Advanced Engineering Informatics 27 (2013) 346–357

Contents lists available at SciVerse ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Constructing a dental implant ontology for domain specific clustering and life span analysis

Charles V. Trappey [a], Tong-Mei Wang [b,*], Sean Hoang [a], Amy J.C. Trappey [c]

[a] Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan
[b] School of Dentistry, National Taiwan University, Taipei, Taiwan
[c] Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Available online 15 May 2013

Keywords:
Clustering
Key phrase extraction
Dental implant ontology
Life span analysis

ABSTRACT

Dental implant and prosthetics is a growing industry that follows the increasing aged populations that incur a higher percentage of tooth loss [1]. The dental implant sector is one of the most technical oriented fields in dentistry with many new techniques, devices, and materials being invented and put to clinical trials. Most innovations and technologies tend to be protected by intellectual property rights (IPRs) through patents. Thus, this research identifies the life spans of dental implant (DI) key technologies using patent analysis. Key patents and their frequently appearing phrases are analyzed for the construction of the DI ontology. Afterward, the life spans of DI technical clusters are defined based on the ontology schema. This research demonstrates the feasibility of using text mining and data mining techniques to extract key phrases from a set of DI patents with different patent classifications (e.g., UPC, IPC) as the basis for building a domain-specific ontology. The case study of ontological sub-clustering for dental implants demonstrates life span mapping of the technology and the ability to use clusters to represent stages of development and maturity in specific technology life cycles.

1. Introduction

Dental implants are a unique technology with a very wide range of applications and a huge market of approximately seven billion US dollar in 2011 [2]. Even though the technology for single tooth implants has been successfully used for over a decade, there are many conditions and uses of implants that are little understood and conditional. Many of the conditions of concern to dentists are long term survival and success rates that are influenced by many factors such as location of the implant, substitution (i.e., denture replacement), denture anchoring, tissue health, bone density, age of recipient, prosthetic complications, implant and abutment types, as well as materials and post-operative medicines [3]. Thus, it is a medical technology field that requires the combination of continuous technical innovation and clinical trials for improving the implant survival rates and reliability as well as reducing failure rates [4].

Huang et al. [5] describe ontology as a model which contains the concepts and the relational links of concepts in a specific domain that reflects the reality of the world. WordNet [6] defines ontology as a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations. Patent documents, and many technology oriented documents, contain domain specific terms which are not covered by common dictionaries. Therefore the advantage of ontology is that it defines a specific domain corpus to help analysts understand the meaning and relationships of the technical terms. Ontology can be seen as a hierarchical or network structure which abstracts domain concepts and relations expressed in terms of domain terminologies using a standard knowledge representation language [7] to facilitate knowledge sharing. Since data is growing rapidly with the creation of new patents, ontology development processes based on patents help keep knowledge bases current.

High technology companies strive to orient and align R&D strategic plans with emerging technologies. Patent documents are often publicly available through government databases and provide information that forms the foundation for technology trend analysis. Patent analysis has been used to formulate economic indicators that relate technology development and economic growth [8]. Recently, it has become strategically important to use patent analysis as a means for high technology companies to evaluate technology trends [9]. Companies face technology information overload and need tools to analyze growth trends of complex innovations and the development of products with increasingly shorter product life cycles. The demand for the rapid creation of new technologies or designs is expected to accelerate as the world marketplace

* Corresponding author. Tel.: +886 223562508.
E-mail addresses: trappey@faculty.nctu.edu.tw (C.V. Trappey), tongmeiwang@ntu.edu.tw (T.-M. Wang), long.hoang87@gmail.com (S. Hoang), trappey@ie.nthu.edu.tw (A.J.C. Trappey).

1474-0346/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aei.2013.04.003

becomes more integrated and information access is facilitated by the Internet [10]. Granstrand [11] points out that, during the industrialized age, tangible assets such as land, factories, and natural resources were the primary concern for management and growth. The knowledge-based age shows that intangible asset such as intellectual property, copyrights, and trademarks are now the focus of wealth creation and are used as leverage in the marketplace.

Clustering is a data analysis technique which classifies patterns of key phrases into categories based on the characteristics of relationship [10]. The objective of clustering is to measure the similarity in data and categorize it into groups that maximize the similarity of specified variables within the same cluster. For this research, the objective is to create homogenous clusters with the same context of data from patent documents so that each patent document belonging to a cluster should be similar and express related claims. Almeida et al. [12] notes that the presence of high connectivity among patent documents indicates a high association between the terms used in the documents. For patents, the challenge is to characterize trends and development for a business or industry [13]. Some researchers [14] use patent data and clustering analysis to analyze technology trends and developments, to track the growth and frequency of patent applications, and to determine or forecast the growth or maturity of patents.

This research uses patent analysis techniques to cluster dental implant patents and analyze the life spans of the clusters. The case study inputs a set of dental implant patents acquired from the United State Patent and Trademark Office (USPTO) to identify potential research opportunities for dental implants and to demonstrate the methodology's utility for technology forecasting. An NTF-based key phrase analysis tool is used to create a domain specific ontology and a visualization schematic is created with Microsoft Visio.

2. Background and related research

This section discusses text mining, patent analysis, clustering, and technology life cycle analysis. The text mining section introduces the advantage of using computers to process data when the volume of data is growing too rapidly to transform into knowledge. The patent analysis section describes using patent documents for the analysis of specific technologies, trend analysis, and the classification of technical terminology. The clustering section introduces the method of grouping patents into meaningful classifications. The technology life cycle analysis introduces the use of patent data to model technology life cycles.

2.1. Text mining

Text mining is a technique developed from data mining to analyze unstructured text documents including patents [15]. Text mining contains techniques to label documents and link them to specific words that facilitate analysis and knowledge creation [16]. A text document is often unstructured yet contains many types of information that can be ordered to represent facts and new knowledge [17]. Most information stored in a database is in the form of text documents. Text mining applies statistical algorithms for automatic knowledge discovery and pattern recognition [18]. Text mining is a broad field that includes information retrieval, text analysis, information extraction, clustering, categorization, visualization, machine learning, and data mining [19].

Recently, text mining has attracted researchers to use the techniques to study patents [9]. For example, text mining techniques have been used to create a patent maps for carbon nano-tubes [20]. Other researchers applied text mining techniques to automatically create categorization features with greater efficiency than human analysts [21]. One of the advantages of using text mining techniques is that large volumes of patent documents can be automatically sifted to extract useful information. Since patent documents are lengthy and contain unique technical terms and document formats, automatic text mining better enables researchers, engineers or managers to make decisions [15]. However, the extracted data should meet specific quality criteria to be understood by humans and to concisely represent the text concepts [9]. Text mining techniques have been applied to text segmentation, text summarization, feature selection, term association, cluster generation, topic identification, text mapping, technology trend analysis, and automatic patent classification.

Many researchers use specific indicators as determinants of patent value. By using different types of patent data sets including information about regional patent offices, particular technology sectors, or particular companies in a given country, new knowledge is acquired. Other researchers have studied patents and their impact on economic growth, technological innovation and development, and a country's overall competiveness [8]. Since on average only about 1 out of 50 patents generates significant financial returns, the identification and acquisition of high value patents with broad technical claims and high citation indexes can increase the financial value of a company. Companies with strong patent portfolios that conduct systematic and strategic patent planning activities are more successful than other companies especially in the fields of mechanical engineering and biotechnology [22]. Patent analysis can be effectively used for companies to gain competitive advantages in the global marketplace [8]. Finally, patents are easily accessible (and often freely) available throughout the world through databases managed by governments that insure their accuracy.

2.2. Patent analysis

Patent documents contain rich and detailed information about research results that are expressed in complex technical and legal terms that is invaluable to the industry, legal practitioners, and policy makers [23]. The detailed content of patent documents, if carefully analyzed, can reveal areas of technology development, inspire novel technical solutions, show technical relations, or stimulate investment policies [21]. Tseng et al. [20] point out that patent analysis has become important at the government level for policy formation. Countries are investing resources to depict technical and commercial information that can be turned into knowledge [23]. Patent documents are often lengthy and require time, effort and expertise to interpret. Tseng et al. [20] emphasize that patent analysts require expertise in information retrieval, knowledge of domain specific technologies, legal knowledge, and business intelligence to be effective.

Patent analysis can be divided into two levels; macro level research of national or industrial technology development and micro-level research of specific technology development for claims analysis and forecasting [23]. Macro level analysis evaluates the economic effect of technological innovations, technological development and the competitiveness of countries [8]. Micro-level analysis identifies the development of specific technologies, the advantages and disadvantages of competitors, aids the strategic planning of R&D activities, and identifies relations between companies and technologies [24].

2.3. Ontology used to represent domain knowledge

Ontology structures concepts that reflect the reality of the world [5] and defines common terms in a domain of interest including the relationships among these terms. Ontology used for knowledge extraction via data mining has been applied to various fields. For example, an ontology tree can be used for automatic patent document summarization which extracts key information into shortened abstracts describing the key concepts [10]. The goal is to use the ontology to create a knowledge base as input for a software program that improves the capturing of information and the creation of knowledge. A biomedical gene ontology that helps researchers accelerate knowledge acquisition, structure complex biological domains and relate data is now considered a significant competitive resource [25].

Ontology links the semantic data between concepts which makes it possible to perform pattern recognition, similarity analysis, and clustering of patent documents with respect to content [26]. To create a domain specific ontology for patents requires key phrases that describe the concepts of patent documents [10]. A variety of methods have been proposed to create knowledge domains and one of the methods suggests a single ontology that integrates all knowledge domains. The potential drawback of this method is the lack of scalability which narrows the usefulness of the information. Researchers recommend creating a small or niche domain ontology and then integrating several into a top level ontology [27]. The same approach has been used to capture patent knowledge and enhance information retrieval.

2.4. Patent document clustering

Clustering facilitates key phrases into categories based on the characteristics of relationship [10]. The similarity in data is measured to create the most suitable clusters. The clusters maximize the similarity of specified variables within and create homogenous content representing similar patent documents. There should be a high level of connectivity among these patent documents with a high association between technical terms [12]. The challenge of patent analysis is to characterize technical trends and development for a business or industry [13]. Researchers have developed clustering techniques for patent documents which analyze technology trends and track the growth and frequency of patent applications to forecast the life cycle of patents [14].

One way to mathematically define the similarity between two objects is based on the Euclidean distance [12]. Other researchers [10] use an equation called the Manhattan distance. Patents may be used to cluster groups of technology based on their knowledge content rather their International Patent Codes (IPCs) or United States Patent Codes (UPCs). Patent technology clustering of this type uses a key phrase correlation matrix as input and applies the K-means algorithm to form the clusters [10]. A more complete discussion on applications of the K-means algorithm is provided by Han et al. [28]. The Root Mean Square Standard Deviation (RMSSTD) is the standard deviation of all variables and represents the minimum variance in the same cluster. Therefore, the value of RMSSTD should be as small as possible to gain optimal results. The R-Square (RS) value describes the maximum variance between different clusters and the value of RS should be as large as possible since RS is the sum of squares between different clusters divided by the total sum of squares for the set of data. Thus, RMSSTD and RS are used to find the optimal number of clusters for a set of data.

Patent document clustering uses the correlation matrix generated from patent technology clustering as the K-means algorithm input [10]. Patent technology clustering splits patent documents into groups according to the similarity of key phrases in each patent document. The key phrases represent the dominant technology depicted in the patent documents. Finally, patent document clustering measures the internal relationship of the key points of the patent document and classifies patent documents based on the similarity of the technologies which enables patent analysts to identify the characteristics of the clusters.

2.5. Technology life cycle analysis

Life cycle analysis, as the name implies, assesses of the development of a product or service, from initial extraction of raw material to the final output or disposal of the product. When companies invest R&D capital on technologies, the investment decision often depends on the current life cycle stage of the technology [29]. Patent documents reveal the technical development and the life cycle stage of an industry [30]. A patent or patentability of a technology is also a precondition of commercial potential. In addition, patent documents contain data about the patent application date which relates to the life cycle of different products and the trends of commercialization and market development. The concept of technology life cycles is similar to product life cycles which include four stages including introduction, growth, maturity, and market decline. Regardless of the reference factor used to define the technology life cycle, patent based life cycles usually begin earlier than product development and commercial cycles [29].

The start of the patent life cycle introduction stage is often fraught with fundamental scientific problems that are not yet fully overcome. These technical problems have to be solved in order to advance and researchers often struggle to achieve radical innovations. At this stage, the patent applications are low but slowly increasing since there is a lot of uncertainty and few pioneer firms are willing to take the R&D risk [14]. During the early patent growth stage, the patent applications per applicant increase since the problems of the innovative technology are resolved. However, the cost may still be too high for customers' acceptance or standardization of the product. During the growth stage, when the fundamental technical problems have been solved and the market uncertainty has been replaced with reliable products, many new competing products are likely to appear stemming from the earlier technological advances. Since the R&D risk has decreased, other inventors attempt to find competing alternative solutions and there is an increase of patent applications. The growing number of patent applications 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 all new features developed for this technology have been commercialized for the market. Thereafter, the technology enters the decline or saturation stage when new products and technologies are introduced.

Patent activity is an important indicator of the current technology life cycle [29] but verification requires a statistical survey of all patent applications of a given technological field [30]. In order to simplify analysis, the S-curve methodology can be used to study niche market developments such as pacemaker technology. All cumulative patent applications for a specific technology over a certain period of time can be plotted as an S-curve and the different technology life cycle stages can be forecasted [30].

3. Methodology for dental implant patent ontology engineering

This section describes the methodology and the research framework to achieve our case research objectives. This section describes the procedure from data selection to key phrase analysis, building the ontology and creating the domain-specific ontological clustering of patents.

3.1. The framework for dental implant patent ontology engineering

There are five steps for building the domain specific ontology based on patent data. Fig. 1 presents the procedural framework for systematic ontology building. This procedure is called Domain Specific Patent Ontology Engineering (DSPOE). The DSPOE is based

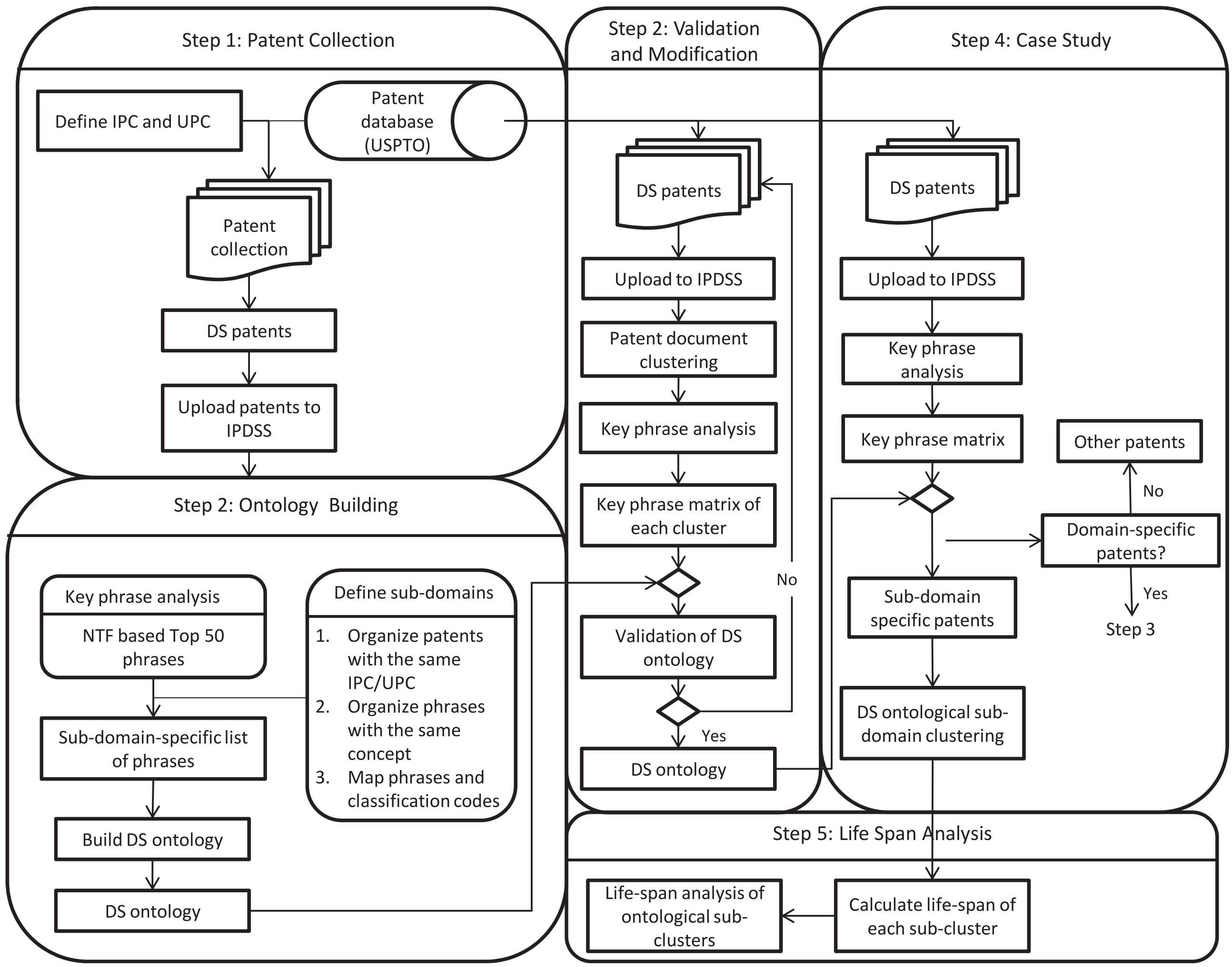

Fig. 1. Steps of building and applying a domain specific (DS) ontology.

on domain specific patent data. The concepts, construction steps, and applications are described in the following sub-sections. The framework identifies the domain of interest and collects the domain specific (DS) patent documents. Afterward, key phrases are extracted from the DS patent documents. The sub-domains are defined that identify the ontological sub-domain concepts and relationships. After building the initial DS ontology, it is verified and modified so that a case study can be conducted using a validated ontology. The Intellectual Property Defense-based Support System (IPDSS) software [31] was used to automatically extract key phrases, build the key phrase matrix, and cluster patent technologies and documents. Fig. 1 shows the steps of building and applying the DS ontology. The detailed procedures are described in Section 3.3.

3.2. Patent key phrase analysis

Most information stored in databases contains text documents. Extracting key phrases makes it possible to determine which document is important and to identify the relation among several documents. Key phrase extraction is useful for document or information retrieval, document clustering, summarization, and text mining [32]. There are many useful applications for key phrase extraction including highlighting key phrases in text, document classification, text compression, or constructing human readable text. Statistical approaches are used to 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 the term frequency approach. Some researchers divide key phrases extraction algorithms into two categories [33] that requires supervised learning and are applied for single documents and unsupervised key phrase extraction using self learning which is also known as knowledge discovery (KDD).

Key phrases extraction has been applied in many different fields, although mainly for summarization purposes [34]. For example, research on the impact of automatic summarization systems based on key phrase extraction compared to human summarization showed that the key phrase frequency methodology generated summaries comparable with humans [35]. Other researchers use a hierarchy and semantic relationships to create a patent summarization system based on the specific domain of the patent document.

In this research, the key phrase analysis applies the normalized TF-IDF (NTF) methodology to extract key phrases. The TF-IDF method calculates weights for frequent key terms in a series of documents to determine relevance. Frequent key terms in one document cannot represent a domain but frequent key terms in a series of documents might represent the concept of the domain [36]. The formula for IDF [37] is defined 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 contain term i. The variable idf_i represents the inverse document frequency (IDF) of the term i. The equation describes idf_i as a value representing term i and if idf_i becomes a significantly high value, then the term i represents a specific document.

The weighting of key phrases using TF-IDF in text documents where TF are weighted in IDF is expressed as:

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

where w_{ik} is defined as the weight of term i in document k of the collection, tf_{ik} is the number of terms i that occur 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 frequently occurring key phrases in a specific text document and are identified as the key phrases for any document k.

Furthermore, it is necessary to normalize TF-IDF because the TF-IDF method does not consider the difference between the number of words in each document, therefore the frequency weights of key phrases are normalized by the number of words in each document. The normalized term frequency (NTF) is expressed as follows:

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

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 r is the total number of documents in the document collection.

The key phrase correlation matrix calculates the correlation of important key phrases (KPs) in each patent document which is used to create the logical link between concept and methodologies. The use of NTF-IDF to calculate the correlation between key phrases to create a key phrase correlation matrix using inner product of vectors is expressed as:

$$\begin{aligned} Correlation(KP_i, KP_j) &= \frac{KP_i \cdot KP_j}{\|KP_i\|\|KP_j\|} \\ &= \frac{\sum_{k=1}^{n} w_{ik} \times aw \times w_{jk} \times aw}{\sqrt{\sum_{k=1}^{n} w_{ik}^2 \times aw^2 \times \sum_{k=1}^{n} w_{jk}^2 \times aw^2}} \end{aligned} \tag{4}$$

where $KP_i = aw(w_{i1}, w_{i2}, \ldots, w_{in})$ and $aw = \frac{\sum_{s=1}^{n} WN_s}{n \times WN_k}$ is the average Word Number (WN). The algorithm consists of four stages. First, the algorithm transforms the patent document into a key phrases vector and analyzes the frequencies of key phrases. Second, it derives the key phrase vector by eliminating unnecessary phrases. Third, the correlation values between key phrases are calculated using Eq. (4). Fourth, the correlation coefficients are calculated based on the number of different key phrases occurring in each patent document.

The key phrase correlation matrix is used as an input for patent technology clustering. The key phrase correlation matrix represents the technology in each patent document and thus represents the internal relationship among patent documents instead of clustering patents according to classification codes such as United States Patent Classification (UPC) or International Patent Classification (IPC).

For 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_n occurring among $Patent_1$ to $Patent_n$. NTFR is the product of NTF and Rate which expresses the relevance of KP_n among the patent collection. The key phrase, KP_n, is a representative phrase in $Patent_n$. If the frequency F_{nm}, is large enough across $Patent_1$ to $Patent_n$, then KP_n is a representative phrase of $Patent_n$. The key phrase and patent correlation matrix are shown in Table 1.

Table 1
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_n	...	...	...	...	F_{nm}	...	...	...

3.3. The steps of DSPOE procedure

This section describes the DSPOE steps proposed in Fig. 1. In Section 4, the DSPOE framework is applied for the dental implant patent analysis.

3.3.1. DSPOE 1: Define and collect domain specific range of patent documents

The first step of the proposed methodology is to identify the range of the ontological domain and focus on specific patents. The approach of using International Patent Classification (IPC) and United States Patent Classification (UPC) is applied to understand the scope of the ontology and define the sub-domains of the ontology. The patents are downloaded from United States Patent and Trademark Office (USPTO) and domain specific patents are extracted based on the patent's first listed UPC. These patents are uploaded to the IPDSS platform to automatically extract key phrases and provide statistical analysis of the document metadata.

3.3.2. DSPOE 2: Key phrase analysis and ontology construction

After the domain range is defined, the next step is to establish a list of key phrases in the specified domain and define sub-domains based on the key phrase list. The key phrase analysis is based on the NTF-methodology which generates a list of the top 50 key phrases from the patent data collected. Table 2 shows an example of the key phrase matrix. Patents are organized and grouped according to their UPC and WordNet is used to define keywords and the relationship between keywords in the key phrase list to organize and group sub-domain phrases. The final step is to classify key phrases and UPCs in a key phrase matrix to generate an overview of phrases expressed in different UPCs. The goal is to form a preliminary matrix of domain specific knowledge and keywords for the ontology building process.

Based on the sub-domain list of key phrases from DSPOE 2, the domain specific ontology is engineered using Microsoft Visio as a building and visualization tool. The list of the top 50 sub-domain phrases are linked based on their concepts and relationships. Top-down classification starts from the upper phrases and then extends to the lower phrases to establish an ontology tree with relationships links.

Table 2
Partial key phrase and patent correlation matrix.

Key phrase/PAT. no.	$Patent_1$	$Patent_2$	$Patent_3$	...	$Patent_n$	NTFR
KP1: Implant	75	55	33	...	...	14,023
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
...	...	...	...	...	...	...
KP_n	...	...	...	...	F_{nm}	$NTFR_n$

3.3.3. DSPOE 3: Validation and modification of the domain specific ontology

This research uses domain specific patent data for patent clustering to create sub-domains of the ontology. Key phrase analysis based on the NTF-methodology is applied to each sub-cluster to generate a sub-domain-specific list of the top 15 phrases. These phrases are used to validate the ontology and check that these phrases are included. If the phrases do not match, the ontology is modified and the process is repeated until the ontology created is strong enough to capture the domain specific knowledge of the sample of patent documents. In this step, domain experts are consulted to verify, validate, and modify the DS ontology schema.

3.3.4. DSPOE 4: A case study of ontological sub-domain clustering

In order to test the domain specific ontology, the methodology requires the input of new domain specific patents that have not been utilized previously and 30 new domain specific patents are downloaded. The methodology applies key phrase analysis based on the NTF-methodology and generates a key phrase matrix with frequencies of each key phrase for each patent document. The list of phrases is used to classify patents into sub-domains for ontological sub-domain clustering. Only dental implant patents are used to create the ontological sub-domain clustering for life span analysis.

3.3.5. DSPOE 5: Life span analysis

After the ontological sub-domain clustering, the average application age is calculated for all patents in each cluster. The application date is the date when a patent application is officially handed into a government agency that issues patents. Each ontological sub-domain cluster is plotted against the average life span of the whole cluster. The age of the patent is calculated using the application date as a starting date and not the issuing date. The average age of each cluster is plotted against the ontological sub-domain clusters. Fig. 2 illustrates the analysis of potential emerging or declining clusters depending on average age. The size of each bubble represents the number of patents. The *Y*-axis plots the ontological sub-domain clusters and the *X*-axis is the average age of each cluster. Cluster 5 in Fig. 2 represents a young cluster of an ontological sub-domain which is a specific sub-domain of dental implants. The mapping method allows researchers to explore which sub-clusters have potential for further development or which sub-clusters may soon become outdated.

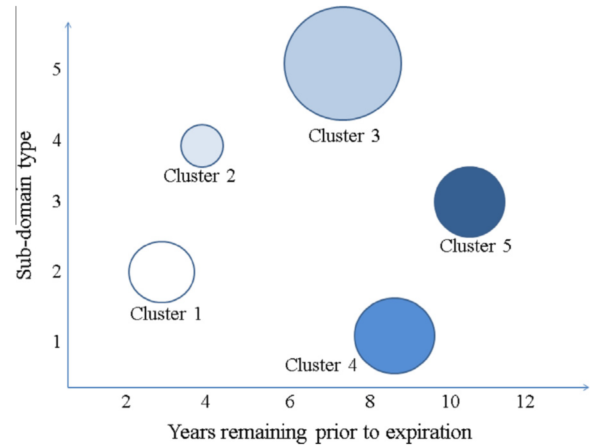

Fig. 2. Proposed life span analyses of dental implant patent clusters.

4. Ontology based clustering for dental implant patents

In this section, dental implant patents are used as a case to demonstrate ontological sub-domain clustering based on patent data. The following discussion describes dental implants and the components. Various components of dental implants are the implant body, the cover screw (prevents bone access), the trans-mucosal abutment (links the implant body to the mouth), the healing abutment (temporarily placed on the implant to maintain the mucosal penetration), the healing caps (temporary covers for abutments), the crowns, bridges, and gold cylinder (to fit an abutment and form part of the prosthesis), and the laboratory analogue (a base metal replica of implant or abutment) [38,39]. The main components of a dental implant also include a screw that connects to a custom-made crown.

4.1. Dental patent document sampling

Patents under the same UPC may be entirely different in technology. Therefore, a large sample including different UPCs is included when collecting data for building a domain specific ontology (Table 3). The IPDSS software completes the data preprocessing and key phrase extraction. Then the key phrase correlation measures are used to create a key phrase and patent correlation matrix. IPDSS uses K-means clustering as its algorithm for patent document clustering.

Table 3
List of dental implant patents in UPC classifications and dimensions.

UPC	Number of patents	UPC definition
433/173	97	**By fastening to jawbone:** Subject matter wherein the denture is secured directly to the jawbone of the patient
433/174	24	**By screw:** Subject matter wherein the denture is secured to the jawbone by an elongated helically ribbed member
433/175	4	**Shape of removed tooth root:** Subject matter wherein the lower portion of the denture that is secured to the jawbone is shaped to correspond to the configuration of the root of a natural tooth which had previously occupied the same position in the mouth
433/176	4	**By blade:** Subject matter wherein the denture is secured to the jawbone by a flat plate-like member extending from the bottom of an artificial tooth
433/172	13	**Holding or positioning denture in mouth:** Subject matter relating to locating or securing one or more artificial teeth in the mouth
433/201.1	5	**Dental implant construction:** Subject matter relating to either the structure or a process of making a dental prosthesis which is adapted to be fixed to the jawbone
433/169	7	**Stress breaker:** Subject matter wherein a denture includes means to redirect or absorb forces during mastication to protect the denture from damage
433/17	2	**Having arch wire enclosing guide (e.g., buccal tube):** Subject matter wherein the bracket includes an elongated member having a passage therein through which the arch wire is placed
Total patents	156	

Table 4
Dental implant key phrase and patent correlation matrix (partial).

Key phrase/PAT. no. UPC	US6312260 433/174	US6039568 433/175	US5297963 433/172	US5362235 433/172
Implant	177.46	165.01	51.01	49.2
Dental	29.09	84.28	25.7	23.74
Bone	7.17	21.37	9.49	10.79
Screw	40.04	8.31	16.61	32.8
Abutment	0	26.12	34.4	71.64
Thread	42.15	14.25	24.91	26.33
Bore	14.75	0	32.03	26.76
Prosthetic	0	4.75	0	0
Cap	140.37	4.75	138.01	91.92
Healing	141.63	7.72	141.97	95.38
Root	0	3.56	12.65	6.9
Tissue	6.32	4.15	11.86	11.22
Healing cap	125.61	4.75	133.27	88.47
Fixture	0	0	45.87	44.88
Cavity	0	0	6.72	7.77
Hole	5.48	8.31	0	6.04
Jaw	5.06	0	8.7	9.93
Jawbone	18.13	20.18	0	0
Implant fixture	0	0	35.59	32.37

4.2. Key phrase and patent correlation matrix

The key phrase and patent correlation matrix is derived from the dental implant patent data. The top 50 key phrases are chosen in chronological order from the highest NTF-value. Table 4 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 implant main components [38,39]. For example, the abutment (support for crown) and the healing cap (covers abutments) are both listed in the matrix. The UPC 433/174 is described as fastening implants to the jawbone by screw and from Table 4 the key phrases listed are jawbone, threads, screw, and hole which conform to the UPC. Another example is UPC 433/172 – holding or positioning the denture in the mouth. Table 4 lists key phrases including embodiment, bore, and implant fixture. Patents with the same classification code may not be expressed by the same set of key phrase which supports the reason to include patents from several different classifications to create an ontology that captures the main concepts of the domain.

4.2.1. Sub-domain definition of key phrases and the patent correlation matrix

The key phrases are sorted and grouped into four large groups. Table 5 shows 2 sub-domains for dental implant dimensions where the key phrases are logically grouped to demonstrate their related concepts. For instance, dental, implant, prosthetic, and artificial are in one group, while screw, threads, and titanium are in another group. The grouping enables the creation of sub-domain clusters for the ontology schema.

Table 5
Two sub-domains and representing key phrases (partial).

Sub-domain	Key phrase	UPC					
1	Implant	433/173	433/174	433/172	433/169	433/175	433/201.1
1	Dental	433/173	433/174	433/172	433/169	433/175	433/201.1
1	Artificial	433/173	433/174	433/172	433/169		433/201.1
1	Prosthetic	433/173	433/174	433/172			433/201.1
2	Screw	433/173	433/174	433/172	433/169	433/175	
2	Threaded	433/173	433/174	433/172	433/169	433/175	433/201.1
2	Thread	433/173	433/174	433/172	433/169		
2	Titanium	433/173	433/174	433/172	433/169	433/175	433/201.1

4.3. Building the ontology

The proposed life span analysis of the dental implant patents uses the ontology as a variable for clustering dental implant patents. The key phrases of each dimension are grouped as shown in Table 5 and represent the sub-domains of the ontology. The ontology in this research is an adapted version of Pritzek's RFID ontology tree [40]. The ontology of dental implants, shown in Fig. 3, only uses phrases from the key phrase matrix to link phrases to their concepts and relationships. Patent documents that contain detailed information about research results are written using complex technical and legal terms, so it is preferred to extract data from patents to build a domain specific ontology. Building the ontology from health industry patents, particularly dental implants, has not yet been studied. Therefore, it is unique to analyze clusters using an ontology based on dental implant patents. The ontology in Fig. 3 shows four preliminary sub-domains of the dental implant dimensions which are classified as geometry, implant fixture, biological, and dental components. The ontology is divided into sub-domains to separate and provide more specific concepts relevant to the dental implant domain.

The ontology is often built by domain experts and is subjective. In this research, part of building the ontology is subjective since the linking concepts and phrases are based on WordNet and the opinion of the researchers of this report. However, constructing a domain specific ontology in the dental implant area based on patent data using objectively extract phrases by computer software creates an ontology that is more robust for the analysis of dental implant patents clusters.

4.3.1. Validation and modification of the ontology

One method to validate the ontology is to use key phrases defined by experts that are familiar with the domain. The experts may also compare illustrations in each patent document if the patents include a figure of an implant body. This initial research only uses key phrases to group dental implant patents and the clusters are based on the similarity of technology. The key phrases extracted for dental implants are shown in Table 6. Patent document clustering is applied and for each cluster, key phrase extraction is used to extract the top 15 key phrases based on NTF-values. If more phrases are extracted, it will only generate a larger and more complex ontology. Therefore, 50 phrases are used to build the ontology and these key phrases are used to validate and modify the ontology (Fig. 4).

The comparison of key phrases from Table 6 with the ontology in Fig. 3 shows that the ontology has to be modified since some key phrases from each cluster in Table 6 do not match the sub-domains in the ontology from Fig. 3. The reason is that each sub-domain includes repeated key phrases to describe the concept of that sub-domain. In Fig. 3, the sub-domain "screw" should also contain links to jaw bone, fixture, attachment, and crown which describe the concept and sub-domain of "screws" more accurately. Analyzing the phrases in cluster 1 (Table 6) shows that the terms are more likely to belong to the sub-domain of "screws" than other sub-domains.

The ontology in Fig. 4 includes four sub-domains which are implant, implant assembly, screw device, and implant fixture. The validation of each dimension of the ontology was completed after repeated validation and modification by the domain expert. Fig. 3 depicts the initial ontology and Fig. 4 is the result modified to include shared phrases that describe the core concept. However,

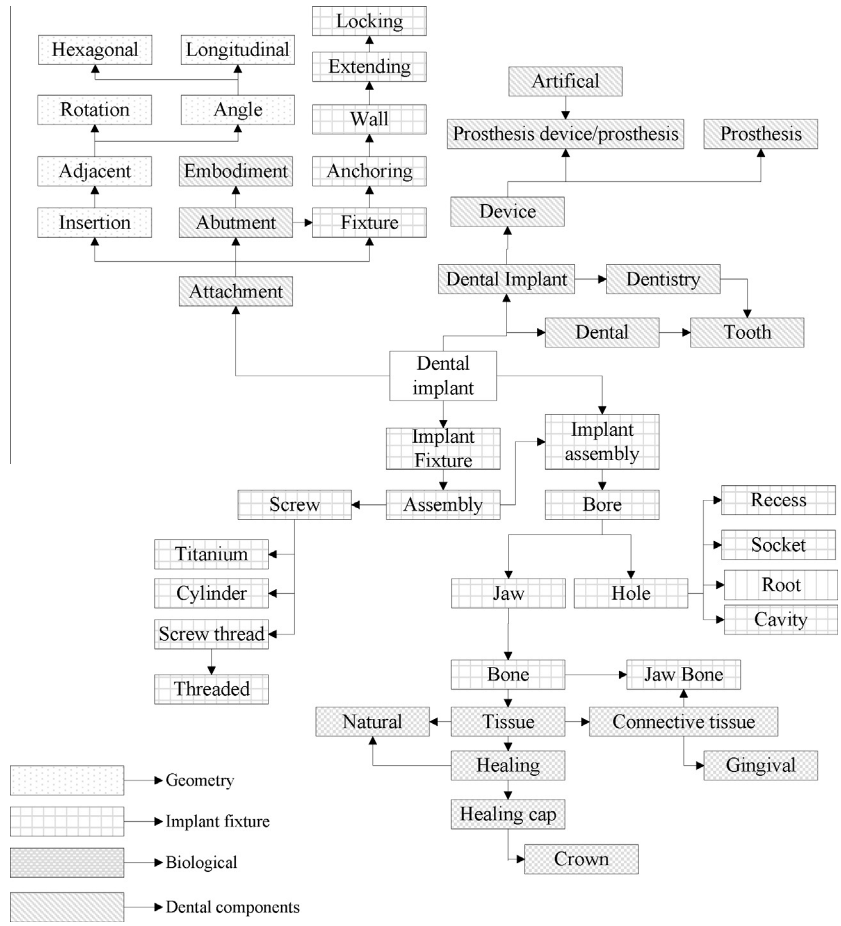

Fig. 3. Preliminary dental implant ontology.

Table 6
Key phrases for validation of implant ontology.

Cluster 1 Key phrases	Cluster 2 Key phrases	Cluster 3 Key phrases	Cluster 4 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: Titanium (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

including too much detail and sharing too many phrases among technology sub-domains weakens the ontology and decreases the ability to build strong and unique clusters.

The implant sub-domain in Fig. 4 includes many shared phrases and includes few unique phrases that are distinct in the ontology whereas the implant assembly sub-domain includes several unique phrases which increase the cluster quality. The screw device sub-domain also has several unique phrases which build a stronger cluster compared with the implant sub-domain. The implant fixture sub-domain includes unique phrases. However, this sub-domain includes distinctive phrases which are easily separated from the screw device or the implant assembly domain. For example, extending, anchoring, rotation, and angle may also be combined with embodiment, insertion and attachment.

4.4. Life span analysis of dental implant clusters

For this research, a case study of the life span analysis was based on the dental implant ontology. Key phrase analysis of 30 test patents created the 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 the dental implant ontology (Fig. 4). 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. Table 7 shows the results of the key phrase analysis of the test patents and the number of patents in each cluster. The analysis includes a column of "other patents" where the dental implant ontology failed (e.g., patents that include key phrases such as "dental implant package" in combination with "healing screw"). The dental implant ontology requires minor modification to include patents where there are potentially overlapping clusters.

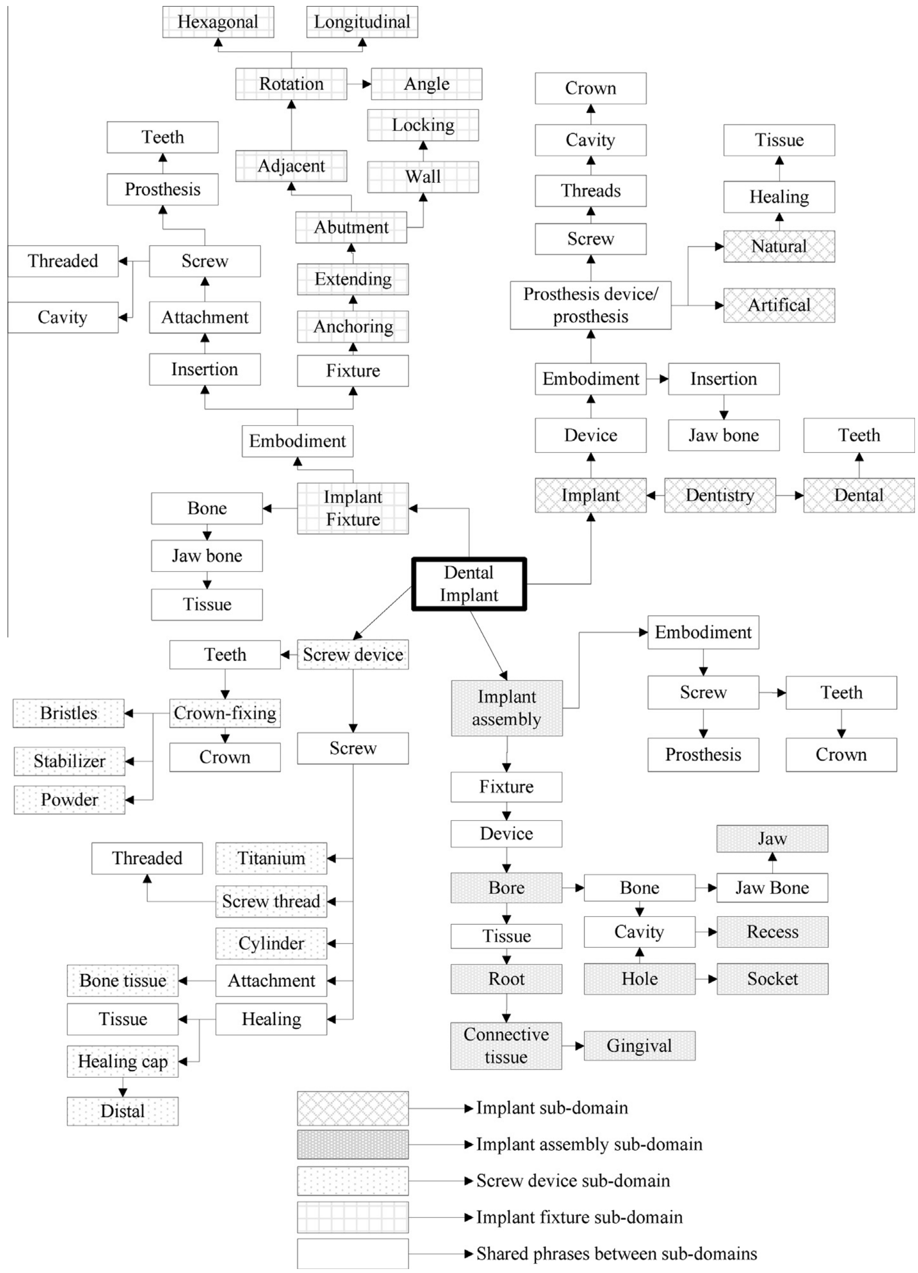

Fig. 4. Modified dental implant ontology.

However, the objective of this research is to focus on dental implants and specified UPCs which results in an initial ontology that captures the most relevant patents and excludes unnecessary patents.

In Table 7, there are new phrases which indicate that the ontology requires further improvements as a result of the initial training patent sample used for building the ontology. The 30 test patents in this case study did not restrict any UPC or IPC, and as long as the title included "dental implant" it was collected for analysis. Results also depend on the age of the patent, for example, patent US5022860 in Table 8 is an expired patent that is 23 years old. Changes in terminology over the years for dental implants also have an impact on the analysis. All test patents use the first UPC that matches the training patents in Table 3 and several relevant classification codes were included. Although the training patents have different UPCs, the dental implant ontology constructed is able to separate patents with the same UPC but different technology sub-domains. This result supports previous research [14] that patents in the same classification codes may be entirely different in technology.

An example of the implant fixture sub-domain is shown in Table 8 and the life span is calculated from the application date of current patents. The sub-domains implant assembly, screw device, and implant fixture include expired patents since the test patents samplings included random dental implant patents. The life span of the sub-domain implant assembly is about 14 years (excluding expired patents, about 11 years) and the life span for screw devices is about 13 years (excluding expired patents, about 12 years). Each ontological sub-domain patent cluster is plotted against their clusters average age including other patents and these are plotted without expired patents. Fig. 5 shows that implant assembly and implant fixture are the two out of three sub-domains that are considered to have potential. In this research the 30 test patents resulted in 7 patents being excluded since it was considered these

Table 7
Partial list of phrases for each ontological sub-domain of test patents.

7 Patents Other patents	4 Patents Implant assembly	5 Patents Screw device	14 Patents 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
Tissue	Embodiment	Embodiment	Embodiment
Prosthesis	Tissue	Tissue	Tissue
Healing screw	Prosthesis	Prosthesis	Prosthesis
Insertion	Insertion	Threads	Threads
Dental implant package	Jawbone	Extender	Extender
Package	Dental prosthesis	Healing screw	Healing screw
Dental prosthesis	Crown	Insertion	Insertion
Implant package	Teeth	Jawbone	Jawbone
Crown	Device	Dental prosthesis	Barrel

Table 8
Implant fixture sub-domain patent information.

Patent No.	Patent title	UPC	Filing date	Age
US5571016	Dental implant system	433/173; 433/169	January 24, 1995	16
US5752830	Removable dental implant	433/173; 433/169	June 20, 1996	15
US5863200	Angled dental implant	433/173	August 7, 1997	14
US5931674	Expanding dental implant	433/173	December 9, 1997	14
US6171106	Cover screw for dental implant	433/173; 433/174	September 9, 1999	12
US6431867	Dental implant system	433/173	August 10, 2000	11
US6500003	Dental implant abutment	433/173	June 14, 2001	10
US7341453	Dental implant method and apparatus	433/173	June 22, 2004	7
US7708559	Dental implant system	433/174	May 14, 2004	7
US6099312	Dental implant piece	433/174	July 15, 1999	12
US5951288	Self expanding dental implant and method for using the same	433/173; 433/175; 433/201.1	July 3, 1998	13
US7112063	Dental implant system	433/174	August 11, 2004	7
US7396231	Flared implant extender for endosseous dental implants	433/173; 433/172; 433/174	March 7, 2005	6
			Average age (in years)	11.1
Expired patents in cluster				
US5022860	Ultra-slim dental implant fixtures	433/174	December 13, 1988	23
			Total average age	12.9

patents were inappropriate. According to the literature review, the sub-domain implant assembly shows potential due to a small number of patents at an early stage of development. The sub-domain implant fixture can be seen as the dominant cluster in this

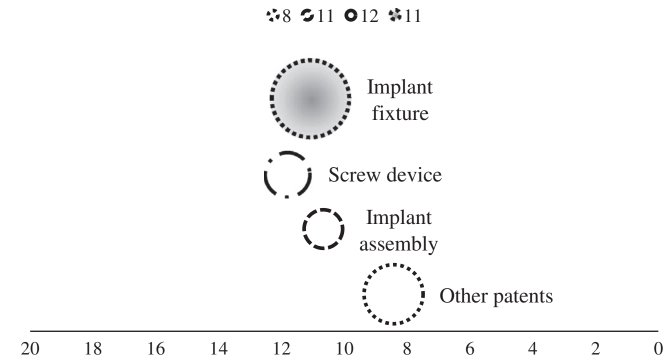

Fig. 5. Life span of dental implant clusters without expired patents (years in reverse scale).

case and with an average age of about 11 years indicating potential opportunities for development and investment. Fig. 6 shows a histogram comparison of the sub-domain clusters.

The comparison in Fig. 6 shows that there are differences when mapping the average age of patents in clusters. For example, the sub-domain cluster screw device is a young cluster with an average age of about 12 years (without expired patents) and the field of implant fixtures has potential for further development. However, including the 23 year old expired patent affects the average age and makes the implant fixture cluster seem less attractive for R&D investments. From these test patents, the implant assembly sub-cluster is the youngest and is in the introductory stage with potential growth opportunities. The similarity of implant assembly and implant fixture might overlap in the ontology, hence, implant assembly focuses more on the surroundings like drilling holes or biological aspects including tissues or a device. Implant fixtures focus more on the implant body 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 since it appears in the introductory stage and its ontological sub-domain includes several unique key phrases which support the strength of the dental implant ontology. However, the results require improvements of the ontology to capture several unique key phrases to better describe the sub-domain. The screw device and implant fixture sub-domains are also strong with several unique phrases. The implant sub-domain appears weak since it did not capture any patents but depends on the test patent samplings. Both screw device and implant assembly sub-domains results demonstrate signs of growth. The small sampling of test patents makes it rather difficult to draw a conclusion whether the clusters are in the frontier or laggards in technology develop-

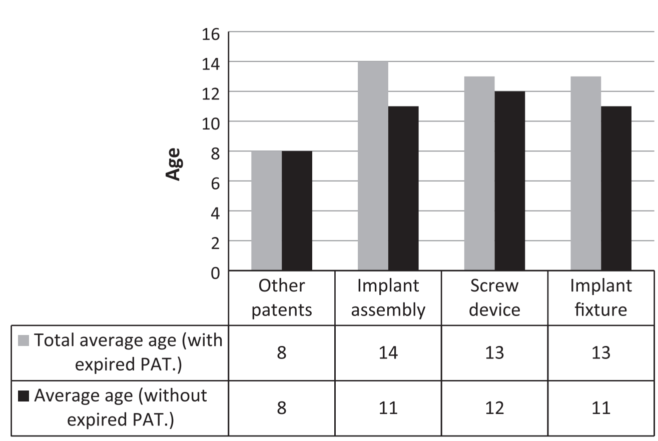

Fig. 6. Comparison of sub-domain clusters (rounded value).

ment. One must also take into consideration that the number of test patents in each sub-cluster is different. A more objective analysis requires a fair number of test patents and an 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 research presents a new and valid means of clustering patents and determining which clusters have the potential of growth or may be declining. Life span analysis of clusters is one of the many lifecycle analysis techniques and can be considered as an overview cluster analysis for mapping domain specific technologies for further detailed analysis of the technology life cycle. Patents have a lifetime of 20 years and depending on the clustering technique used, may reveal which cluster is moving towards growth or maturity. However, a mature cluster may enter the growth stage again if the patent activity increases for that cluster. Therefore, it requires the researcher must constantly update the clusters and create a timeline before concluding its stage of the life cycle and potential.

Life span analysis of domain-specific clusters includes patents within a limited time period which map the growth of each cluster and the change in average age. For example, by including patents from the years 1995 to 2005 and comparing with patents from 2000 to 2010, the historical development of technologies are analyzed. Furthermore, it is possible to map historical technology barriers critical to overcome or avoid. As such, a cluster in the mature stage such as a screw device (Fig. 5) can return to the growth stage by increased patent activity in this domain.

5. Conclusion

This research studies the feasibility of using patent analysis techniques to build and verify a domain specific ontology using patent analysis techniques. A case study is used to cluster dental implant patents using the dental implant ontology and examine the life span of these clusters. The analysis supports the use of text mining techniques to extract key phrases to build a domain specific ontology. The validation methodology is reliable and feasible although it requires further research to gain increasingly significant results. The case study of the dental implant ontology demonstrates that the patent sample consisting of several patent classifications has similar technology even though classified in different classes. The dental implant ontology also demonstrates a means to create specific sub-domains to sub-cluster dental implant patents with the same classification code including clustering patents in other classifications even though not included as training patents. The construction of the dental implant ontology based on patent data provides a means of clustering patents based on their technology concepts. The ontology is flexible and new key phrases can be added, deleted and adapted for creating a more specific domain ontology.

The life span analysis of patent clusters is based on the technological life cycle and by mapping these clusters potential opportunities for future development can be identified. With a consistent clustering technique, the life span analysis of patent clusters provides an overview of potential or future trends in technology development. However, a potential problem may possibly be that patents or technologies can overlap in clusters which will require developing a methodology to separate these overlaps. The ontology only captures relevant patents and excludes patents that did not match the sub-domain concepts. The results indicate that the dental implant ontology is robust and domain specific for dental implants. Other domains may be included in the ontology for improvement since this research only focuses on the dental implant body, abutment, crown, and fixture and excludes patents that focus on dental implant packages.

The life span analysis of ontological sub-domain clusters provides an overview of the domain specific clusters and their current life span position to support R&D decision making. Each cluster can gain competitive advantage again through increased patent activity which lowers the average age of each cluster. Consideration of expired patents in future research should provide a detailed analysis of sub-cluster development over time. The advantage is to provide a visualization of the development of technology barriers (historical and current) to determine if sub-clusters gain competitive advantage through increased patent activity.

Acknowledgements

This research was partially supported by National Science Council research projects. The authors express their gratitude to Dr. Chun-Yi Wu for his assistance in running the case analysis using the IPDSS software tool (www.wheeljet.com.tw/edu/).

References

[1] World Health Organization, Oral health, 2012. <http://www.who.int/mediacentre/factsheets/fs318/en/> (retrieved 01.12.12).
[2] Ceramic Industry, Dental implants and prosthetics market continues growth, 2012. <http://www.ceramicindustry.com/articles/92515-dental-implants-and-prosthetics-market-continues-growth> (retrieved 25.11.12).
[3] C. Mangano, A. Piattelli, G. Lezzi, A. Mangano, L. La Colla, Prospective clinical evaluation of 307 single-tooth Morse taper-connection implants: a multicenter study, The International Journal of Oral and Maxiofacial Implants 25 (2) (2010) 394–400.
[4] R.E. Jung, B.E. Pjetursson, R. Glauser, A. Zembic, H. Zwalen, N.P. Lang, A systematic review of the 5-year survival and complication rates of implant supported single crowns, Clinical Oral Implants Research 19 (2008) 119–130.
[5] C.-J. Huang, A.J.C. Trappey, C.Y. Wu, Develop a formal ontology engineering methodology for technical knowledge definition in R&D knowledge management, in: Proceedings, 15th ISPE International Conference on Concurrent Engineering (CE 2008), August 18–22, Belfast, N. Ireland, UK, Springer-Verlag, London, 2008, pp. 495–502, ISBN 978-1-84800-971-4.
[6] Princeton University, WordNet, 2011. <http://wordnet.princeton.edu/> (retrieved 15.11.11).
[7] V.W. Soo, S.Y. Lin, S.Y. Yang, S.N. Lin, S.L. Cheng, A cooperative multi-agent platform for invention based on patent document analysis and ontology, Expert Systems with Applications 31 (2006) 766–775.
[8] Z. Grilliches, Patent statistics as economic indicators: a survey, Journal of Economic Literature (1990) 1661–1707.
[9] B. Yoon, Y. Park, A text-mining-based patent network: analytical tool for high-technology trend, The Journal of High Technology Management Research 15 (2004) 37–50.
[10] C.V. Trappey, A.J.C. Trappey, Wu, CY, Clustering patents using non-exhaustive overlaps, Journal of Systems Science and Systems Engineering 19 (2) (2010) 162–181.
[11] The. Granstrand, Economics and Management of Intellectual Property: Towards Intellectual Capitalism, Edward Elgar Publishing Limited, Cheltenham, UK, 1999.
[12] J.A.S. Almeida, A.A. Barbosa, C.C. Pais, S.J. Formosinho, Improving hierarchical cluster analysis: a new method with outlier detection and automatic clustering, Chemometrics and Intelligent Laboratory Systems 87 (2007) 208–217.
[13] J.R. Kettenring, A patent analysis of cluster analysis, Applied Stochastic Models in Business and Industry 25 (2009) 460–467.
[14] C.V. Trappey, H.-Y. Wu, F. Taghaboni-Dutta, A.J.C. Trappey, Using patent data for technology forecasting: China RFID patent analysis, Advanced Engineering Informatics 25 (2011) 53–64.
[15] S. Lee, B. Yoon, Y. Park, An approach to discovering new technology opportunities: keyword-based patent map approach, Technovation 29 (2009) 481–497.
[16] R. Kostoff, D. Toothman, H. Eberhart, J. Humenik, Text mining using database tomography and bibliometrics: a review, Technological Forecasting and Social Change 68 (2001) 223–252.
[17] T. Nasukawa, T. Nagano, Text analysis and knowledge mining system, IBM Systems Journal 40 (4) (2001) 967–984.
[18] S. Weiss, N. Indurkhya, T. Zhang, F. Damerau, Text Mining Predictive Methods for Analyzing Unstructured Information, Springer, Berlin, 2005.
[19] A.H. Tan, Text mining: the state of the art and the challenges, Proceedings of the PAKDD 696 (2011) 65–70.
[20] Y.H. Tseng, C.J. Lin, Y.I. Lin, Text mining techniques for patent analysis, Information Processing and Management 43 (2007) 1216–1247.
[21] Y.H. Tseng, Y.M. Wang, D.W. Juang, C.J. Lin, Text mining for patent map analysis, in: Proceedings, IACIS Pacific 2005 Conference, 2005/4/16-17, Taipei, Taiwan, 2005.

[22] H. Ernst, Patent applications and subsequent changes of performance: evidence from time-series cross-section analyses on the firm level, Research Policy (1995) 143–157.
[23] C. Choi, S. Kim, Y. Park, A patent-based cross impact analysis for quantitative estimation of technological impact: the case of information and communication technology, Technological Forecasting and Social Change 74 (2007) 1296–1314.
[24] M.E. Mogee, R.G. Kolar, International patent analysis as a tool for corporate technology analysis and planning, Technology Analysis Strategic Management 6 (4) (1994) 485–503.
[25] D.L. Rubin, N.H. Shah, F.N. Natalya, Biomedical ontologies: a functional perspective, Briefings in Bioinformatics 9 (9) (2007) 75–90.
[26] L. Wanner, R. Baeza-Yates, S. Brugmann, J. Codina, B. Diallo, E. Escorsa, Towards content-oriented patent document processing, World Patent Information 30 (1) (2008) 21–33.
[27] S. Taduri, G.T. Lau, K.H. Law, H. Yu, J.P. Kesan, Developing an Ontology for the US Patent System, in: Annual International Conference on Digital Government Research, University of Maryland, College Park, USA, June 12–15, 2011.
[28] J. Han, M. Kamber, J. Pei, Data Mining – Concepts and Techniques, Morgan Kaufmann Publishers, Elsevier, Waltham USA, 2011.
[29] R. Haupt, M. Kloyer, M. Lange, Patent indicators for the technology life cycle development', Research Policy 36 (2007) 387–398.
[30] H. Ernst, Use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry, Small Business Economics 9 (1997) 361–381.
[31] Wheeljet.com, IPDSS – Intellectual Property Defense Support System, 2012. <http://www.wheeljet.com.tw/edu/> (retrieved on 1.11.12).
[32] Y. Matsuo, M. Ishizuka, Keyword extraction from a single document using word co-occurrence statistical information, FLAIRS Associations for the Advance Artificial Intelligence (2003) 392–396.
[33] K.M. Hammouda, D.N. Matute, M.S. Kamel, CorePhrase: Keyphrase extraction for document clustering, Lecture Notes in Artificial Intelligence (LNAI), in: Proceedings Conference MLDM 3587, 2005, pp. 265–274.
[34] P.D. Turney, Learning algorithms for key phrase extraction, Information Retrieval 2 (2000) 303–336.
[35] A. Nenkova, L. Vanderwende, K. McKeown, A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization, in: Proceedings of SIGIR (2006) 573–580, August 6–11, 2006.
[36] S.E. Robertson, K. Sparck Jones, Relevance weighting of search terms, Journal of the American Society for Information Science 27 (3) (1976) 129–146.
[37] A.J.C. Trappey, C.V. Trappey, An R&D knowledge management method for patent document summarization, Industrial Management & Data Systems 108 (2) (2007) 245–257.
[38] Astra Tech Dental, Dental Implant, 2011. <http://www.likenaturalteeth.us/Main.aspx/Item/781594/navt/83633/navl/83642/nava/83644> (retrieved 08.08.11).
[39] Free Dental Implant Information, 2011. <http://www.free-dental-implants.com/dental-implant-components/> (retrieved 30.8.11).
[40] S. Pritzek, An Ontology for the RFID Domain, 2005. <http://www.competencies.at/Ontologies/RFIDOntology0903/RFIDOntology-Report.pdf> (retrieved on 15.12.11).