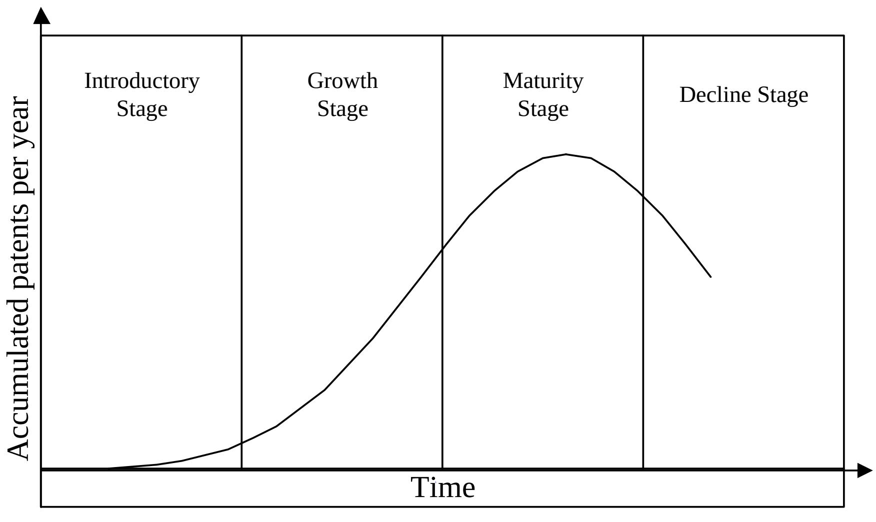


Figure 6. S-curve of TLC stages. X-axis represents a period of time and Y-axis represents accumulated patents over the time period (Adapted from Trappey et al. 2010).

The technology life cycle analysis is important for companies to evaluate the timing of R&D or other investment opportunities of technologies. It is strategically important to account for technology life cycle analysis when for example at the introduction stage. Companies should aggressively apply for patent families of their core invention (patent) to strengthen their position at the marketplace. If however, the technology is at the growth phase it is important for companies to search for core technologies in that field and develop their own applications. At the mature stage it is important to evaluate the boundaries of patents to avoid infringement issues or create alliances to trade IP. Finally the declining stage implies that new technology is replacing the old and is the beginning of a new technology life cycle (Trappey et al. 2010; Trappey and Wu 2008; Haupt et al. 2007; Ernst 1997).

2.8. Research framework

This research proposes a domain-specific patent ontology methodology for technology clustering and life-span analysis. The methodology steps are based on previous research by Trappey et al. (2010), as shown in Table 6. Trappey et al. (2010) referenced a modified ontology to extract key phrases from patent documents. The first step is to define a patent domain and select IPC and UPC in this domain. The second step is to collect domain specific patent documents from USPTO database. These steps are completed by using data mining software Intellectual Property Defense-based Support System (IPDSS) (Wheeljet, 2011).

Table 6. Methodology outline of this research

Method by Trappey et al. (2010)	**Method in this research**
1. Data preprocessing	1. Define domain (IPC & UPC)
2. Key phrase analysis (TF-IDF)	2. Data preprocessing
3. Key phrase correlation measure	3. Key phrase analysis (NTFR)
4. Patent technology clustering	4. Process ontology
5. Patent document clustering	5. Ontological sub-domain technology clustering
6. Lifecycle analysis	6. Life-span analysis of clusters

3. METHODOLOGY

This section describes the methodology development of this research and is divided into four parts patent domain definition, key phrase analysis process, processing domain-specific ontology, and technology life-span analysis of patent clusters. The IPDSS (Wheeljet, 2011) is used as a tool for data mining, key phrase extraction and clustering. Figure 7 and 8 shows the IPDSS software used in this research.

Figure 7. Login page of IPDSS

Figure 8. Workspace of IPDSS

3.1. Patent domain definition

The first step is to define a specific patent domain and select relevant UPCs or IPCs. As shown, in Figure 9. As described at the literature review section, patent under the same classification code may be entirely different in technology (Taghaboni-Dutta et al. 2010).

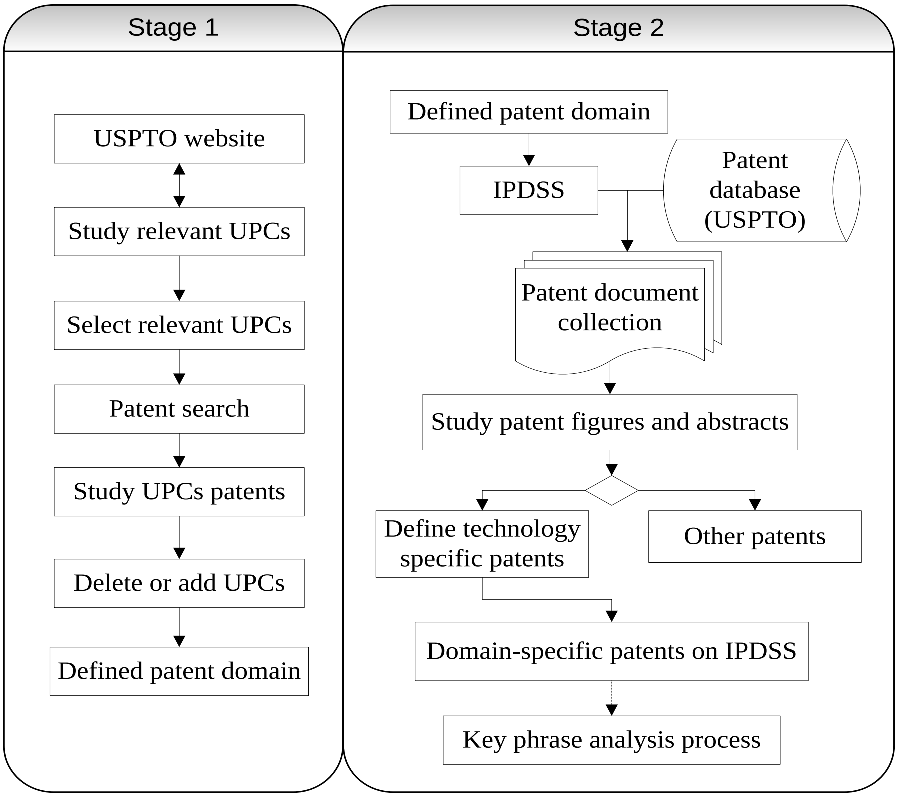


Figure 9. Processing of patent domain definition

Stage 1 is to select patent domain based on patent classifications following these steps:

1. Use USPTO or WIPO website and study UPC or IPC definitions for specific domain chosen. Patent classifications, UPC and IPC, are described at USPTO and WIPO, respectively (USPTO 2011; WIPO 2011).
2. Select relevant IPCs or UPCs. (Note: Use either IPC or UPC, not both).
3. Include 5 patent classifications to 15 patent classifications.
4. Search for 3-4 patents for each individual IPCs or UPCs on USPTO or WIPO.
5. Study those patents to determine if chosen IPCs or UPCs (from step 2) are relevant or not for chosen domain. Delete or add patent classifications to your domain.
6. Patent domain defined.

Stage 2 is to collect training domain-specific patent documents according to UPCs or IPCs. Thereafter define the technology specific patents and exclude other. Following these steps:

1. Use Intellectual Property Defense-based Support System (IPDSS) to download 150 training patents from USPTO if UPC is chosen (WIPO if IPC is chosen) (Note: The patents have to be according to patent classifications chosen).
2. Study patent figures and abstracts to define technology specific patents according to chosen IPCs or UPCs. Delete other patents. (Note: Patents under same classification code might represent different technology).
3. Key phrase analysis process of domain-specific patents

In this research, a dimension represents a domain, for example dental implant can also have dental implant tools and dental implant materials. One dimension includes several important classification codes to represent key concepts and technology. However, it requires limiting classifications to be technology specific. The choice of IPC or UPC depends on specific domain and technology.

The software IPDSS has a function to connect with USPTO database (or WIPO) to perform patent search and download patents to IPDSS database. IPDSS also automatically preprocess all patent documents into standard format which means that spaces between words and phrases is removed to automatically perform frequency count of words and phrases of each patent document. For each dimension, key phrases are separately extracted from dental patent documents to build a domain-specific ontology of dental implants.

3.2. Key phrase analysis process

The key phrase analysis process generates a list of frequent and important phrases from each patent document. These phrases are used to form a logical link between concepts. In this research, key phrases analysis, key phrase correlation matrix and key phrase and patent correlation matrix are derived using IPDSS which apply the methodology normalized term frequency – inverse document frequency (NTF), as shown in Figure 10. The following sections will discuss NTF, key phrase correlation measure, key phrase and patent correlation matrix.

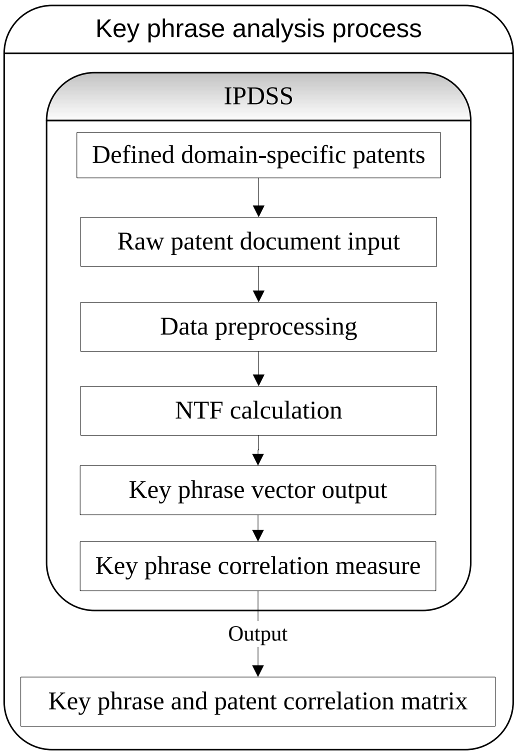


Figure 10. The key phrase analysis process

Normalized term frequency – inverse document frequency (NTF)

After IPDSS perform data preprocessing, the weight of each term is calculated using IPDSS that apply normalized term frequency – inverse document frequency (Trappey et al. 2008). As described at the literature review section, TF-IDF is a weighting method that weight frequent terms in a series of documents to represent text documents (Aizawa 2002). However, TF-IDF is a method that does not consider the difference of number of words in each document (Trappey et al. 2010). Therefore, normalization is applied to the weights frequency of key phrases by the number of words in each document. Robertson (2004) points out those high frequent terms represent a document better. The normalized TF-IDF (NTF) can be expressed as following:

$$NTF = tf_{ik} \times \frac{\sum_{s=1}^{n} WN_s}{n} \times \frac{1}{WN_k} \tag{6}$$

tf_{ik} = the number of term i that occurs in document k of the collection

WN_k = the words number of document k

n = the total number of documents in the document collection.

The frequency (F_{nm}), Rate, NTF, and NTFR are calculated and tabulated at Table 7 which is the output of the key phrase analysis.

Table 7. Key phrases and patent correlation matrix

	$Patent_1$	$Patent_2$	$Patent_3$	…	$Patent_n$	NTF	Rate (%)	NTFR
KP_1	$F_{1,1}$	$F_{1,2}$	$F_{1,3}$	..	…	…	…	…
KP_2	$F_{2,1}$	$F_{2,2}$	…	…	…	…	…	…
KP_3	$F_{3,1}$	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
KP_m	…	…	…	…	F_{nm}	…	…	…

The formulas are described as following:

$$F_{nm} = \sum_{i=1}^{F(KP_i)} KPF_i$$

$F(KP_i)$ = The number of key phrases (belonging to KP_m) that are included in $patent_n$

KPF_i = The frequency of the key phrase m (belonging to KP_m) of document j

The NTFR-value is expressed as

$$NTFR = NTF \times Rate \tag{7}$$

$$Rate = \frac{\sum_{n=1}^{Patent_n} X_{nm}}{n} \qquad \text{if } F_{nm} = 0;\ X_{nm} = 0 \text{ OR } F_{nm} > 0;\ X_{nm} = 1$$

Key phrase correlation measure

IPDSS calculates the correlation values between key phrases to create a key phrase correlation matrix using inner product vector expressed as:

$$\mathrm{Correlation}(KP_i, KP_j) = \frac{KP_i \cdot KP_j}{\|KP_i\|\|KP_j\|} = \frac{\sum_{k=1}^{n} w_{ik} \times aw \times w_{jk} \times aw}{\sqrt{\sum_{k=1}^{n} w_{ik}^2 \times aw^2 \times \sum_{k=1}^{n} w_{jk}^2 \times aw^2}} \tag{8}$$

where $KP_i = aw(w_{i1}, w_{i2}, \dots, w_{in})$ = the vector of key phrase i

$KP_j = aw(w_{j1}, w_{j2}, \dots, w_{im})$ = the vector of key phrase j

$$\mathrm{aw} = \frac{\sum_{s=1}^{n} \mathrm{WN_s}}{\mathrm{n} \times \mathrm{WN_k}}$$ =average Word Number (WN)

$$w_{ik} = tf_{ik} \times idf_i \tag{9}$$

tf_{ik} = the number of term i that occurs in document k of the patent collection

df_i = the number of documents in the collection which containing term i

$$idf_i = log_2\left(\frac{n}{df_i}\right) \tag{10}$$

n = the total number of documents in the collection

First, the algorithm transforms the patent document into a key phrases vector and analyzes the frequency of key phrases. Second, derive the key phrase vector by eliminating unnecessary key phrases. Third, the correlation values between key phrases are calculated using Equation (11). Fourth, the correlation coefficients are derived by the number of different key phrases occurring in each patent document. The correlation coefficient is calculated according to the formula below:

$$\mathrm{Correlation_k}\left(\mathrm{KP_i}, \mathrm{KP_j}\right) = \frac{\sum_{k=1}^{n} \mathrm{Rev}\left(\mathrm{KP_i}, \mathrm{KP_j}\right)}{\mathrm{n}} \tag{11}$$

where $\mathrm{Correlation_k}\left(\mathrm{KP_i}, \mathrm{KP_j}\right)$ = the correlation value of key phrase i and key phrase j in document k

n = the total number of documents in the patent collection

After the correlation coefficient is calculated, it can be shown as the key phrases correlation matrix in Table 8. The frequency is calculated for all terms and KPf_i is used for the frequency of key phrase KP_i in the document. RPf_{ij} is used to represent the frequency of related phrases RP_{ij}. The correlation of RP_{ij} and KP_i are listed as R_{ij}. The final frequency of KP_i can be calculated as following:

$$KPF_i = KPf_i + \sum_{j=1}^{k} RPf_{ij} \times R_{ij} \tag{12}$$

A vector is created after all the KPF is calculated for all key phrases and are listed as following:

$$[KPF_1, KPF_2, \ldots, KPF_n] \tag{13}$$

This vector is used as the input of patent technology clustering and patent document clustering.

Table 8. Key phrases correlation matrix

	KP_1	KP_2	KP_3	…	KP_n
KP_1	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	..	…
KP_2	$R_{2,1}$	$R_{2,2}$	…	…	…
…	…	…	…	…	…
KP_m	…	…	…	…	R_{mn}

Source: Trappey et al. (2010)

3.4. Processing domain-specific ontologies

The domain-specific ontology is build by using Microsoft Visio 2007 (MS Visio) as a visualization tool for transferring domain-specific ontological schema (Huang et al. 2008). Ontology can be visualized as a pyramid and on the top of the pyramid represent the domain concept. In this research, the ontology structure is based on RFID ontology tree in Figure 4 and top 50 NTFR-values of key phrases in the key phrase matrix (from the key phrase analysis process) are chosen to build the ontology. The processing of the ontology is described by the following steps and the overview processing is shown in Figure 11.

Step 1: Organize patents with the same patent classification codes

This step utilizes the key phrase matrix output from the key phrase analysis process. First, patents with same classification, for example $Patent_1$, $Patent_{6-10}$, and $Patent_{23-28}$ has the same patent classification are grouped together. Second, analyze the frequency (F_{nm}) of each key phrase (KP_n) of the key phrase matrix and determine which KP_n is expressed in which UPC. This enables an overview of which classifications uses the same key phrases and this can be tabulated as shown in Table 9.

Step 2: Map classification codes – TYPE I ontology

From previous step, a new matrix is constructed, as shown in Table 10, to visualize which key phrase is expressed for each UPC. The Type I ontology is constructed using MS

Visio, as shown in Figure 11, the key phrases are placed out to visualize common phrases among patents. Key phrases that are expressed in the same number of UPCs, for example UPC_1, UPC_2, and UPC_3 are colored using your own defined coloring scheme and a different color for UPC_1 & UPC_2. It helps R&D engineers to visualize common phrases in this domain by coloring scheme of key phrases.

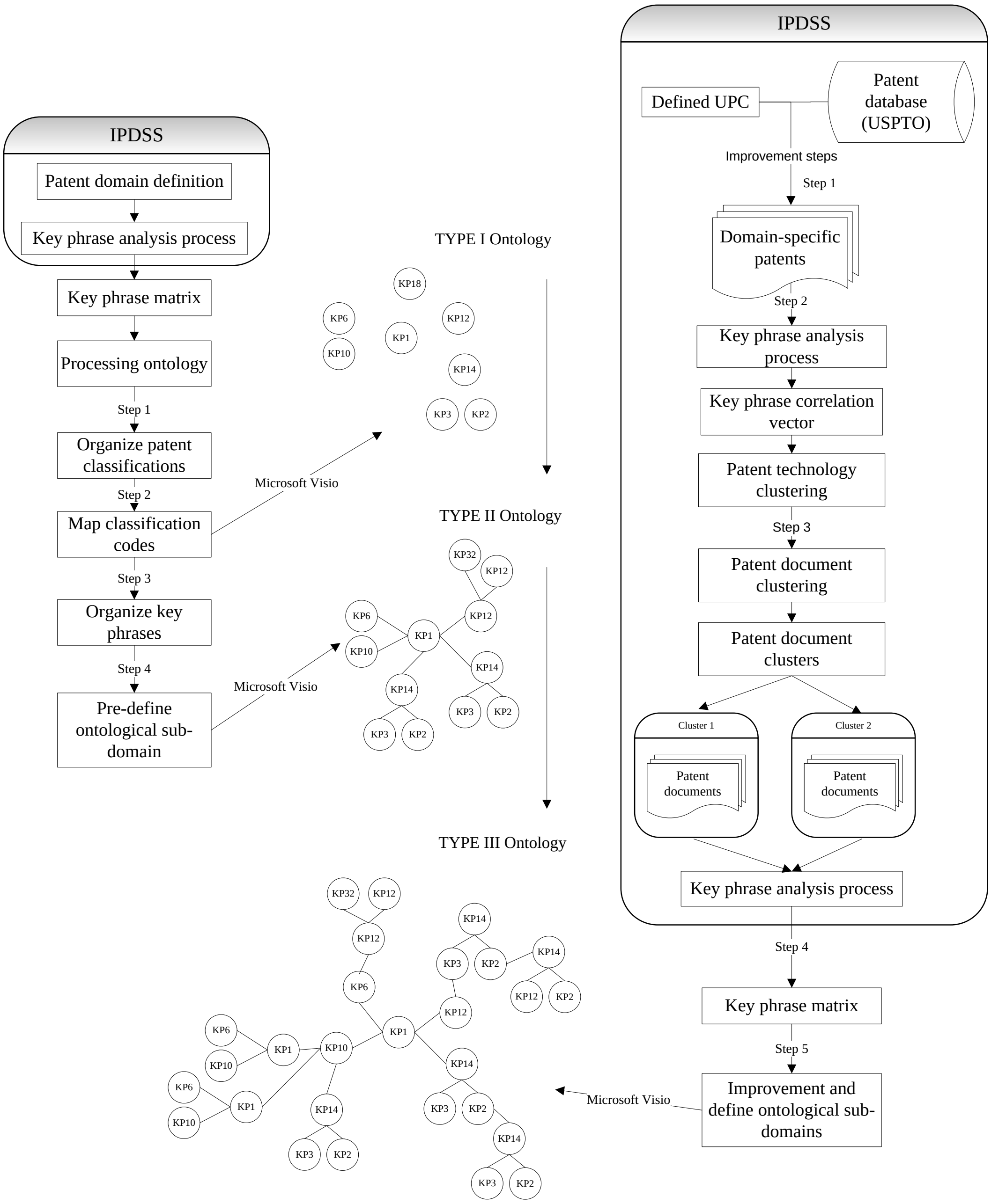


Figure 11. Processing of ontology

Step 3: Organize key phrases

The key phrases from Table 10 are organized and grouped together according to their relationship and logical link, as shown in Table 11. Online dictionary WordNet (Princeton

University, 2011) and studying patent documents are used to understand the meaning of the key phrases. This procedure is based on personal experience and interpretation of words. The goal is to create 4-5 large groups of key phrases. For example, in Table 11, KP_3, KP_2, and KP_6 are one group. The first draft of ontology is improved using the group key phrases to provide better concept relationship.

Table 9. Patent and UPC matrix

	$Patent_1$	$Patent_2$	$Patent_3$	$Patent_4$	…	$Patent_n$	NTF	Rate (%)	NTFR
UPC	UPC_1	UPC_1	UPC_1	UPC_2	…	UPC_z			
KP_1	$F_{1,1}$	$F_{1,2}$	$F_{1,3}$	..	..	…	…	…	…
KP_2	$F_{2,1}$	$F_{2,2}$	…	…	…	…	…	…	…
KP_3	$F_{3,1}$	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
KP_{50}	…	…	…	…	…	F_{n50}	…	…	…

Note: UPC_z = the UPC code of each patent

Table 10. Key phrase and UPC matrix

KP_1	UPC_1	UPC_2	UPC_3	..	…
KP_2	UPC_1	UPC_2	…	…	…
KP_3	UPC_1	…	…	…	…
…	…	…	…	…	…
KP_{50}	…	…	…	…	UPC_n

Note: Each column has the same classification code (IPC or UPC)

Table 11. Key phrase organization matrix

KP_3	UPC_1	…	…	…	…
KP_2	UPC_1	UPC_2	…	…	…
KP_6	UPC_1	UPC_2	UPC_3	..	…
…	…	…	…	…	…
KP_{50}	…	…	…	…	UPC_n

Note: Each column has the same classification code (IPC or UPC)

Step 4: Pre-define ontological sub-domain – TYPE II ontology

The final step is to pre-define the key phrase groups in Table 11 by studying phrases, patent classification definitions, and patent technology to assign appropriate definition of the ontological sub-domains. The processing of these steps is shown in Figure 12.

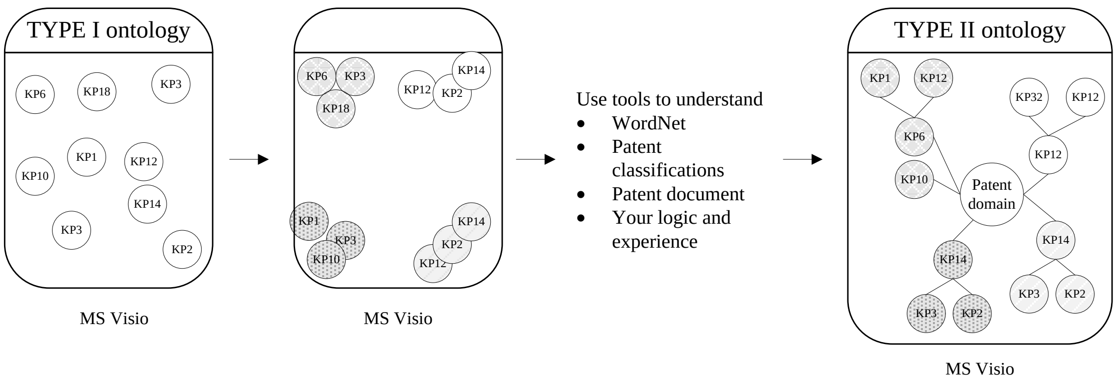


Figure 12. Processing TYPE II ontologies

The following steps describe the construction of Type II ontology using MS Visio:

1. Use MS Visio to group key phrases according to groups in Step 3 for Type I ontology
2. Start with the patent domain at the center in MS Visio.
3. Use WordNet, patent classifications, patent document, and your logic to determine which key phrases are strongest associated with the patent domain.
    a. Start to link key phrases from the center (patent domain) and outwards using MS Visio. Ontology is hierarchical, use your logic and link the phrases of the

pre-defined sub-domains. (Note: Linking key phrases to create are reliant on patent domain, technology, and understanding which requires study of patent documents and domain).

b. Use a color scheme to color the key phrases of pre-defined ontological sub-domains using MS Visio.

<u>Improvement of domain-specific ontologies</u>

This stage use methods to improve the Type II ontology, as shown in Figure 13, following steps describe the process to create Type III ontology.

Step 1: New domain-specific patents

Collect 50 new domain-specific patent documents from USPTO based on patent domain definition and use IPDSS to preprocess the data and carry out key phrase analysis process, patent technology clustering, and patent document clustering.

Step 2: Patent technology clustering

The correlation matrix derived from the key phrases correlation analysis in IPDSS is used as input for technology clustering. A feature of patent technology clustering is to discover the relationships of patents. Since the key phrases represent the concept or technology for each patent document, the key phrase correlation matrix and key phrases extracted are important for the key phrase collected. Patent technology clusters are generated by applying K-means algorithm of the key phrase correlation matrix. This technique can help researchers to select technology clusters to analyze, however, in this research it is used as an input for patent document clustering which is describe in the next step.

Step 3: Patent document clustering

The vector output of patent technology clustering is derived and used as an input for patent document clustering. A matrix is constructed as input for patent document clustering, as shown in Table 12. Patents under the same classification code can be entirely different and patent document clustering derives the internal relationship based on technologies. The output of this method is similar patens are clustered together to create a homogenous cluster. Therefore it is important and useful for researchers want to group technologies for analysis. Patents are clustered according to the formula below.

$$N_{ij} = \sum_{m=1}^{N(KP_i)} KPF_m \qquad (16)$$

$N(KP_i)$ = The number of key phrases (belonging to TC_m) that are included in $patent_n$

KPF_m = The frequency of the key phrase m (belonging to TC_m) of document j

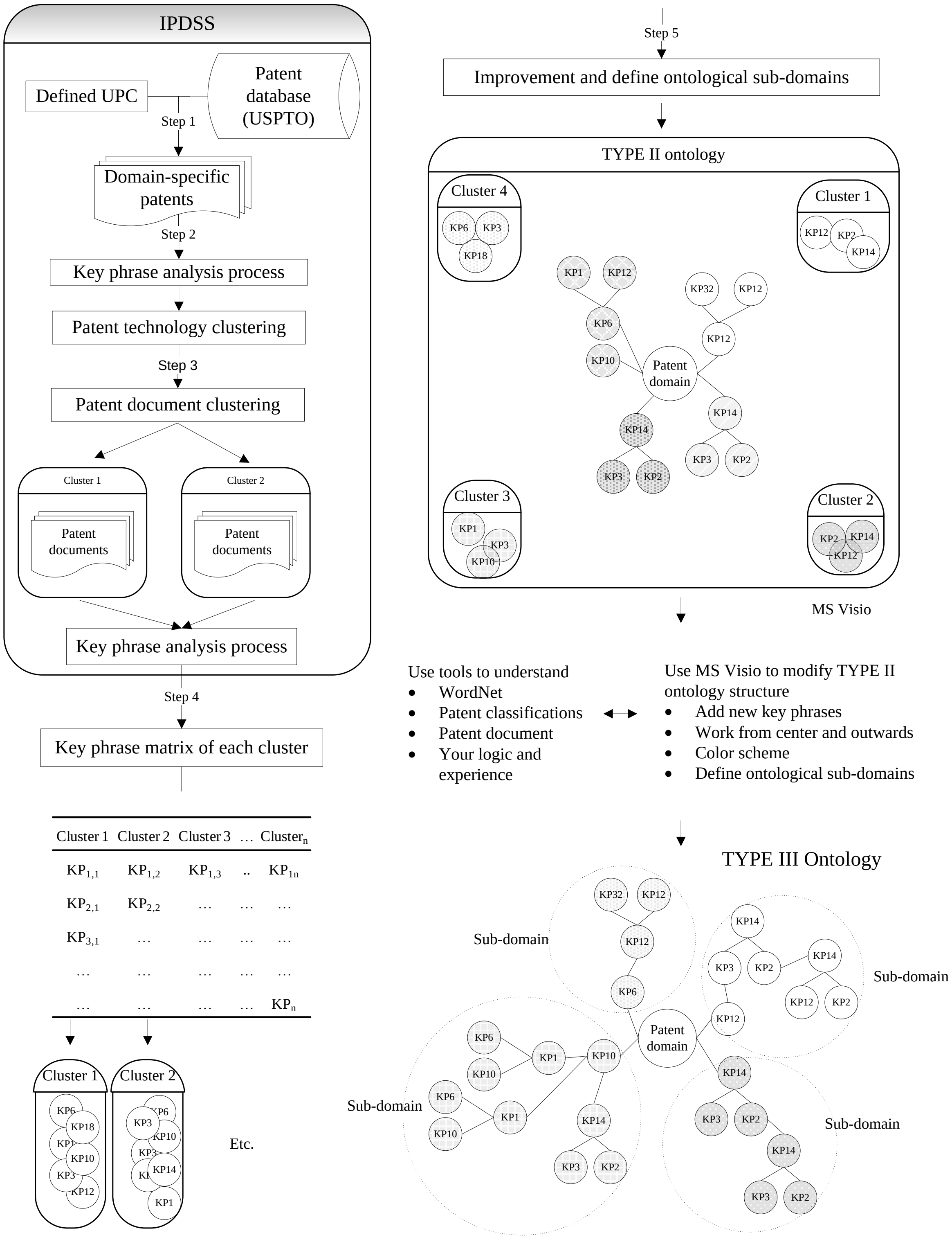


Figure 13. Improvement steps of domain-specific ontologies

Step 4: Key phrase extraction of document clusters

The patent document clustering creates clusters of the new patent documents according to step 3. The patent document clusters are separately subjected to the key phrase analysis process using IPDSS which applies NTFR and the top 15 NTFR-valued key phrases are extracted, tabulated as in Table 13. The key phrases are used to improve the pre-defined ontological sub-domain phrases.

Table 12. Key phrases and patent correlation matrix

	$Patent_1$	$Patent_2$	$Patent_3$	…	$Patent_n$
TC_1	$N_{1,1}$	$N_{1,2}$	$N_{1,3}$	..	N_{1n}
TC_2	$N_{2,1}$	$N_{2,2}$	…	…	…
…	…	…	…	…	…
TC_m	…	…	…	…	N_{nm}

Table 13. Key phrases of each cluster

Cluster 1	Cluster 2	Cluster 3	…	$Cluster_n$
$KP_{1,1}$	$KP_{1,2}$	$KP_{1,3}$	..	KP_{1n}
$KP_{2,1}$	$KP_{2,2}$	…	…	…
…	…	…	…	…
…	…	…	…	KP_n

Step 5: Modification and define ontology sub-domain – TYPE III ontology

The key phrases collection of each individual cluster in Table 13 is used to compare with key phrases in the Type II ontology. The following steps describe the procedure.

1. Use MS Visio – add the key phrases from Table 13 to the Type II ontology
    a. Note: Do not link phrases yet, group phrases according to its clusters
2. Choose each cluster with key phrases from Table 13 to compare with the Type II ontological sub-domains.
    a. For example, key phrases by key phrase from cluster 1 in Table 13 are used to compare key phrase by key phrase of the Type II ontology. The more matching phrases, the better the cluster represent that sub-domain.
    b. Assign the best matching cluster to the Type II ontological sub-domains
3. Use WordNet, patent classifications, patent documents and you logic to understand as well as determine if these clusters from Table 13 are relevantly grouped with the Type II ontology
    a. Use MS Visio to rearrange key phrases of Type II ontology and modify the structure of the pre-defined sub-domains (if it makes sense) on the Type II ontology
    b. Use Ms Visio to start linking new key phrases from each assigned clusters in the sub-domains of the Type II ontology. Work from center and outwards. Try to create sub-domains and from each sub-domain create hierarchical tree structure.

       Comments: The pre-defined sub-domains from Type II ontology can be deleted and new sub-domain definition is created. New key phrases from Table 13 are also added and linked and the shared key phrases from other sub-domains. Shared key phrases are usually the most common key phrases in one patent domain which can help engineers to understand the domain concept better.
    c. Type II ontology structure is modified and colored with a color scheme according to the sub-domain definition (next step)
4. Define the ontological sub-domains which depends on previous step and each sub-domain (depends on cases to case) can be for example describing the main components of a technology and can be separated in several major parts.

3.5. Processing life-span analysis in patent clusters

In this research, the procedure of clustering technology and life-span analysis is done by a case study. In this research, ontology is used to cluster patents according to its sub-domain-concepts by assigning each patent individually to each ontological sub-domain-cluster, as shown in Figure 14. Key phrases from each patent and compare with each sub-domain-cluster of the ontology which includes key phrases that are considered to be key concepts.

Step 1: Test patents

Thirty new domain test patents are downloaded from USPTO and processed in IPDSS. These test patents are new and have not been used in training the system, build or improved the ontology. Patent classifications on these patents are not extracted restricted.

Step 2: Key phrase analysis process

IPDSS apply NTFR-methodology to analyze key phrases of the patent documents and the output is key phrase matrix. The key phrase matrix lists the frequencies of all key phrases for each patent.

Step 3: Ontological sub-domain clustering

The list of key phrases for each patent is compared with the sub-domains of the Type III ontology. Key phrase by key phrase is compared with the sub-domain key phrases. The patents are assigned to that specific sub-domain if the key phrases describe the concepts and relationship of the sub-domain ontology most consistent. The patents assigned to each sub-domain are clustered together and this is called ontological sub-domain clustering. The clusters are named after the sub-domain definition.

Step 4: Life-span analysis

For each ontological sub-cluster, the age of the patent is calculated from the filing date as a starting date, not issuing date, and up to today. Patents are protected from the filing date and when granted it is called issue date. It can take up to two or three years before it is issued. The average age is the sum of each patent age divided by the number of patents in the sub-domain cluster and an example of the information is shown in Table 14. Average age is calculated according to the formula below:

$$Average\ age = \frac{\sum Age}{n}$$

where n = the total number of patents in a sub-domain

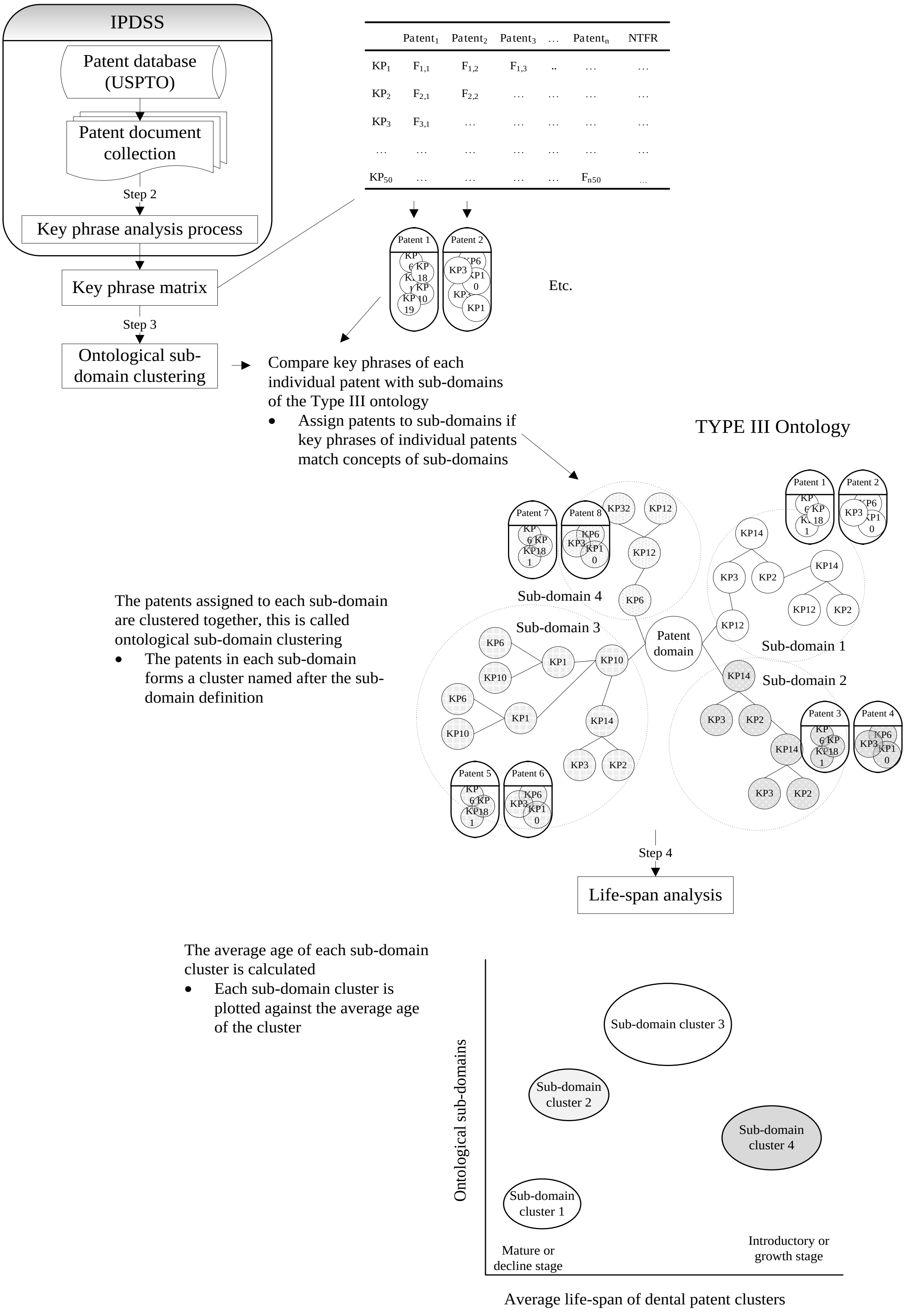


Figure 14. Processing of life-span analysis

Table 14. Patent information and average age

Sub-domain 1

Patent No.	Patent title (PT_1)	UPC	Filing date	Age
P_1	PT_1	UPC_1; UPC_2; etc.	Month, Year	Age
P_2	PT_2	UPC_1; UPC_2; etc.	Month, Year	Age
			Average age	AA

The average age of each cluster is plotted against the ontological sub-domain clusters, Figure 15 illustrate the analysis of potential emerging or declining clusters depending on its average age. The size of each bubble represents the number of patents, Y-axis is ontological sub-domain clusters and X-axis is the average age starting from 0-20 years (from right to left on the X-axis). Cluster 4 on Figure 15 represents a young cluster of a specific ontological sub-domain, in other words, a specific sub-domain technology in dental implants. This mapping method allows researchers to explore which sub-clusters have potential for further development or which sub-clusters are poor and soon to be outdated.

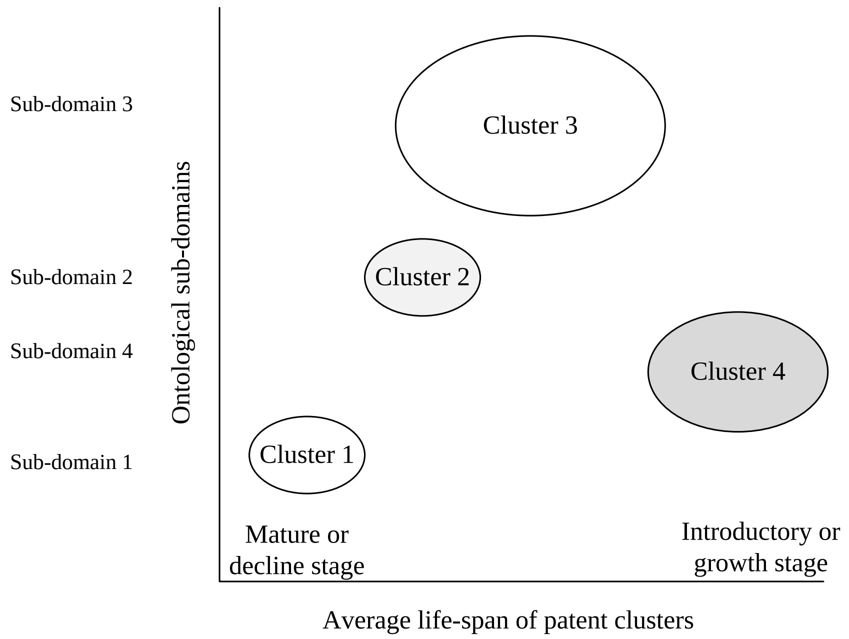


Figure 15. Proposed life-span analyses of dental implant patent clusters

In this research, the proposed analysis is to study the lifespan of each patent cluster to determine which clusters are outdated or growing. By mapping clusters using ontology as a dimension in the analysis can demonstrate clearer picture of clusters that are outdated or growing.

4. CASE STUDY AND ANALYSIS

This section includes five parts that describes the results obtained for the case study of dental implant dimension. Part one describes the training patent samplings and boundaries. Part two describes the key phrase and patent correlation matrix obtained and the key phrases defined. The third part describes the key phrases used to build dental implant ontology. Part four introduces the case study of dental implant ontology obtained, validation and the modification of the ontology. The final part describes the results obtained using ontological sub-domain clustering and analyzing the life-span of these clusters.

4.1. Dental patent documents samples

A total of 400 training patent documents were downloaded from USPTO. By defining patent classification boundaries, deleting irrelevant patents and only including relevant patent classifications, the original total number of training patents decreased to 304. The list of samplings and classifications boundaries is shown in Table 15. The difference in number of patents in each dimension and patent classification is because domain-specific technologies may have different patent classification code (IPC and UPC). Taghaboni-Dutta et al. (2009) also indicate that patents may be place under the same code but are entirely different in technology. Difference in number of patents in each classification code may cause weighting segmentation in key phrase analysis, in other words, some domain key phrases might have higher statistical significance by the large number of patents in a classification code. The focus of the case study is ontological sub-domain clustering and life-span analysis for dental implant therefore more dental implant patents to create a consistent domain-specific ontology for this dimension. Tools and materials ontologies were created for future research.

4.2. Dental implant technology key phrase analysis processing results

The key phrase and patent correlation matrix is derived for each dimension of dental implants. Top 50 key phrases are chosen on a chronological order of highest NTFR-value since high frequent key phrases are considered to be more relevant to the concept of the content as Trappey et al. (2008) proposes. Table 16 shows a partial key phrase and patent correlation matrix with the top 28 key phrases and four different patents with frequency values for each key phrase in each patent. Key phrases extracted from these training patents match most of the dental implants main components listed on Astra Tech Dental (2011) and Free Dental Implant (2011). For example, abutment (support for crown) or healing cap (covers abutments) is both listed on the matrix. The UPC 433/174 is described as “fastening implants to the jawbone by screw” and from Table 16 phrases listed are jawbone, threads, screw, hole, etc. which conform with the UPC. Another example is UPC 433/172 – holding or positioning denture in mouth, Table 16 list phrases such as embodiment, bore, implant fixture, etc. Patents with same IPC or UPC may not express same key terminologies which is a reason to include patents from several different classifications to create ontology that capture the main concepts of that domain. The partial key phrase matrix is shown in Appendix 1.

Table 15. List of patents in each classification and each dimension

Classifications and samplings

Dental implant		**Dental implant tool**		**Dental implant material**	
Patent classifications	Number of patents	Patent classifications	Number of patents	Patent classifications	Number of patents
433/173	97	433/173	28	A61C5/08	9
433/174	24	433/141	18	A61C8/00	12
433/175	4	433/163	3	C03C8/00	1
433.176	4	433/144	1	A61C13/225	14
433/172	13	433/172	14	A61C5/00	2
433/201.1	5	433/177	1	A61C13/00	16
433/169	7	433/174	3	A61C13/275	1
433/17	2	433/176	1	A61C13/20	3
		433/215	4	A61K6/083	2
		433/76	1		
		433/7	2		
		433/27	2		
		433/32	1		
		433/50	1		
		433/75	8		
Total patents	156		88		60

Table 16. Part of dental implant key phrase and patent correlation matrix without NTF, Rate, and NTFR values

Key phrase/Patent No.	**US6312260**	**US6039568**	**US5297963**	**US5362235**
UPC	**433/174**	**433/175**	**433/172**	**433/172**
KP1: Implant	177.46	165.01	51.01	49.2
KP2: Dental	29.09	84.28	25.7	23.74
KP3: Dental Implant	20.65	73.01	18.98	20.71
KP4: Bone	7.17	21.37	9.49	10.79
KP5: Screw	40.04	8.31	16.61	32.8
KP6: Abutment	0	26.12	34.4	71.64
KP7: Threaded	42.15	14.25	24.91	26.33
KP8: Bore	14.75	0	32.03	26.76
KP9: Prosthesis	0	4.75	0	0
KP10: Cap	140.37	4.75	138.01	91.92
KP11: Healing	141.63	7.72	141.97	95.38
KP12: Root	0	3.56	12.65	6.9
KP13: Tissue	6.32	4.15	11.86	11.22
KP14: Healing cap	125.61	4.75	133.27	88.47
KP15: Fixture	0	0	45.87	44.88
KP16: Implants	5.48	9.5	0	0
KP17: Embodiment	13.07	0	0	0
KP18: Threads	3.37	8.31	14.24	8.2
KP19: Cavity	0	0	6.72	7.77
KP20: Hole	5.48	8.31	0	6.04
KP21: Crown	0	0	0	0
KP22: Jaw	5.06	0	8.7	9.93
KP23: Jawbone	18.13	20.18	0	0
KP24: Device	0	5.94	0	0
KP25: Implant fixture	0	0	35.59	32.37
KP26: Prosthetic	0	0	0	0
KP27: Artificial	0	0	0	0
KP28: Thread	4.22	0	0	0

Define sub-domain of dental implant key phrase and patent correlation matrix

The key phrases are sorted and organized to be group together in four to five large groups. An example dental implant dimension is shown in Table 17, where key phrases are logically grouped to demonstrate the concept of this dimension. Even though, for example, “dental prosthesis” is not expressed by some UPC, it is still grouped with “dental” and “implant” because it is more likely to be associated with that sub-domain more than in the group of “screw” and “threads” (separated by a line in Table 17). By mapping the UPCs (or IPCs) for each dimension will give an overview of which key phrases are most fitted to which group which also depends if several classification codes are expressed by the same key phrases or not. The grouping creates sub-domains for the ontology. It is a form of clustering. The complete defined sub-domain key phrase matrix is shown in Appendix 2.

Table 17. Part of dental implant sub-domain key phrase matrix

Key phrase/UPC						
KP1: Implant	433/173	433/174	433/172	433/169	433/175	433/201.1
KP2: Dental	433/173	433/174	433/172	433/169	433/175	433/201.1
KP3: Dental implant	433/173	433/174	433/172	433/169	433/175	433/201.1
KP49: Dental implants	433/173	433/174	433/172	433/169	433/175	433/201.1
KP16: Implants	433/173	433/174	433/172	433/169	433/175	433/201.1
KP17: Embodiment	433/173	433/174	433/172	433/169	433/175	433/201.1
KP34: Attachment	433/173	433/174	433/172	433/169	433/175	433/201.1
KP39: Attached	433/173	433/174	433/172	433/169	433/175	433/201.1
KP27: Artificial	433/173	433/174	433/172	433/169		433/201.1
KP9: Prosthesis	433/173	433/174	433/172	433/169	433/175	433/201.1
KP26: Prosthetic	433/173	433/174	433/172			433/201.1
KP38: Dental prosthesis	433/173	433/174	433/172	433/169		433/201.1
KP24: Device	433/173	433/174		433/169	433/175	433/201.1
KP44: Teeth	433/173	433/174	433/172	433/169	433/175	433/201.1
KP6: Abutment	433/173	433/174	433/172	433/169	433/175	433/201.1
KP5: Screw	433/173	433/174	433/172	433/169	433/175	
KP7: Threaded	433/173	433/174	433/172	433/169	433/175	433/201.1
KP18: Threads	433/173	433/174	433/172	433/169	433/175	
KP28: Thread	433/173	433/174	433/172	433/169		
KP45: Titanium	433/173	433/174	433/172	433/169	433/175	433/201.1

4.3. Dental implant technology key phrases

The list of top 50 key phrases is extracted for each dimension and Table 18 shows the top 15 key phrases in dental implants, tools, and material. Many key phrases are the same although different dimensions, however it was expected that the top 10 key phrases are used to describe components features of dental implants and can be considered as the fundamental terminologies. However, key phrases of implants and tools share many of the same phrases which indicate an overlap in sub-domains and require more detailed tool ontology to separate the concepts apart. As Turney (2000) mentioned, key phrases enable the user to quickly determine if the key phrases are in the field of interest and can be used for relevant indexing, in this research it is used to build an ontology. The complete top 50 key phrases are shown in Appendix 1.

Table 18. Part of key phrase matrix for each dimension

	Implants	**Tools**	**Material**
Number of patents	156	88	60
Patent classifications	8 UPCs	15 UPCs	9 IPCs
Top 15 key phrases	KP1: Implant	KP1: Implant	KP1: Dental
	KP2: Dental	KP2: Dental	KP2: Teeth
	KP3: Dental implant	KP3: Bone	KP3: Crown
	KP4: Bone	KP4: Screw	KP4: Porcelain
	KP5: Screw	KP5: Dental implant	KP5: Resin
	KP6: Abutment	KP6: Abutment	KP6: Weight
	KP7: Threaded	KP7: Healing	KP7: Implant
	KP8: Bore	KP8: Cap	KP8: Glass
	KP9: Prosthesis	KP9: Healing cap	KP9: Prosthesis
	KP10: Cap	KP10: Device	KP10: Bone
	KP11: Healing	KP11: Bore	KP11: WT
	KP12: Root	KP12: Prosthesis	KP12: Particles
	KP13: Tissue	KP13: Threaded	KP13: Plate
	KP14: Healing cap	KP14: Teeth	KP14: Powder
	KP15: Fixture	KP15: Distal	KP15: Temperature

4.4. Dental technology ontology

The proposed life-span analysis of dental implant patents uses ontology as a variable for mapping clusters. The key phrases of each dimension are already grouped in previous step as shown in Table 17 and these groups represent the sub-domains of the ontology. The ontology in this research is an adapted version of Figure 4. The ontology of dental implant, shown in Figure 16, uses only phrases from key phrase matrix of dental implants to link phrases based on their concepts and relationships, using Microsoft Visio as a building tool. As Huang et al. (2008) describes ontology as a model that includes concepts and relationships in a specific domain that reflects the reality of the world. Thus, patent document contain detailed information about research results that are in complex technical and legal terms (Tseng et al. 2007). It is therefore more consistent to extract data from patents to build a domain-specific ontology for analysis. As Robertson (2004) and Trappey et al. (2008) suggests, high frequent key phrases in a text document are more relevant to the concept of the content and represent it better. Moreover, the key phrase analysis (NTF) and the correlation methodology applied in this research search for high frequent key terms in a series of documents to represent the concept of the domain better as Robertson and Sparck Jones (1976) proposes. Building ontology based on patent data is uncommon and for dental implant, according to my research up to date, it has not yet been explored. Therefore, it can be a huge interest in analyzing clusters using ontology based on dental implant patents. The ontology in Figure 16 shows four preliminary sub-domains of dental implant dimension and is named as geometry, implant fixture, biological, and dental components. The names of sub-domains are based on the phrases and should describe the concepts of the sub-domain. The ontology is divided into sub-domains to separate concepts in the dental implant domain, making it more specified. The list of key phrases is shown in Appendix 1 and the ontologies for dental implant tools and material are shown in Appendix 3.

Ontology is often built by domain experts and it is often subjective (Trappey et al. 2010). In this research, part of building ontology is subjective since linking concepts and phrases are based on the help from WordNet (Princeton University 2011) and opinion of the researcher of this report. However, constructing a domain-specific ontology in the dental implant area based on patent data that objectively extract phrases using computer software creates an ontology that is more relevant for this research and the analysis of dental implant patents clusters.

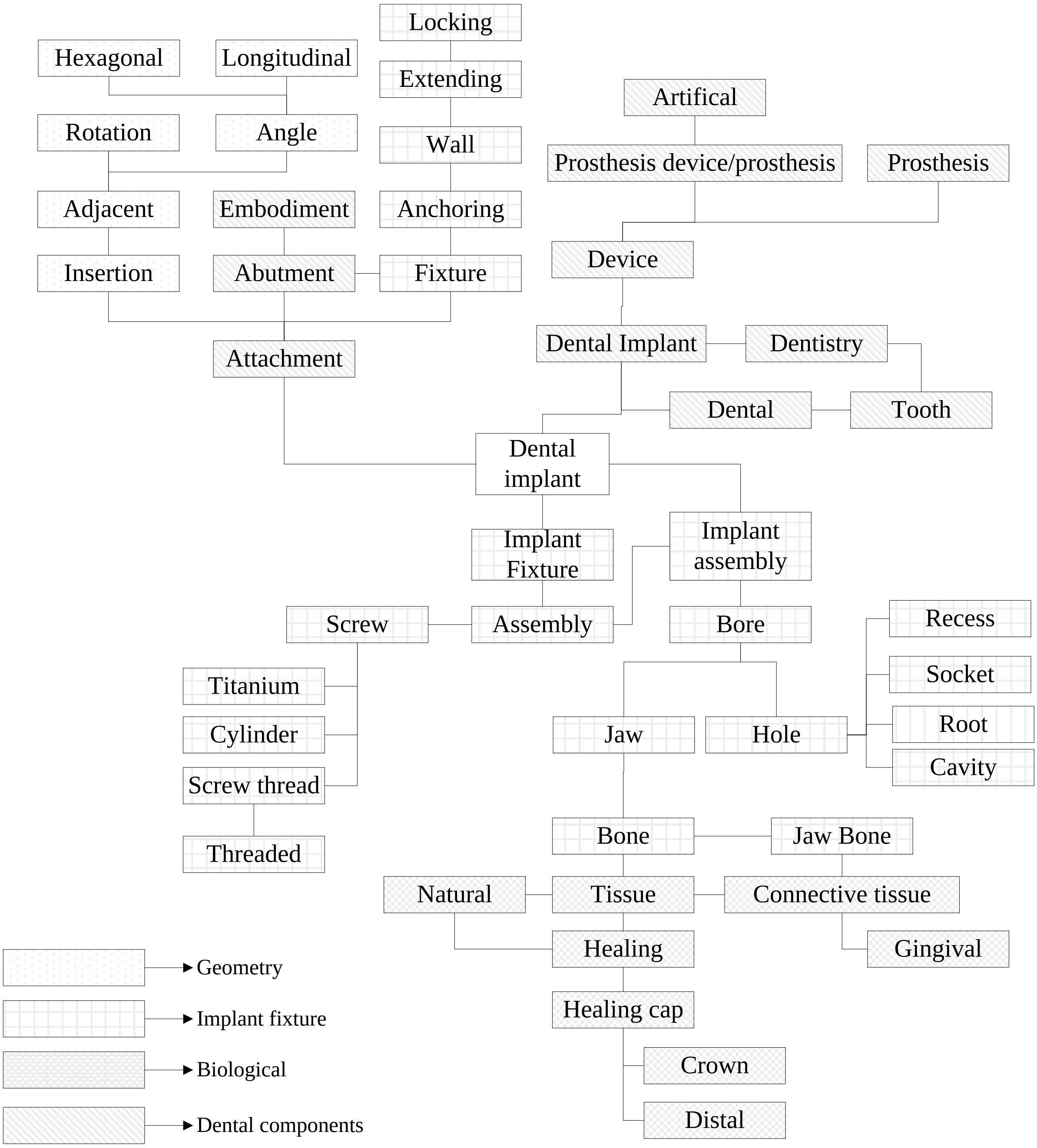


Figure 16. Dental implant ontology tree

Improvement of dental implant ontology

The ontology is often built using phrases defined by experts since ontology is domain specific (Trappey et al. 2010). One method to validate ontology is for example to control figures in each patent document. Thus, it is applicable only for some areas such as dental implants which have a figure of an implant body. However, this research applies a different approach by using patent document clustering methodology and K-means algorithm to cluster the new dental implant patents based on internal relationship of the documents and the output is patents with similar technology is clustered in the same group (Taghaboni-Dutta 2009). The dental implant ontology tree in Figure 16 is divided in four sub-domains and each represents a specific technology of dental implants. Therefore, patent document clustering is used to create clusters of the new dental implant patents which should represent the sub-domain technology of the ontology in Figure 16. Thereafter, using key phrases extracted from each individual cluster to compare with current ontology (Figure 16) if the phrases extracted match.

Currently, there exists no dental implant ontology based on patent data in journals. Therefore this research adapts the dental implant ontology based on Figure 4. The key phrases extracted for dental implant is shown in Table 19. For each dimension, patent document clustering is applied and from each cluster, key phrase extraction is used to extract the top 15 key phrases based on NTFR-values. If extracting more phrases, it will only generate a larger and more complex ontology since 50 phrases is used to build ontology. These key phrases are used to validate and modify the ontology from Figure 16.

Table 19. Key phrases for improvement of implant ontology

Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Key phrases	**Key phrases**	**Key phrases**	**Key phrases**
KP1: Implant	KP1: Implant	KP1: Implant	KP1: Implant
KP2: Dental	KP2: Dental implant	KP2: Dental	KP2: Bone
KP3: Dental implant	KP3: Dental	KP3: Dental implant	KP3: Dental
KP4: Tissue	KP4: Bone	KP4: Screw	KP4: Dental implant
KP5: Bone	KP5: Healing	KP5: Bone	KP5: Jaw
KP6: Bone tissue	KP6: Embodiment	KP6: Prosthesis	KP6: Tissue
KP7: Crown-fixing	KP7: Tissue	KP7: Dental prosthesis	KP7: Fixture
KP8: Ti	KP8: Prosthetic	KP8: Threads	KP8: Implant fixture
KP9: Device	KP9: Screw	KP9: Jaw	KP9: Jaw bone
KP10: Bristles	KP10: Threads	KP10: Embodiment	KP10: Embodiment
KP11: Powder	KP11: Insertion	KP11: Fixture	KP11: Device
KP12: Attachments	KP12: Cavity	KP12: Jawbone	KP12: Crown
KP13: Stabilizer	KP13: Prosthesis	KP13: Teeth	KP13: Prosthesis
KP14: Crown	KP14: Teeth	KP14: Cavity	KP14: Teeth
KP15: Teeth	KP15: Jawbone	KP15: Tissue	KP15: Screw

The comparison of key phrases from Table 19 with ontology in Figure 16 shows that the ontology has to be modified. Since some key phrases from each cluster in Table 19 does not match with the sub-domains in the ontology from Figure 16. The reason may be that each sub-domain may include repeated key phrases to describe the concept of that sub-domain better. For example in Figure 16, the sub-domain “screw” should also contain links to jaw bone, fixture, attachment, crown etc. which describe the concept and sub-domain of “screws” more consistent. Therefore, for example by analyzing the phrases from cluster 1 in Table 19 shows that it is more likely to belong to the sub-domain of “screws” than other sub-domains.

The ontology in Figure 17 includes four sub-domains which are implant, implant assembly, screw device, and implant fixture. The validation of each dimension of ontologies was done once which indicate it requires several repeated validation and modification for a domain-specific ontology. Figure 16 shows a simple ontology and after modification, Figure 17 shows a more detailed ontology which makes more sense in concepts because each sub-domain do share phrases that describe its core concept. However, making the sub-domains too detailed and sharing too many phrases among other sub-domains makes the ontology weak and not functional for clustering since each sub-domain include too many identical phrases. As Taduri et al. (2011) explains that using a single ontology that integrates all knowledge domains has some drawbacks, for example lack of scalability, and a suggestion is to create small niche ontologies and integrate it to a top level ontology. Therefore this research divided dental implant into three dimensions (dental implants/tools/material) to create niche ontology of each dimension to enhance the information retrieval. Also as Taduri et al. (2011) mention is that technical terminologies in patent documents for specific domains is presented in several forms but using ontology allows the application to understand the associations and avoid terminological inconsistencies. The modified ontologies of dental implant tools and materials are shown in Appendix 3.

The sub-domain implant in Figure 17 includes many shared phrases and includes few unique phrases to make it distinct in the ontology. The sub-domain implant assembly includes several unique phrases which increases the possibility of higher quality clusters. Screw device sub-domain also has several unique phrases which show a stronger concept compared with implant sub-domain. The implant fixture sub-domain includes some unique phrases, however this sub-domain includes distinctive phrases which are easily separated from screw device or implant assembly domain. For example, extending, anchoring, rotation, angle, etc. combined with embodiment, insertion and attachment.

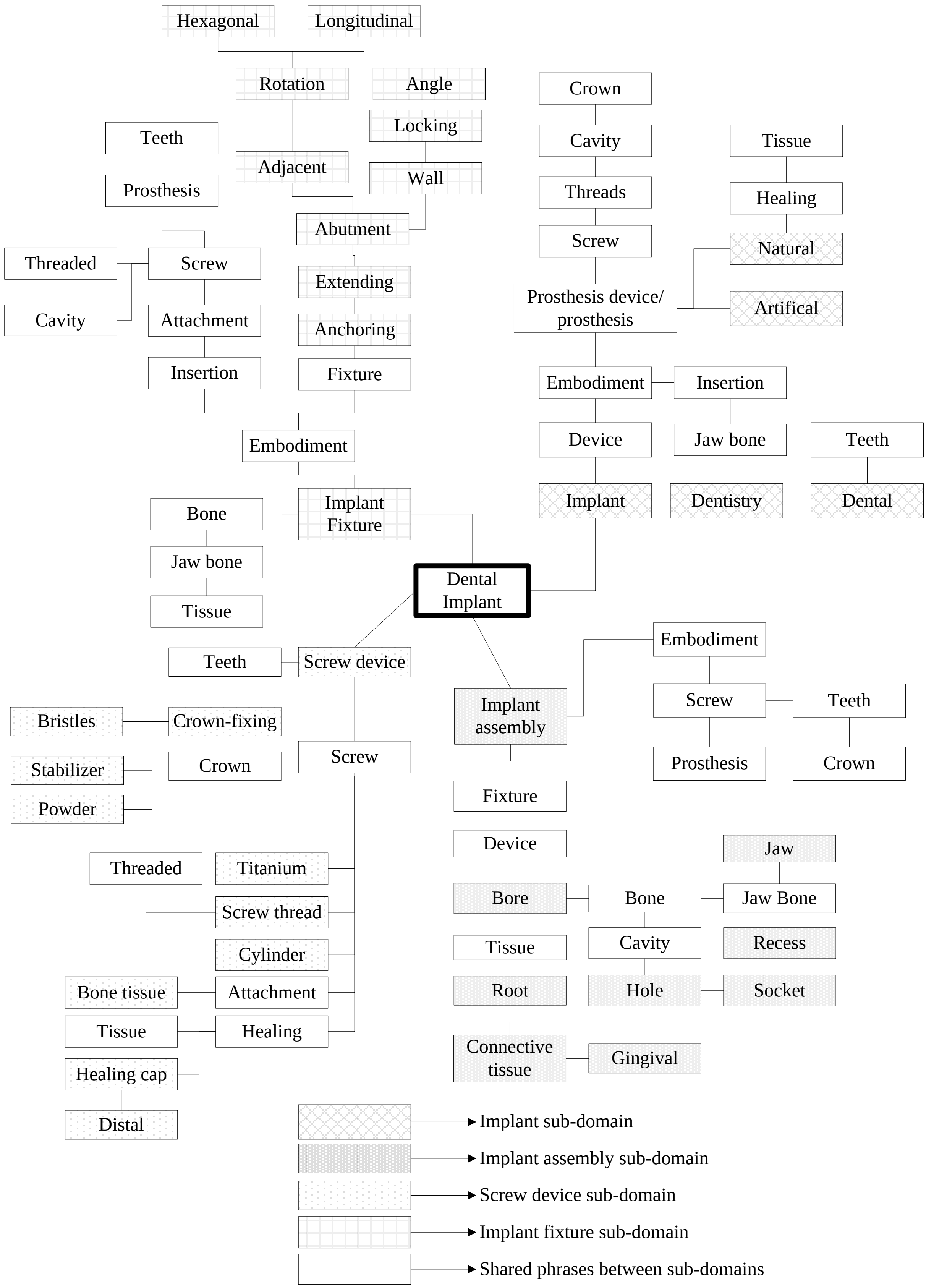


Figure 17. Dental technology ontology tree

4.5. Ontology-based technology clustering of dental implant patents

A case study of the life-span analysis was made only on dental implant ontology. Key phrases are extracted from thirty test patents to create a key phrase and patent correlation matrix. For each individual patent, the frequency and list of key phrases are analyzed and compared with the sub-domains of dental implant ontology (Figure 17) to determine if the patent is suitable for specific sub-domain. Thereafter, each patent is assigned to the ontological sub-domain of implant, implant assembly, screw device, or implant fixture. This is called ontological sub-domain clustering. The ontology links phrases between concepts which makes it possible to perform clustering of patent documents (Wanner et al. 2008). Table 20 shows the key phrases that at least one or more patents in each sub-domain have expressed and the number of patents in each ontological sub-domain cluster. The key phrases are not ranked according to NTFR-value.

Table 20. List of phrases for ontological sub-domains of test patents

7 patents	**4 patents**	**5 patents**	**14 patents**
Other patents	**Implant assembly**	**Screw device**	**Implant fixture**
Implant	Implant	Implant	Implant
Dental	Dental	Dental	Dental
Dental implant	Dental implant	Dental implant	Dental implant
Screw	Screw	Screw	Screw
Fixture	Fixture	Fixture	Fixture
Bone	Bone	Bone	Bone
Cavity	Implant fixture	Implant fixture	Implant fixture
Healing	Cavity	Cavity	Cavity
Embodiment	Healing	Healing	Healing
Prosthesis	Embodiment	Embodiment	Embodiment
Healing screw	Tissue	Tissue	Tissue
Insertion	Prosthesis	Prosthesis	Prosthesis
Dental implant package	Threads	Threads	Threads
Package	Insertion	Extender	Extender
Dental prosthesis	Jawbone	Healing screw	Insertion
Implant package	Dental prosthesis	Insertion	Jawbone
Teeth	Crown	Jawbone	Barrel
Accommodation	Teeth	Dental prosthesis	Dental prosthesis

Table 20 includes a column of other patents because the dental implant ontology was not able to capture its concepts since these patents include key phrases such as “dental implant package” in link with “healing screw”. This result indicates that the dental implant ontology requires improvement or modification to be able to capture these patents too. As Trappey et al. (2010) mention, domain-specific ontology requires key phrases that describe the concepts of the patents. However, the conditions of this research was to focus on dental implants and specific UPCs which results in an ontology that only capture relevant patents and exclude unnecessary patents. It can also depend on the training patent samplings for building the ontology and the 30 test patents in this case study did not restrict any UPC or IPC, only if it included the keyword “dental implant”, it was collected for analysis. The purpose is to examine if the dental implant ontology (Figure 17) is able to capture the concepts of the test patents. Different classification code, for example, patent US5967305 in Table 21 have UPC 206/63.5 which is not included in our training patent collection can result in new phrases.

Table 21. Patent information and dental implants patents over the years

Other patents

Patent No.	Patent title	UPC	Filing date	Age
US5967305	Dental implant component package and holder	206/63.5; 206/486	Sep 17, 1997	14
US7854316	Dental implant package	206/63.5; 206/369; 433/174	Dec 13, 2006	5
US7770722	Dental implant package including a plug	206/63.5; 206/368; 433/174	Nov 17, 2008	3
US5975903	Dental implant and implant system	433/173; 433/174	Apr 2, 1998	13
US6048204	Self tapping screw type dental implant	433/174	Dec 22, 1998	13
US7090493	Dental implant system	433/173	Jan 14, 2003	8
US8021150	Method for dental implant placement	433/72; 433/173	Sep 11, 2008	3
			Average age	8

Furthermore, it can also depend on the age of the patent, for example, patent US4293302 in Table 22 is an expired patent with an age of 31, and thus changes in terminologies for dental implant may have an impact in the key phrase analysis. The patent title, UPC, filing date, and etc. for the sub-domains implant assembly, screw device, and implant fixture, are shown in Table 22, 23, and 24, respectively. Table 21 shows other patents excluded from the ontological sub-domain clustering. A considerable number of test patents have the same UPC as the training patents in Table 15 which indicate several relevant classification codes were included. Although the training patents have different UPC, the dental implant ontology constructed is able to separate these into specific-sub-domains even for patent with the same UPC. This supports the statement of Taghaboni-Dutta et al. (2009) that patents in the same classification codes may be entirely different in technology.

Table 22. Implant assembly sub-domain patent information

Implant assembly sub-domain

Patent No.	Patent title	UPC	Filing date	Age
US6358050	Dental implant systems	433/173	Oct 29, 1998	13
US6273720	Dental implant system	433/173; 433/172; 433/176; 433/201.1	Apr 20, 1999	12
US7125253	Dental implant system and method	433/173; 433/167; 433/215	Oct 1, 2004	7
Expired patent in cluster			Average age	10.7
US4293302	Tooth implants	433/173	Mar 26, 1980	31
		Total average age		14

Table 23. Screw device sub-domain patent information

Screw device sub-domain

Patent No.	Patent title	UPC	Filing date	Age
US5599185	Dental implant healing abutment	433/173; 433/174	Sep 28, 1994	17
US6234797	Dental implant and method for installing the same	433/174	Oct 12, 1999	12
US6287117	Endosseous dental implants including a healing screw and an optional implant	433/173	Apr 22, 1999	12
US7300284	Dental implant system	433/173; 433/172; 433/174; 433/177; 433/182	Aug 3, 2005	6
Expired patent in cluster			Average age	12
US3797113	Dental implant	433/173; 433/201.1	Feb 24, 1972	39
			Total average age	13

Table 24. Implant fixture sub-domain patent information

Implant fixture sub-domain

Patent No.	Patent title	UPC	Filing date	Age
US5571016	Dental implant system	433/173; 433/169	Jan 24, 1995	16
US5752830	Removable dental implant	433/173; 433/169	Jun 20, 1996	15
US5863200	Angled dental implant	433/173	Aug 7, 1997	14
US5931674	Expanding dental implant	433/173	Dec 9, 1997	14
US6171106	Cover screw for dental implant	433/173; 433/174	Sep 9, 1999	12
US6431867	Dental implant system	433/173	Aug 10, 2000	11
US6500003	Dental implant abutment	433/173	Jun 14, 2001	10
US7341453	Dental implant method and apparatus	433/173	Jun 22, 2004	7
US7708559	Dental implant system	433/174	May 14, 2004	7
US6099312	Dental implant piece	433/174	Jul 15, 1999	12
US5951288	Self expanding dental implant and method for using the same	433/173; 433/175; 433/201.1	Jul 3, 1998	13
US7112063	Dental implant system	433/174	Aug 11, 2004	7
US7396231	Flared implant extender for endosseous dental implants	433/173; 433/172; 433/174	Mar 7, 2005	6
Expired patent in cluster			Average age	11
US5022860	Ultra-slim dental implant fixtures	433/174	Dec 13, 1988	23
			Total average age	13

The life-span is calculated from the application date of the patents to today. The other patent sub-cluster (Table 21) do not have expired patents. The other sub-domains implant assembly, screw device, and implant fixture have expired patents. Test patents samplings randomly selected patents from dental implants. The life-span of the sub-domain implant assembly is 14 years (without expired patent 11 years) and screw device 13 years (without expired patent 12 years). Each ontological sub-domain patent cluster is plotted against their clusters average age. The sub-domains implant assembly, screw device, and implant fixture are first plotted without expired patents, shown in Figure 18, and in Figure 19 include the expired patents. Figure 20 shows a comparison of the average age with and without out of date patents in a staple diagram.

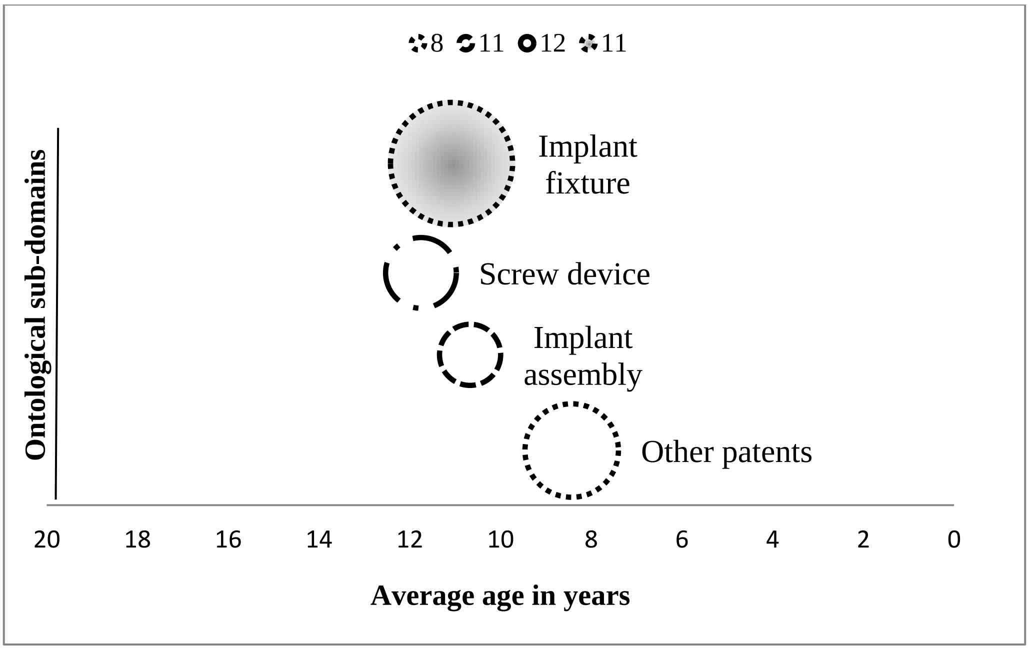


Figure 18. Life-span of dental implant clusters without expired patents

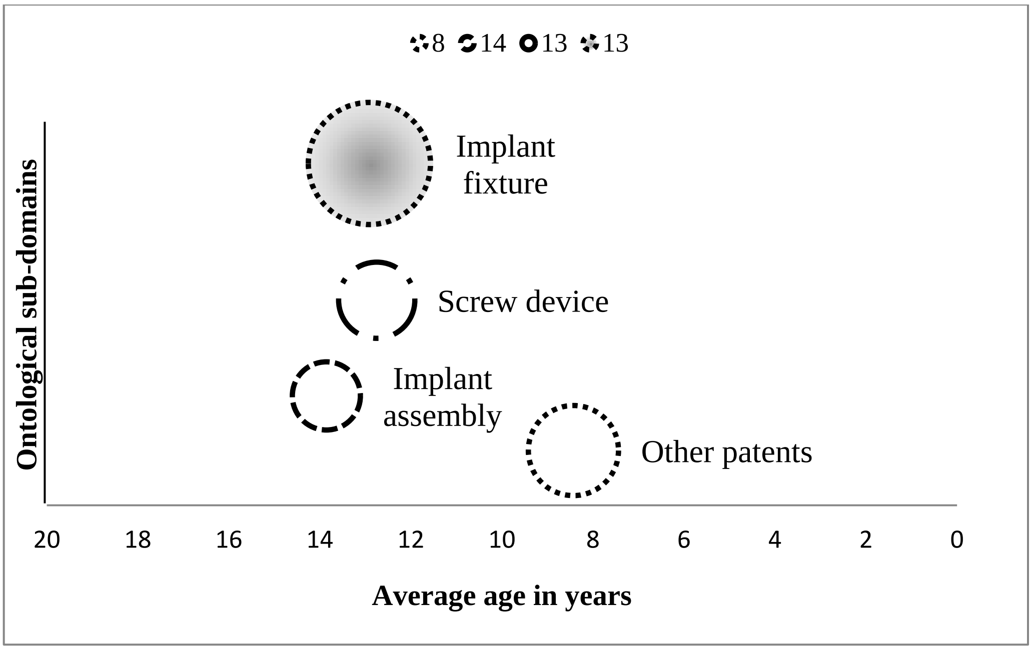


Figure 19. Life-span of dental implant clusters including expired patents

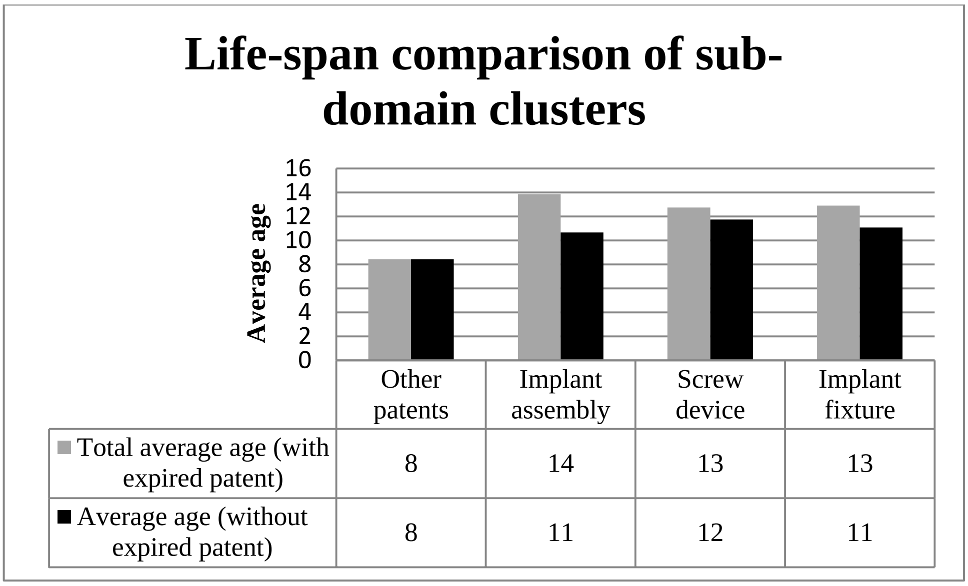


Figure 20. Life-span comparisons of dental implants over the years

<u>Life-span analysis of dental implant patent clusters</u>

The comparison of Figure 18 and Figure 19 shows that there a huge difference when mapping the average age of patents in clusters. For example, the sub-domain implant assembly cluster is a young cluster with an average age of 11 years (without expired patents). However, if including the expired patent of 31 years old affect the average age (3 years difference) and make the implant assembly cluster seem less attractive for R&D investments. As Haupt et al. (2007) states, when companies invest R&D capital on technologies, it often depends on current life cycle stage of the technology. From these test patents it clearly shows that the implant assembly sub-cluster is the youngest and is in the introductory/growth stage with potential growth opportunities for commercial potential and as Haupt et al. (2007) point out, patent based life cycles starts earlier than product/sales based one since the application process is time-consuming (~2-3 years).

The similarity of implant assembly and implant fixture might overlap in ontology, hence, implant assembly focuses more on the surroundings for example drilling a hole or biological including tissues or a device. For implant fixture focuses more on the implant body of attaching the implant crown (artificial teeth) to the jawbone. In this sampling of test patents, it is clear that the implant assembly sub-cluster has great potential for development and its ontological sub-domain include several unique key phrases which support the strength of the dental implant ontology. However, it requires improvements of the ontology to capture several unique key phrases which describe the sub-domain even better and also historical patent activity is an important indicator of current life cycle (Haupt et al. 2007). The screw device and implant fixture sub-domains are also strong with several unique phrases. The implant sub-domain indicates to be weak since it did not capture any patents but it can also depend on the test patents samplings. Both screw device and implant assembly sub-domains results demonstrates signs of growth pattern. Due to the small sampling of test patents makes it rather difficult to draw a conclusion whether which clusters are in the frontier or laggard in technology. It also has to take into consideration that the amount of test patents in each sub-cluster is different and to gain an objective analysis requires a fair volume of test patents and almost equal number of test patents in each sub-cluster. Unquestionably, the ontology has to be taken into consideration since it defines the sub-clusters.

This is a new way of clustering patents and calculating which clusters have the potential of growth or decline. Since patents have a lifetime of 20 years, the life cycle often starts before products and services are being commercialized which must be kept in mind when analyzing potential clusters for R&D investments. Also as Haupt et al (2007) and Trappey et al. (2008) point out that the beginning of a technology life cycle (introduction stage) the new technology is developing scientific fundamental problems and after these technical problems have been solved it will rapidly progress in technological advancement and during this time period awaits radical innovations. Therefore, by analyzing the life-span of domain-specific technology clusters makes it possible to gain quick overview snapshot of technology development and identify potential competitors and their relation between technologies. Furthermore, patents include for example owner of the invention which makes it possible to identify competitors' advantages and disadvantages as well as strategic plans of R&D activities, for example a company have several high impact patents in a mature cluster (soon to be expired) and soon open for commercial usage can be used to create new innovative technologies (Narin and Noma 1987). On average only 1-3 patents out of 100 can generate significant financial returns, although only a few patents have commercial success, most patents are developed by follow-up patenting into significantly important technologies (Ernst 1997). Ontological sub-domain clustering and life-span analysis can help companies to map its own patent portfolio and compare it with the market place to identify potential threat or opportunities for R&D investments, for example strategically invest in young clusters to claim the leadership position and at mature stage evaluate boundaries to avoid infringement issues. As Grilliches (1990) states, patent analysis can be effectively used for companies to gain competitive advantages at market place.

Life-span analysis of clusters is one of the many life cycle analysis techniques. It can be considered as an overview technology cluster analysis for mapping domain specific technologies for further detailed analysis of technology life cycle or by mapping the historical patent activity in each cluster to study technology development in a company, industry, country, etc. (Tseng et al. 2005). Furthermore, increased patent activity (newly issued patents) in a cluster will lower the average age of patents therefore a technology cluster average may be old but can later enter growth stage again by increased patenting activity. According to Haupt et al. (2007) patents can be used to study technology life cycle, the research results demonstrate a snapshot of the current life cycle position of patent clusters and for example in Figure 18 using the S-curve principles discussed by Ernst (1997), the implant assembly sub-cluster has most potential because of young average age and in this case less accumulated patents in the cluster.

For the life-span analysis of dental implant clusters also requires studies of patent activity of each cluster to understand the technology development and to be able to identify potential R&D investment opportunities. Life-span analysis of domain-specific clusters can be used to systematically include patents in limited time period to map the growth of each cluster and the change in average age. For example, include patents only from year 1995 to 2005 and comparing it with patents from 2000 to 2010, which can generate interesting results of historical development of technologies. Furthermore, it can be used to map historical technological barriers which are valuable for R&D to map barriers to overcome or avoid. It should be noted that a cluster in the mature stage such as implant fixture (from Figure 18) can return to the growth stage by increased patent activity in this domain.

5. DISCUSSION AND CONCLUSION

This chapter discusses results obtained and conclusion of this research and the final part ends with future research suggestions.

5.1. Discussion

The patent data are only from USPTO and the boundaries are within specified patent classifications in the dentistry domain. Largest number of patents collected for dental implant is in the patent classifications 433/172 (holding or positioning denture in mouth), 433/ 173 (by fastening to jawbone) and 433/174 (by screw). Consequently some key phrases extracted can have higher statistical significance by the large number of patents collected in these domains. Therefore it is important to define the desired domain and restrict training patents within specific domains as well as collect large volumes of patents in these domains as Huang et al. (2008) describes the ontology should reflect a specific domain. For example in this research, the focus was on dental implants. The focus on specific patents will generate relevant key phrases for building the ontology therefore based on Trappey et al. (2008) NTF-methodology which generates high frequent key terms in patent documents can therefore represent the concept of the document (Robertson 2004; Trappey et al. 2008). As suggested by Trappey et al. (2010), to construct a domain-specific ontology requires domain-specific data collection since the key phrase analysis extracts key phrases for building the ontology and it was done in this research.

As Taduri et al. (2011) proposes to create several small niche knowledge domains of ontologies and then integrate it to a top level ontology. This research created four sub-domains of dental implant technology ontology and the modification step which made the ontology more detailed to enhance information retrieval. The dental technology ontology structure was based on RFID-ontology tree found in Trappey et al. (2009) whom used it for automatic patent document summarization and this research adapted the RFID-ontology structure. By using WordNet (Princeton University 2011) to understand the relationships between words and the analysis of which patent classification the key phrase is expressed in. The dental implant ontology structure also depends on how the researcher defines the sub-domains of the key phrases.

The dental implant ontology validation methodology has effectively utilized patent analysis techniques developed by C.V Trappey et al. (2010) which used key phrase analysis and patent document clustering with K-means algorithm to cluster technology into groups according to the internal relationship of the concepts of the patent documents. It resulted in four relevant sub-domains of the dental implant ontology. The results indicate that a sub-domain share several phrases and it demonstrate patents in different sub-domains share phrases but can be separated by unique phrases expressed in that sub-domain. As Wanner et al. (2008) states that ontology should link concepts to make it possible to perform pattern recognition and clustering of patent documents with respect to its content which was done in this research. In other words, to describe a concept (sub-domain) also requires the general phrases (as for other sub-domains) for it to make sense, for example, abutment, attachment, implant, prosthesis, etc. (generally the top 10 phrases from the key phrase matrix). This indicates that a domain-specific ontology requires iterative steps of improvement as shown in step 3 in Figure 21.

The case study of dental implant patent (step 4 in Figure 21) can be seen as a validation step but the methodology is not the same. This step will confirm if the ontology constructed is strong enough to capture concepts of random dental implants. The dental implant ontology in Figure 17 has demonstrated sub-domain clustering ability of dental implant patents. The dental implant ontology also excluded several patents since these patents did not match the concepts of the sub-domains. It is an indication of a well constructed ontology that is able to capture relevant patents and exclude the other patents which consistent with literature (Wanner et al. 2008; Trappey et al. 2010). However if, patents excluded are domain-specific it can be used to modify the ontology. This research developed a general methodology shown in Figure 21 which summarizes the general steps in this research to create a domain-specific ontology for ontological sub-domain clustering and life-span analysis for dental implant patents. As Tseng et al. (2007) mentioned that patent documents are often lengthy and require time, effort and expertise to interpret the result into a technology development analysis, this research has used automated patent analysis techniques to create a dental implant ontology which is able to perform sub-domain clustering and life-span analysis of these clusters to identify potential R&D opportunities for dentistry companies to pursue R&D investments.

The results of the life-span analysis show that the sub-cluster implant assembly is the youngest and most promising (in this case). However, this can be considered as a snap shot of current dental implant market. The life-span analysis gives a view of which clusters are in the growing or maturing but it can also give false signs since it clearly change depending on the patents application age (young or expired). The size of the sampling in each sub-cluster will also affect the outcome since the average age is calculated. To gain a competitive analysis requires key patents and similar sample size of each sub-cluster. Also to gain competitive knowledge from the life-span analysis requires study of the historical patent activity of these sub-clusters because increased patent activity in old sub-cluster will decrease the average age rapidly making the sub-cluster to enter the growth phase again. However, based on the conditions of this research, the results indicate that implant assembly sub-domain is the youngest and has growth potential. Implant fixture cluster is not much older but the size of the cluster compared with the other clusters is less attractive (more competitors).

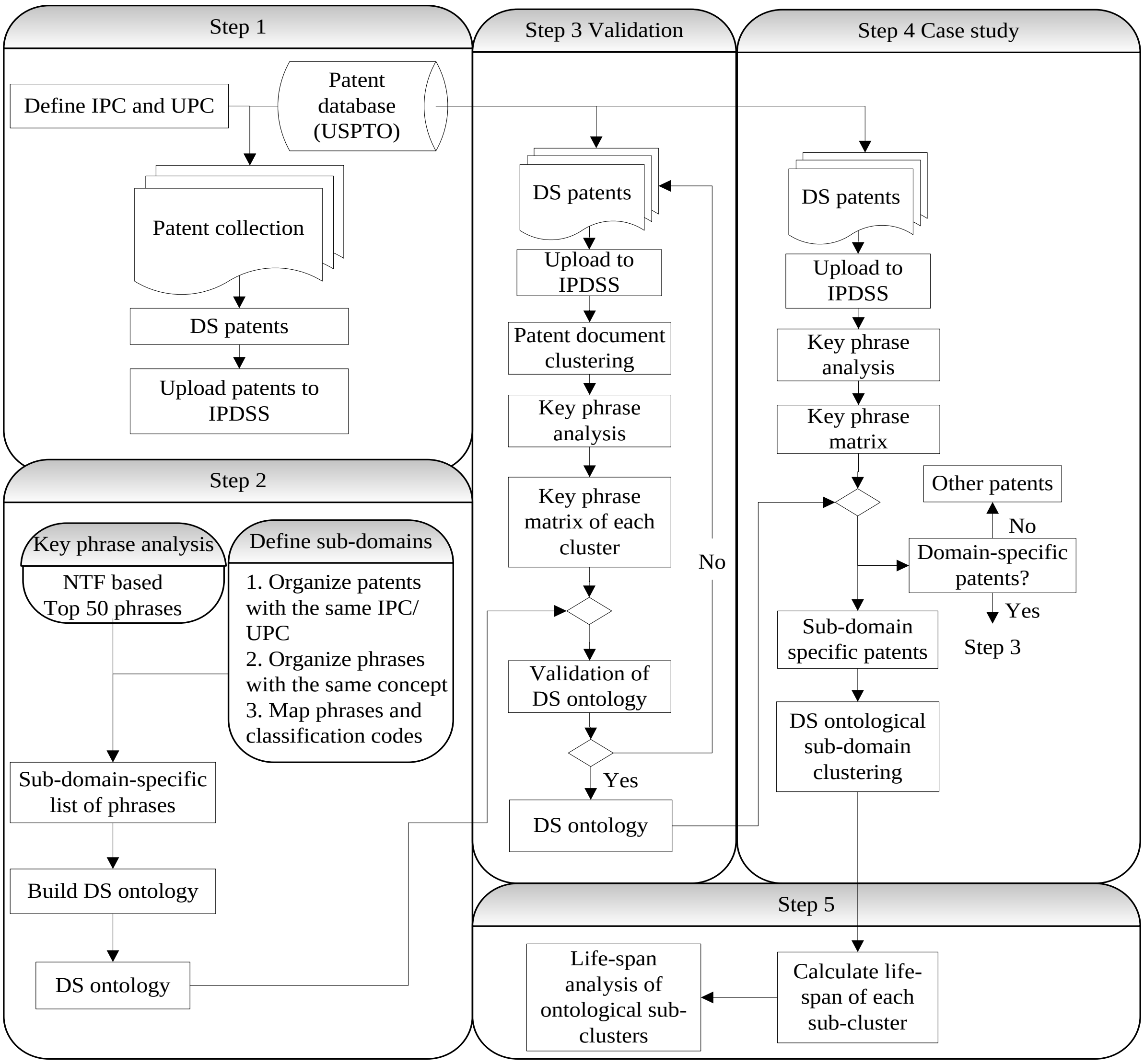


Figure 21. Process for building a domain-specific ontology

5.2. Conclusion

The objectives of this research are the development of a domain-specific patent ontology for technology clustering and life-span analysis. Syntheses of methodologies of patent analysis, procedures were developed to create a domain-specific ontology and the procedures for ontological sub-domain technology clustering and technology life-span analysis were developed. A case study of dental implant patent demonstrates that R&D managers can gain a quick overview snapshot of technology development and identify potential competitors. Furthermore, it can help R&D managers to identify the advantages and disadvantages of technologies of their competitors. In addition, dental implants R&D engineers can quickly create and modify a domain-specific ontology to perform patent analysis to study technology development as well as strategically plan R&D activities. In this research, dental implant technology ontology was created and according to the background research dental implant patent ontology has not been created before. R&D engineers or managers can use the dental technology ontology for capturing new dental implant patents in other patent classifications or map its own patent portfolio and compare it with market place to identify potential threat or opportunities for investments. By studying the life-span of a technology a company can minimize the investment risk in R&D because the status of technology life cycle is often an important factor in decision making for technology investments.

The results of this research demonstrates the development of a domain-specific patent ontology procedure by using data mining techniques to perform patent analysis which is faster than traditional patent analysis techniques. The developed approach in this research can help R&D engineers as well as managers to gain a quick overview of potential technology clusters by studying the life-span of these clusters. Ontology is useful because it covers the main concepts in a domain and relations between concepts. It enables R&D engineers in Asia or non-English speaking countries to better understand a technology or at least the key concepts in comparison with traditional patent analysis which is highly expensive. Data mining techniques in this research applied the same NTF-key phrase analysis technique as Trappey et al. (2008) and key phrases extracted describes the concept of dental implants well since most of the phrase are listed on Astra Tech Dental (2011). The case study of dental implant ontology also shows that the patent samplings which consists of several patent classifications has similar technology even though classified in different classes, which also proves Taghaboni-Dutta et al. (2009) statement that patents in the same classification codes may be entirely different in technology. The steps developed in Figure 21 also show the flexibility and possibility to quickly adapt the domain-specific ontology for clustering and life-span analysis. The dental implant ontology also demonstrate the ability to cluster dental technology and exclude unnecessary patents, in other words, the dental implant ontology only captured domain-specific patents.

5.3. Future research suggestions

Future research should be conducted in improving the methodology of building a domain-specific ontology and methods of validating it. The approach taken in this research has potential for development. A case study can be done to recreate a domain-specific ontology in an area where the ontology has already been validated by domain experts. There is also a need for experts in specific domain to validate the ontology or help building it for gaining a competitive clustering analysis. Two other dental ontologies were created for this research, tools and material. These can be used for case study and also as Taduri et al. (2011) mentioned, integrated to a top level ontology. The dental implant ontology created in this research can be applied in other research for example in Trappey et al. (2010) that uses ontology for key phrase extraction of patent documents.

The life-span analysis made in this research is an interesting approach to map clusters to gain insights whether it has potential or not. However, it requires a strong ontology or other high quality clustering techniques to conduct a good life-span analysis. As mentioned in the conclusion, mapping clusters as being static does not help researchers to understand its development, it is therefore important to map its historical movement throughout several years. The main reason is that a cluster can “grow” young again by increased patent activity. This can be done by creating time-series plots for studying the evolving effect of domain-specific patent technology clusters. By adding a 3rd dimension into the plot, how the clusters are growing or maturing through time and the effects (curves) of technology clusters can be studied. By using the principles of S-curve discussed by Ernst (1997), life-span analysis of domain-specific clusters can be applied to map the technology cluster life cycle. The steps created in Figure 21 can be applied to a company’s patent portfolio and using the suggestion of S-curve, map its own technology development compared with competitors.

Other research suggestions are to compare this developed methodology of constructing a domain-specific ontology with other methods to determine the significance of these methodologies.

REFERENCES

Almeida, JAS, Barbosa, AA, Pais, CC & Formosinho, SJ 2007, '*Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering*', Chemometrics and Intelligent Laboratory Systems, Vol. 87, pp.208-217

Aizawa, A 2002, '*An information-theoretic perspective of tf-idf measures*', Information Processing and Management, Vol.39, pp.45-65

Andersen, B 1999, '*The hunt for S-shaped growth paths in technological innovation: a patent study*', Journal of Evolutionary Economics, Vol.9, pp.487-526

Astra Tech Dental, viewed 8 August 2011, <http://www.likenaturalteeth.us/Main.aspx/Item/781594/navt/83633/navl/83642/nava/83644>

Austin, D 1993, *'An event-study approach to measuring innovative output: the case of biotechnology*', American Economic Review 83, pp.253-258

Ayres, RU 1995, '*Life cycle analysis: A critique'*, Elsevier Science, Vol.95, pp. 199-223

Brocair Partners Industry Survey Series, '*Overview of the Dental Products Marke*t', viewed 10 August 2011, <http://www.brocair.com/downloads/Dental_Products_Market.pdf>

Choi, C, Kim, S & Park Y 2007, '*A patent-based cross impact analysis for quantitative estimation of technological impact: The case of information and communication technology*', Technological Forecasting and Social Change, Vol. 74, pp.1296-1314

Daim, T, Rueda, G, Martin, H & Gerdsri P 2006, '*Forecasting emerging technologies: use of bibliometrics and patent analysis*', Technological Forecasting & Social Change, Vol. 73, pp. 981-1012

Dunham, M 2003, '*Data mining: introductory and advanced topics'*, Upper Saddle River, Prentice Hall, New Jersey

Ernst, H 1995, *'Patent applications and subsequent changes of performance: evidence from time-series cross-section analyses on the firm level'*, Research Policy, pp.143-157

Ernst H 1997, '*Use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry'*, Small Business Economics, Vol. 9, pp.361-381

Fayyad, U, Piatesky-Shapiro, G & Smyth, P 1996, '*From Data Mining to Knowledge Discovery in Databases*', American Association for Artificial Intelligence, Vol.17, no.3, pp.37-54

Free Dental Implant Information, viewed 8 August 2011, < http://www.free-dental-implants.com/dental-implant-components/>

Granstrand, O 1999, '*The economics and management of intellectual property: Towards Intellectual Capitalism'*, Edward Elgar Publishing Limited, Cheltenham, UK.

Grilliches, Z 1998,' *Patent Statistics as Economic Indicators: A Survey'*, National Bureau of Economic Research, University of Chicago Press, pp.287-343.

Grilliches, Z 1990, '*Patent statistics as economic indicators: a survey*', Journal of Economic Literature, pp.1661-1707

Hammouda, KM, Matute, DN & Kamel, MS 2005, '*CorePhrase: Keyphrase Extraction for Document Clustering*', MLDM, LNAI 3587, pp.265-274

Han, J, Kamber, M & Pei, J 2011, 'Data mining – Concepts and Techniques', Morgan Kaufmann Publishers, Elsevier, Waltham, U.S.A.

Haupt, R, Kloyer, M & Lange, M 2007, '*Patent indicators for the technology life cycle development*', Research policy, Vol.36, pp.387-398

Her Majesty's (HM) Government 2011 , Department for Business Innovation and Skills, '*Strength and Opportunity – The landscape of the medical technology, medical biotechnology and industrial biotechnology sectors in the UK'*, viewed August 2011, p.16, <http://www.bis.gov.uk/files/file53947.pdf>

Hirschey, M & Richardson, V.J. 2001 , '*Valuation of patent quality: A comparison for Japanese and U.S. firm* ', Pacific-Basin Finance Journal, Vol.9, issue 1, pp.65-82

Huang, CJ, Trappey, AJC & Wu, CY 2008, '*Develop a Formal Ontology Engineering Methodology for Technical Knowledge Definition in R&D Knowledge Management*', Advanced Concurrent Engineering, Vol.12, pp.495-502

Jun, S & Uhm, D 2010, '*Technology Forecasting Using Frequency Time Series Model: Bio-Technology Patent Analysis*', Journal of Modern Mathematics and Statistics, Vol.4, no.3, pp.101-104

Kettenring, JR 2009, '*A patent analysis of cluster analysis*', Applied Stochastic Models in Business and Industry, Vol.25, pp.460-467

Khelif, K, Dieng-Kuntz, R, & Barbry, P 2007, *'An Ontology-based Approach to Support Text Mining and Information Retrieval in the Biological Domain'*, Journal of Universal Computer Science, Vol.12, no.12, pp.1881-1907

Kim, YG, Suh, JH & Park, SC 2008, '*Visualization of patent analysis for emerging technology'*, Expert Systems with Applications, Vol.34, no.3, pp.1804-1812

Kostoff, R, Toothman, D, Eberhart, H & Humenik J 2001, '*Text mining using database tomography and bibliometrics: a review'*, Technological Forecasting and Social Change, Vol. 68, pp. 223-252

Lai, KK &Wu, SJ 2005, '*Using the patent co-citation approach to establish a new patent classification system*', Information Processing & Management, Vol.41, no.2, pp.313-330

Lau, GT, Taduri, S, Law, KH, Yu, H & Kesan, JP 2011, '*Developing an Ontology for the U.S. Patent System'*, Annual International Conference on Digital Government Research, University of Maryland, College Park, June 12-15, 2011

Lee, S, Yoon, B &Park, Y 2009, '*An approach to discovering new technology opportunities: Keyword-based patent map approach*', Technovation, Vol.29, pp.481-497

Lerner, J. 1995, '*Patenting in the Shadow of Competitors*', Journal of Law and Economics, Vol. 38, no. 2, pp.463-465

Matsuo, Y & Ishizuka M 2003, '*Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information*', FLAIRS Associations for the Advance Artificial Intelligence, pp.392-396

Mogee, ME 1991, '*International Patent Data for Technology Analysis and Planning'*, Center for Innovation Management Studies, Lehigh University.

Nasukawa, T & Nagano, T 2001, '*Text analysis and knowledge mining system'*, IBM Systems Journal, Vol.40, no.4, pp.967-984

Navigli, R, & Velardi, P 2004, 'Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites', Computational Lingustics, Vol.30, no.2, pp.151-179

Nenkova, A, Vanderwende, L & McKeown, K 2006, '*A compositional Context Sensitive Multi-document Summarizer: Exploring the Factors That Influence Summarization'*, In Proceedings of SIGIR, pp.573-580

Norton, MJ 2001, *'Introductory Concepts in Information Science*', ASIA, New Jersey

Noy, NF & McGuiness, DL 2001, '*Ontology Development 101: A guide to Creating Your First Ontology'*, Development Stanford K, pp.1-25

Paterson, H, Zamanian, K 2009, '*The global dental implant market to experience strong growth despite the economic downturn*', iData Research Inc., Vol.2, no.4,

Princeton University, WordNet, viewed 11 November 2011, <http://wordnet.princeton.edu/>

Puja Dental Group, 67 N Dobson Rd. Suite A-104, Mesa, AZ, USA, viewed 5 November 2011, <http://www.pujadentalgroup.com/dental-implants-phoenix.htm>

Robertson, S 2004, '*Understanding inverse document frequency: on theoretical arguments for IDF*', Journal of Documentation, Vol.60, no.5, pp.503-520

Robertson, SE & Sparck Jones, K 1976, '*Relevance Weighting of Search Terms*', Journal of the American Society for Information Science, Vol.27,no.3, pp.129-146

Rubin, DL, Shah, NH & Natalya, FN 2007, '*Bomedical ontologies: a functional perspective*', Breiefings in Bioinformatics, Vol.9, no.9, pp.75-90

Sapsalis, E, Pottelsberghe, B & Navon, R 2006, '*Academic versus industry patenting: An in-depth analysis of what determines patent value*', Research Policy, Vol.35, pp.1631-1645

Shahnaz, F, Berry, MW, Pauca, VP & Plemmons, RJ 2006, '*Document clustering using nonnegative matrix factorization*', Information Processing and Management, Vol.42, pp.373-386

Smith, H 2002, '*Automation of patent classification*', World Patent Information, Vol.24, pp.269-263

Soo, VW, Lin, SY, Yang, SY, Lin, SN & Cheng, SL 2006, '*A cooperative multi-agent platform for invention based on patent document analysis and ontology'*, Expert Systems with Applications, Vol.31, pp.766-775

Taduri. S, Lau, GT, Law, KH, Yu, H & Kesan, JP 2011, '*Retrieval of Patent Documents from Heterogeneous Sources using Ontologies and Similarity Analysis*', Fifth IEEE International Conference on Semantic Computing (ICSC 2011), Stanford University, Palo Alto, CA, USA, Sept. 18-21, pp.1-8

Taghaboni-Dutta, F, Trappey, AJC, Trappey, CV & Wu, HY, '*An exploratory RFID patent analysis*', Management Research News, Vol.32, no.12, pp.1163-1176

Tan, AH 2011, '*Text mining: The state of the art and the challenges*', Proceedings of the PAKDD, Vol.696, pp.65-70

Trappey, CV, Trappey, AJC & Wu, CY 2010, '*Clustering Patents Using Non-exhaustive Overlaps*', Journal of Systems Science and Systems Engineering, Vol.19, no.2, pp.162-181

Trappey, AJC, Trappey, CV, Wu, CY 2009, '*Develop Non-Exhaustive Overlapping Partioning Clustering for Patent Analysis Based on the Key Phrases Extracted Using Ontology Schema and Fuzzy Adaptive Resonance Theory*', Advance Concurrent Engineering, Vol.7, pp.349-360

Trappey, CV, Wu, CY,Taghaboni-Dutta, F & Trappey, AJC 2011, '*Using patent data for technology forecasting: China RFID patent analysis*', Advanced Engineering Informatics, Vol.25, issue1, pp.53-64

Trappey, AJC, Trappey, CV, Hsu, FC & Hsiao, DW 2009, '*A Fuzzy Ontological Knowledge Document Clustering Methodology*', IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol. 39, no. 3, pp. 1083-4419

Trappey AJC, Trappey CV, Wu CY, Lin CW 2011, '*A Patent Quality Analysis for Innovative Technology and Product Development*', Advanced Engineering Invormatics, In Press

Trappey, AJC, Trappey, CV & Wu, CY 2009, '*Automatic patent document summarization for collaborative knowledge systems and services*', Journal of Systems Science and Systems Engineering, Vol.18, no.1, pp.71-94

Trappey, AJC, Trappey, CV & Wu, CY 2008, '*A Semantic Based Approach for Automatic Patent Document Summarization'*, Advanced Concurrent Engineering, Part 12, pp.485-494

Trappey, AJC & Trappey, CV 2007, '*An R&D knowledge management method for patent document summarization*', Industrial Management & Data Systems, Vol.108, no.2, pp.245-257

Tseng, YH, Wang, YM, Juang, DW & Lin, CJ 2005, '*Text Mining for Patent Map Analysis'*, IACIS Pacific 2005 Conference Proceedings

Tseng, YH, Lin, CJ & Lin, YI 2007, '*Text mining techniques for patent analysis*', Information Processing and Management, Vol.43, pp.1216-1247

Turney, PD 2000, '*Learning algorithms for key phrase extraction*', Information Retrieval, Vol.2, pp.303-336

United States Patent Trademark Office (USPTO), '*A Patent Technology Monitoring Team Report*', Parts A1, A2, B, viewed 30 August 2011, <http://www.uspto.gov/web/offices/ac/ido/oeip/taf/apat.htm#PartA1>

University of Ottawa 2011, *Medical Devices – Challenges and Opportunities for Enhancing the Health and Wealth of Canadians*, Medical Devices Innovation Institute (MDI), Ottawa, Canada

Wanner, L, Baeza-Yates, R, Brugmann, S, Codina, J, Diallo, B & Escorsa, E, '*Towards content-oriented patent document processing*', World Patent Information, Vol.30, pp.21-33

Weiss, S, Indurkhya, Zhang N, Damerau F 2005, '*Text mining predictive methods for analyzing unstructured information*', Springer, Berlin

WheelJet, viewed 19 December 2011, <http://140.114.53.242/IPDSS/login.aspx>

World Intellectual Property Orgranisation (WIPO), viewed 01 September 2011, <http://www.wipo.int/portal/index.html.en>

Yoon, B & Park, Y 2004, '*A text-mining-based patent network: Analytical tool for high-technology trend*', The Journal of High Technology Management Research, Vol.15, pp. 37-50

Yoon, B & Park, Y 2007, '*Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information*'. IEEE Transaction on Engineering Management, Vol.54, no.3, pp.588-599

Yoon, B & Lee, S 2008, '*Patent analysis for technology forecasting: Sector-specific applications*', Proceedings of the IEEE International Conference on Engineering Management, June 28-30, IEEE International, pp. 1-5

APPENDIX 1 Key phrase and patent correlation matrix

Tables show part of the key phrase and patent correlation matrix for dental implant. Because of the large volume of patent documents collected it will only cause confusion if the complete table is presented in a word document. Therefore MS Excel files are available on request. Please contact me by mail.

Appendix 1-1 Dental implant key phrase and patent correlation matrix

Key phrase /PAT.NO.	**US3672058**	**US4626214**	**US4195409**	**…**	**NTFR**
UPC	**433/174**	**433/174**	**433/175**	**…**	
KP1: Implant	75	55	33	…	14023
KP2: Dental	29	82	24	…	6422
KP3: Dental Implant	12	45	21	…	3808
KP4: Bone	10	0	67	…	3628
KP5: Screw	35	0	0	…	1872
KP6: Abutment	0	0	0	…	1056
KP7: Threaded	19	31	0	…	1575
KP8: Bore	0	37	0	…	932
KP9: Prosthesis	29	43	29	…	830
KP10: Cap	0	0	0	…	436
KP11: Healing	0	0	0	…	413
KP12: Root	6	0	48	…	399
KP13: Tissue	0	0	21	…	710
KP14: Healing cap	0	0	0	…	130
KP15: Fixture	0	0	0	…	218
KP16: Implants	17	0	5	…	822
KP17: Embodiment	0	4	6	…	775
KP18: Threads	15	0	0	…	446
KP19: Cavity	21	0	19	…	381
KP20: Hole	6	16	0	…	402
KP21: Crown	0	0	54	…	306
KP22: Jaw	0	0	5	…	456
KP23: Jawbone	33	0	0	…	268
KP24: Device	0	0	0	…	257
KP25: Implant fixture	0	0	0	…	97
KP26: Prosthetic	0	0	0	…	199
KP27: Artificial	12	0	0	…	233
KP28: Thread	0	12	0	…	198
KP29: Extending	8	14	0	…	371
KP30: Jaw bone	0	0	5	…	171
KP31: Socket	0	0	25	…	192
KP32: Distal	0	0	0	…	107
KP33: Anchoring	4	0	0	…	116

KP34: Attachment	0	0	5	…	161
KP35: Wall	12	0	0	…	119
KP36: Recess	10	23	3	…	142
KP37: Assembly	0	0	0	…	107
KP38: Dental prosthesis	4	35	0	…	106
KP39: Attached	0	0	16	…	195
KP40: Hexagonal	0	0	0	…	131
KP41: Longitudinal	10	4	0	…	129
KP42: Angle	0	0	0	…	121
KP43: Teeth	10	0	8	…	152
KP44: Titanium	0	0	0	…	132
KP45: Rotation	0	8	0	…	123
KP46: Natural	0	0	27	…	107
KP47: Adjacent	0	0	0	…	104
KP48: Insertion	0	10	5	…	91
KP49: Dental implants	4	0	0	…	101
KP50: Art	0	0	5	…	121

APPENDIX 2 Sub-domain key phrase matrix

The table shows the top 50 key phrases defined in different sub-domains. For dental implant there are 4 sub-domains where key phrases are group together according to its sub-domain. For example, from KP1: implant to KP6: abutment belongs to the same sub-domain. For dental implant tool and material there are 5 sub-domains each and key phrases are also grouped together.

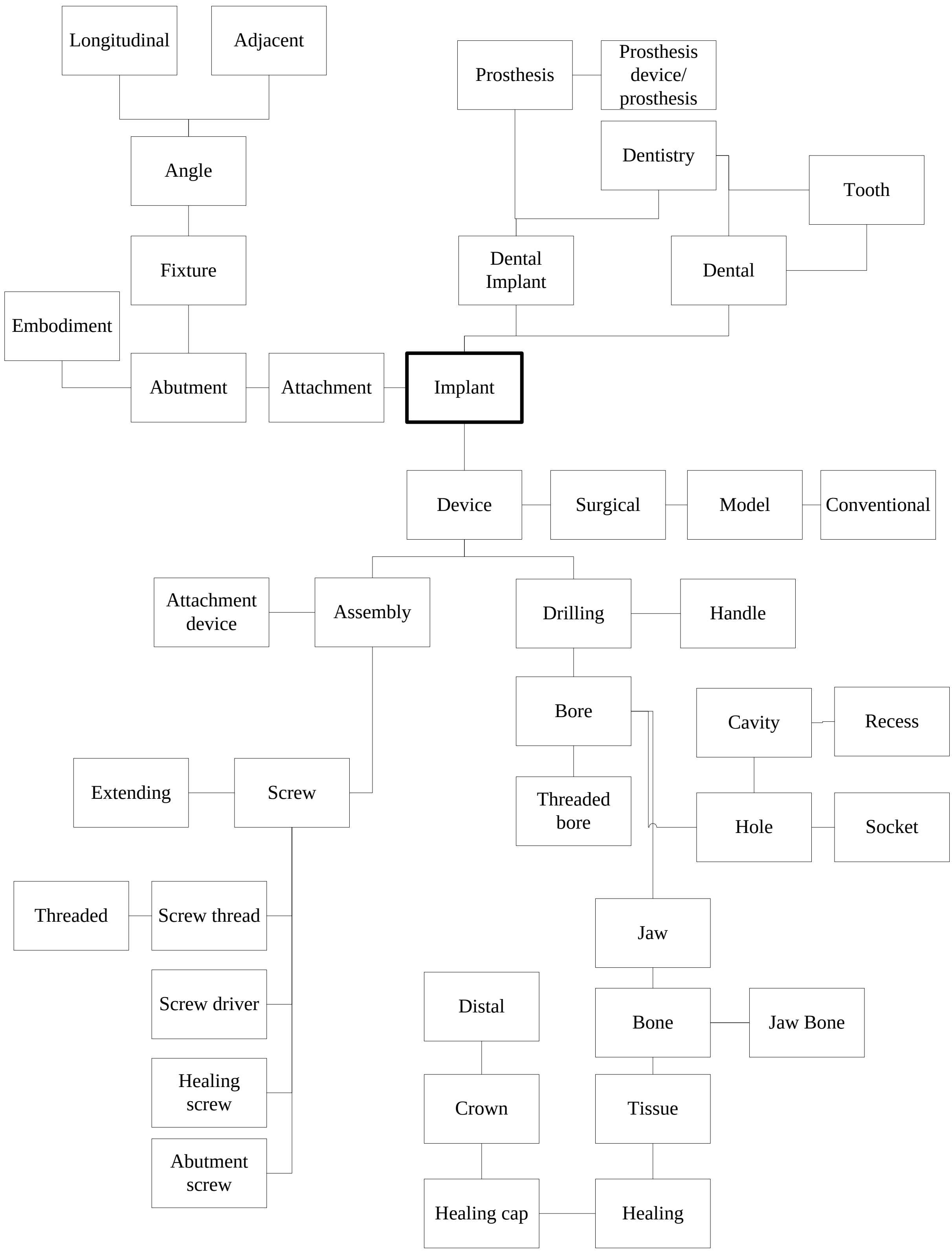
Appendix 2-1 Dental implant sub-domain key phrase matrix

Key phrase/UPC						
KP1: Implant	433/173	433/172	433/175	433/201.1	433/17	433/176
KP2: Dental	433/173	433/172	433/175	433/201.1	433/17	433/176
KP3: Dental implant	433/173	433/172	433/175	433/201.1	433/17	433/176
KP49: Dental implants	433/173	433/172	433/175	433/201.1		433/176
KP16: Implants	433/173	433/172	433/175	433/201.1	433/17	433/176
KP17: Embodiment	433/173	433/172	433/175	433/201.1	433/17	433/176
KP34: Attachment	433/173	433/172	433/175	433/201.1		
KP39: Attached	433/173	433/172	433/175	433/201.1		
KP27: Artificial	433/173	433/172		433/201.1		433/176
KP9: Prosthesis	433/173	433/172	433/175	433/201.1		433/176
KP26: Prosthetic	433/173	433/172		433/201.1		
KP38: Dental prosthesis	433/173	433/172		433/201.1		
KP24: Device	433/173		433/175	433/201.1		433/176
KP44: Teeth	433/173	433/172	433/175	433/201.1	433/17	433/176
KP6: Abutment	433/173	433/172	433/175	433/201.1	433/17	433/176

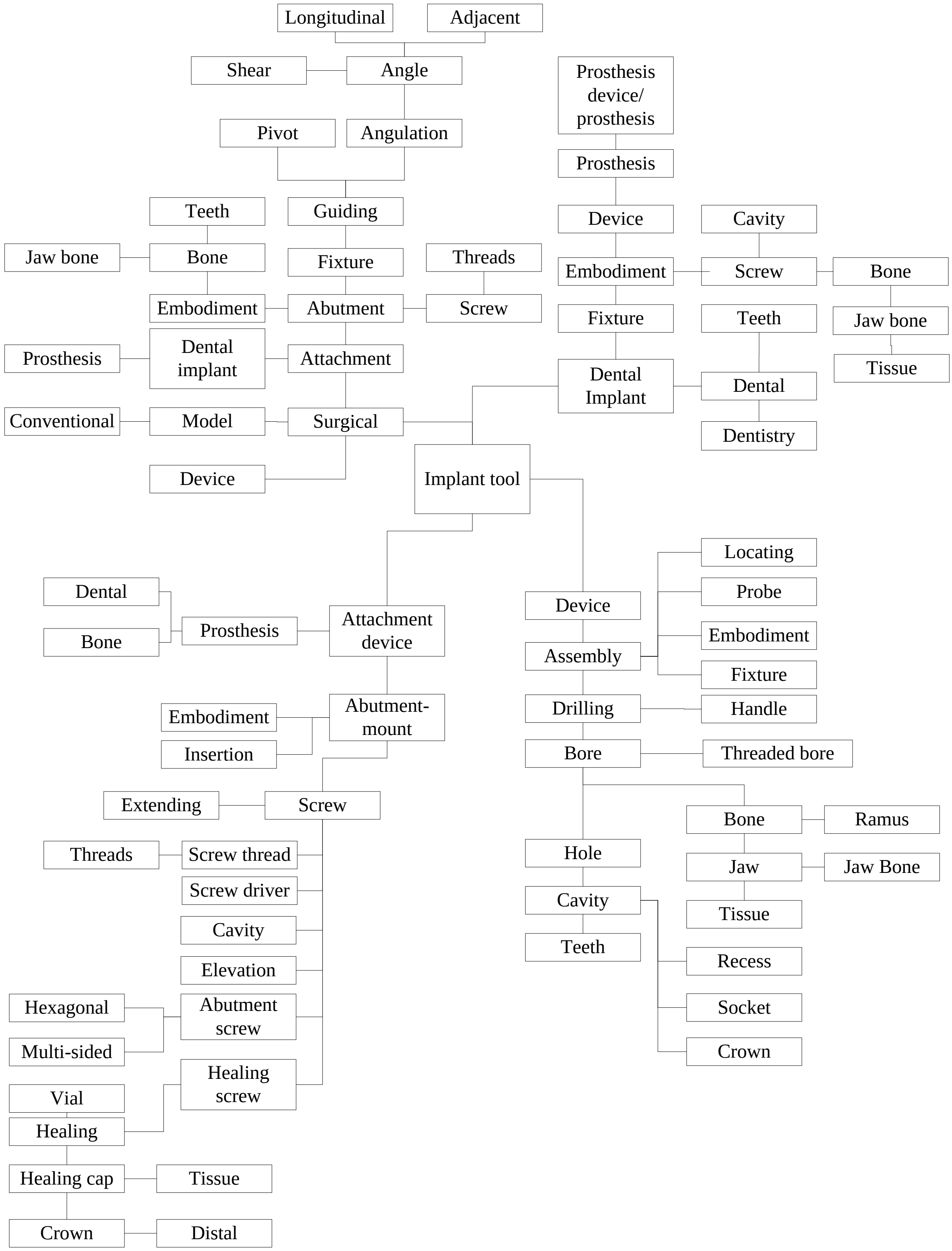
KP5: Screw	433/173	433/172	433/175			
KP7: Threaded	433/173	433/172	433/175	433/201.1		433/176
KP18: Threads	433/173	433/172	433/175			
KP28: Thread	433/173	433/172				
KP45: Titanium	433/173	433/172	433/175	433/201.1	433/17	433/176
KP29: Extending	433/173	433/172		433/201.1	433/17	433/176
KP33: Anchoring	433/173			433/201.1	433/17	
KP25: Implant fixture	433/173	433/172			433/17	
KP15: Fixture	433/173	433/172			433/17	
KP37: Assembly	433/173	433/172				433/176
KP35: Wall	433/173	433/172				433/176
KP36: Recess	433/173	433/172	433/175			
KP8: Bore	433/173	433/172	433/175		433/17	433/176
KP20: Hole	433/173	433/172	433/175	433/201.1	433/17	433/176
KP31: Socket	433/173	433/172	433/175	433/201.1	433/17	433/176
KP19: Cavity	433/173	433/172	433/175	433/201.1		433/176
KP4: Bone	433/173	433/172	433/175	433/201.1	433/17	433/176
KP22: Jaw	433/173	433/172	433/175	433/201.1		433/176
KP30: Jaw bone	433/173	433/172	433/175	433/201.1		433/176
KP47: Natural	433/173	433/172	433/175	433/201.1	433/17	
KP12: Root	433/173	433/172	433/175	433/201.1		
KP13: Tissue	433/173	433/172	433/175	433/201.1		433/176
KP21: Crown	433/173	433/172	433/175	433/201.1	433/17	
KP32: Distal	433/173	433/172				433/176
KP10: Cap	433/173	433/172	433/175			
KP11: Healing	433/173	433/172	433/175			433/176
KP14: Healing cap	433/173	433/172	433/175			
KP49: Adjacent	433/173		433/175	433/201.1	433/17	
KP40: Hexagonal	433/173	433/172				433/176?
KP41: Longitudinal	433/173	433/172	433/175			
KP43: Angle	433/173	433/172				
KP53: Art	433/173	433/172	433/175	433/201.1		433/176
KP46: Rotation	433/173	433/172		433/201.1	433/17	
KP50: Insertion	433/173		433/175	433/201.1		

APPENDIX 3 Validation and modification of ontology

Appendix 3-1 Dental implant tool ontology based on training patents



Appendix 3-2 Modified dental implant tool ontology



Appendix 3-3 Dental implant material ontology based on training patents

Appendix 3-4 Modified dental implant material ontology