INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED15
27-30 JULY 2015, POLITECNICO DI MILANO, ITALY

# AN EXPLORATORY STUDY TO EVALUATE THE PRACTICAL APPLICATION OF PSS METHODS AND TOOLS BASED ON TEXT MINING

**Marques, Caio Augusto Nunes; Matsuno, Ivone Penque; Sinoara, Roberta Akemi; Rezende, Solange Oliveira; Rozenfeld, Henrique**
University of São Paulo, Brazil

## Abstract
Product Service System (PSS) is an approach in which the focus is on producing and delivering value in use instead of transferring physical goods ownership. In other to design PSS, some authors have been adapting methods and tools from traditional product design and others have been developing new specifically methods and tools. The description of those methods and tools is available in many sources of information, like thousands of publications in journals, and can only be analyzed based on Text Mining (TM) techniques. This paper reports an exploratory study whose purpose was to develop a procedure based on Text Mining techniques to support the identification of PSS design methods and tools which have been already applied in practical real cases. The research comprises the development of the first version of the procedure, employing TM techniques such as Named Entity Recognition, Association Rules and Bag of Expressions of the Domain. A Proof of Concept (PoC) was performed to verify whether the procedure is feasible and to identify future works possibilities. The PoC showed promising results and opened several possibilities to improve the procedure.

**Keywords**: Product-service systems (PSS), Text Mining, Information management

**Contact**:
Caio Augusto Nunes Marques
University of São Paulo
São Carlos School of Engineering
Brazil
caiomarques@usp.br

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 20th International Conference on Engineering Design (ICED15), Vol. nn: Title of Volume, Milan, Italy, 27.-30.07.2015

# 1 INTRODUCTION

Product-Service System (PSS) is an approach that changes the focus of the business from the development and selling of physical products to a system composed by products and services, which can comply with clients' specific needs (Manzini and Vezzoli, 2003; Shimomura et al., 2009). Tan (2010) defines PSS as a new approach that integrates activities and competences of an organization and intensifies the relationship between clients and partners in the value chain.

Environmental and economic benefits can result from PSS implementation (Mont, 2002; Unep, 2002; Manzini and Vezzoli, 2003; Tukker, 2004), e. g. when the PSS provider owns the product, it can have parts recovered through end-of-life strategies, such as recycling, remanufacturing and reconditioning (Barquet et al., 2013). Thereby, after recovered, it can be commercialized again in another use phase. The utilization of less resource to put the product back in the market results in both environmental and economic benefits (Mont, 2002; Baines et al., 2007, Tan et al, 2007). In addition, economical motivation to adopt PSS is related to the increase of profit by delivering services and not just selling products (Meier et al., 2010; Tan, 2010; Peruzzini et al., 2014).

New challenges are associated with the development and commercialization of PSS concerning the definition of new business models, which are essential to successfully implement a PSS (Mont et al., 2006; Tan, 2010; Meier and Bossla, 2012; Barquet et al., 2013; Reim et al., 2014). As the adoption of PSS oriented business models are yet limited, new methods for PSS development are expected to be proposed (Baines et al., 2007; Meier et al., 2010; Reim et al., 2014). However, some authors believe that the development of PSS is an expansion of product development scope (Manzini and Vezzoli, 2003; Tan, 2010), what would suggest that the traditional methods for product design and/or service design could be used for PSS development.

It seems that in the last years several researches have been carried out to develop design methods and tools but there weren't reported practical application cases, i.e. many of them were only theoretical development (McMahon, 2012). This situation might leverage to the low adoption and low impact of those methods and tools in industry, so that one should propose a study to better understand real industrial application of them. This was the issue discussed in the debate in the last Design Conference (Eppinger et al., 2014).

That debate and the publication of McMahon were the inspiration of this research. There is an open question whether indeed the methods and tools for PSS design can be differentiated from the product design ones. The proposals of new methods and tools are constantly growing and the sources of knowledge are spread out in publications and also in the web. Thus there might be an approach to continuously update the information about new methods, tools, practices and cases, in order to corroborate their real world practical application.

To face with these challenges there is an ongoing research project that intends to build an open access portal of knowledge about design practices, methods and tools. This portal is an evolution of a previous version that supported the collaborative work of a community of practice on product development (Rozenfeld et al., 2009). The actual version is only in Portuguese and can be visited at www.portaldeconhecimentos.org.br. The concept of this portal aims at offering more than the content of Wikipedia, since a visitor can download free methods and tools for applying in her/his organization. Additionally many knowledge elements (such as white papers, reports, case description, blogs, specialists, URL, discussions, etc.) are associated to them. The main goal is to open the possibility of a collaborative evaluation of those practices, methods and tools concerning its practical application in real cases by researches, consultants and practitioners as illustrated in Figure 1. Thus a practitioner could decide to use in her/his organization practices, methods and tools already validated.

One potential limitation is the steady creation of methods and tools whose description can be found in many sources of information like thousands of publications in journals and web (blogs, news, white papers, etc.). Manual monitoring of which new method or tool is appropriated to a real application for practitioners becomes impossible. So, the research question is: Is it possible to identify methods and tools for PSS design and assess its applicability at practical real cases in an automatic manner?

This big data can only be analysed based on Text Mining techniques. After a method or tool is automatically identified, it can be verified and be inserted in the portal. Afterwards the community can add information, as mentioned before, and its practical application can be described and evaluated (Figure 1). This paper reports an exploratory study with the purpose of develop a procedure based on

Text Mining techniques to support the identification of PSS design methods and tools which have been already applied in practical real cases. This is represented with the area marked at Figure 1.

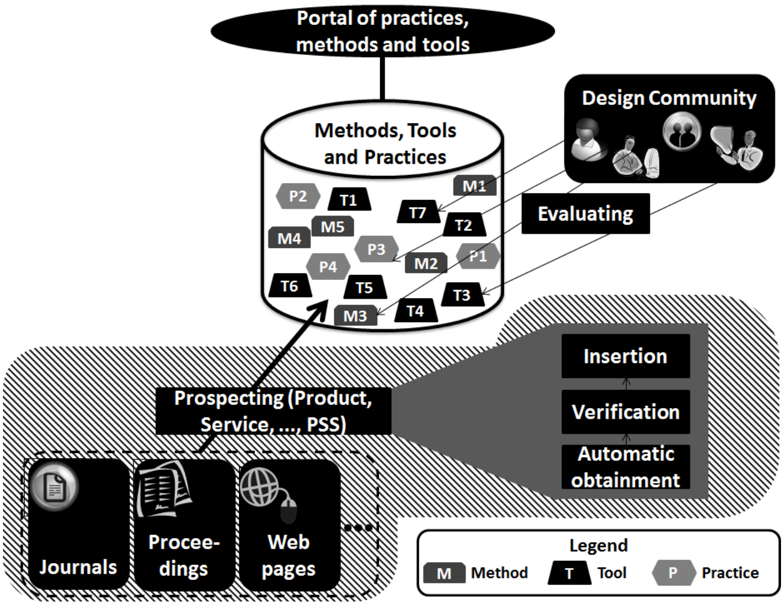


*Figure 1. Proposal of the automatic identification of methods and tools with practical application in the context of the portal of knowledge*

This proposal can be one of the elements for acquiring knowledge for the portal of knowledge described. The focus is on methods and tools for PSS development, but at the same time features about generic design methods and tools are tested in order to examine whether those traditional methods and tools are employed in PSS design. In this first version only indexed journals are considered.
The next section presents fundamental references about methods and tools for PSS design and about Text Mining. Section 3 describes the procedure adopted. A Proof of Concept is presented and discussed in section 4. The final remarks with the conclusions and description of the next steps of this research are presented in section 5.

## 2 THEORETICHAL FOUNDATION

### 2.1 Methods and tools to develop Product-Service Systems

The challenge of designing PSSs has driven part of the scientific community to put efforts to identify appropriated methods and tools to design PSSs. Some researchers have adopted the parsimony principle and have been studying if what are already developed to pure products can be successfully used in PSS design, even with adjustments. Morelli (2006), e.g., identified some methods and tools from others disciplines that he judged as useful in PSS design projects in which he took part.
Cavalieri and Pezzota (2012) affirm that usually companies adopt approaches based on pure product design, which are rigorously applied at the "tangible" part of the system, but still develop the "intangible" part in an intuitive way. Given the systemic profile of a PSS, adaptations should be made in order to meet the stakeholders needs (Beuren et al., 2013). Morelli (2006) and Baines et al. (2007) agree that there is still a lack of well-developed and standardized methods and tools which may provide the companies with a PSS' implementation guide. Meier et al. (2010) stated that companies need new methods and tools at both managerial and engineering levels to be capable of designing Industrial Product-Service Systems satisfactorily.
Many authors perceived this demand and have been researching new methods and tools specifically to design PSS. In summary, there are a range of methods and tools described in PSS literature, but still there is a lack of critical and in depth evaluation about its relevance (Baines et al., 2007). The keyword is application. There is already a set of methods and tools derived from product design theory or specifically developed for PSS design, however there is still a need for thorough and rigorous evaluation of their performance in real cases (Baines et al., 2007; Beuren et al., 2013). Baines et al. (2007) argues that the completeness of these methods and tools claims for evidences. Cavalieri and

Pezzota (2012) affirm that there are few utilization reports of these methods and tools in real cases. According to Morelli (2006), new case studies and more detailed applications may help to define a methodological approach to PSSs design.

### 2.2 Text Mining

A large amount of documents is stored daily in textual format in digital media. As the volume of information generated is greater than the human capacity can process, Text Mining (TM) techniques have become essential for supporting knowledge extraction (Gupta and Lehal, 2009; Aggarwal and Zhai, 2012). TM process is usually divided into five steps: problem identification, preprocessing, pattern extraction, post-processing and knowledge usage (Rezende, 2003; Feldman and Sanger, 2007; Han et al., 2011). As shown in Figure 2, this process is iterative. The internal steps may be executed several times, depending on the post-processing results.

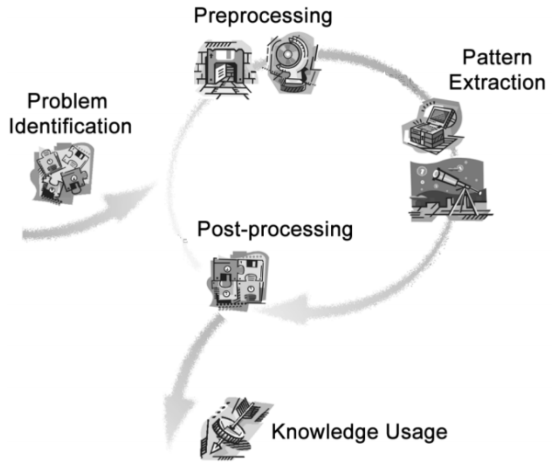


*Figure 2. Steps of the Text Mining process. Adapted from: Rezende (2003)*

The problem identification step will guide the whole process. In this step the objectives of the TM process are defined and the text collection is identified. In the preprocessing step, the text collection is structured in a format that is proper to knowledge extraction. The goal of the pattern extraction step is to identify patterns in the preprocessed data. There are several tasks that can be performed in this step. Classification, clustering and association rules are examples of those tasks. The knowledge extracted (i.e., the patterns) are evaluated in the post-processing step. If the knowledge matches the objectives of the TM process, it is prepared to the knowledge usage final step. If the knowledge does not match the TM process objectives, the cycle preprocessing-pattern extraction-post-processing is restarted.

This work proposes the instantiation of TM process in three different moments, in order to support the knowledge extraction about the application of methods and tools for designing products, services and Product-Service Systems. Each instantiation is performed with a different purpose and therefore, using different techniques. Details are presented in the next section.

## 3 PROPOSED PROCEDURE

Figure 3 shows the main contribution of this work, which is a procedure to identify and evaluate the applicability of methods and tools for designing products, services and PSS from a collection of documents. This is an initial proposal and might be refined as the research progresses. The procedure is an instantiation of the TM process (Figure 2) and is detailed in the following.

### 3.1 Problem Identification

This step is to prepare the database, selecting a set of documents dealing with methods and tools for product, service or PSS design. After defining this text collection, the preprocessing step is started.

### 3.2 Preprocessing

In the preprocessing step, three activities are performed: preparation of documents; extraction of domain terms; and generation of a bag-of-expressions-of-domain.

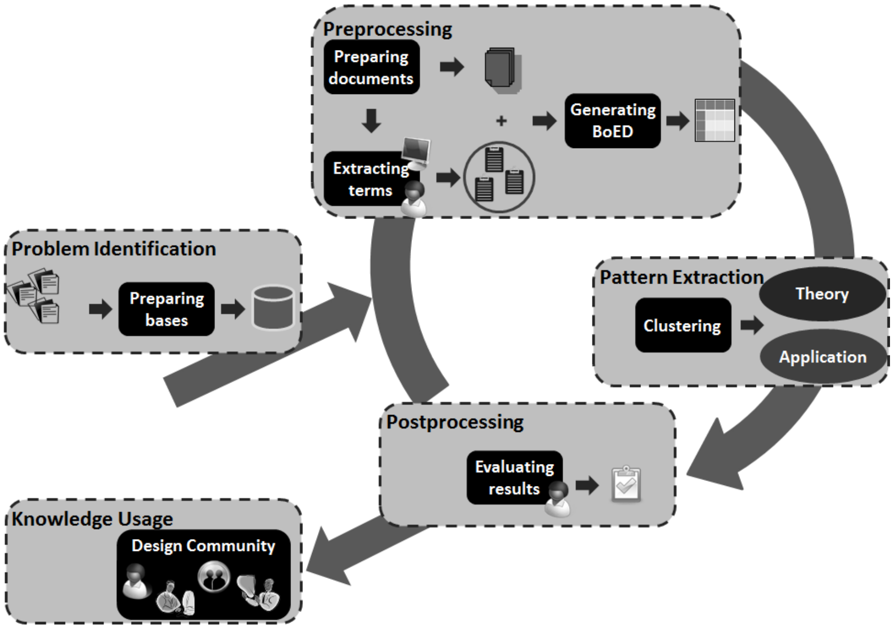


*Figure 3. Procedure proposed to identify and to evaluate the applicability of methods and tools for designing products, services and PSSs*

### 3.2.1 Preparing the documents

The text collection must be prepared to the next activities. The text collections are normally found in PDF or HTML files. In order to allow the next steps, each document must correspond to a single file in plain text format (TXT files).

### 3.2.2 Extracting terms

The term extraction activity aims to generate three lists of words (or phrases) of the domain: i) a list of names of methods and tools; ii) a list of words that the authors use to indicate that a method or tool has been applied; and iii) a list of words that authors use to present a theoretical work about particular method or tool. This activity can be done manually or with the support of TM techniques.

The manual identification consists of selecting a set of reference articles and perform read them to generate the three lists. The terms identification supported by TM can be performed using Named Entity Recognition (NER) methods or association rules generation. Those are two additional instantiation of TM process. These techniques can be used with the purpose of automatically extract the terms and domain expressions from the documents and generate a list to be evaluated by the domain expert.

NER involves processing a text to identify and classify the occurrences of words or expressions considered Named Entities (NE) (Grishman and Sundeheim, 1996). NEs include all entities that can be identified by a proper name, such as people, organizations, locations, brands and products, as well as temporal and numeric expressions. In the approach proposed in this paper, NER is used to extract the names of methods and tools for product or service development process, thus supporting the construction of the first list of terms (list of methods and tools).

The association rule mining was originated in market basket analysis (Agrawal et al., 1993), with the extraction of rules such as "customers who buy the products $x_1$ and $x_2$ also buy the product y". However the applicability of association rules is not restricted to the analysis of market baskets. In general, an association rule is an implication of the form A⇒B, in which A is the antecedent and B is the consequent. A and B are disjoint subsets of a set of items (or in our particular case, words) $I = \{i_1, i_2, \ldots, i_m\}$ and the association rule is generated from a set of transactions, where a transaction is a subset of I. The association rule A⇒B occurs in the transactions set with confidence conf if in 100*conf% of the transactions that occurs A also occurs B. The rule A⇒B has support sup if 100*sup% of transactions occurs A⇒B.

In this paper, the set of items consists of the words found in the text collection and the transactions can be created by two different approaches: i) sentences, i.e., each sentence of the articles form a transaction; or ii) sliding windows, where transactions are extracted by sliding windows in each document (for example, considering a window size of 10, each transaction is formed considering up to 10 words in sequence). Our proposal suggests using association rules for support the extraction of a set of terms that may indicate the words related to theoretical development or application of a method.

#### *3.2.3 Generating a Bag of Expressions of the Domain (BoED)*

In preprocessing step of TM process, textuals collections are usually represented as a bag-of-words (BoW). BoW is a document-term matrix, where each row represents a document, each column represents a word present in the collection of documents and each cell contains a measure of frequency of the word in the document. In our approach, we propose the construction of a bag-of-expressions-of-domain (BoED), whose difference from the traditional bag of words is that words are replaced by domain expressions.

A domain expression is formed by terms from the three lists of terms described in section 3.2.2 and formally presented below:

- $M = \{m_1(s_{11}, \cdots, s_{1i}), m_2(s_{21}, \cdots, s_{2i}), \cdots, m_k(s_{k1}, \cdots, s_{ki})\}$, set that contains the name of the selected k methods or tools for analysis and their respective synonyms ($s_i$);
- $A = \{a_1(s_{11}, \cdots, s_{1i}), a_{21}(s_{21}, \cdots, s_{2i}), \cdots, a_p(s_{p1}, \cdots, s_{pi})\}$, set that contains the p expressions indicating an application of a method or tool and their respective synonyms ($s_i$);
- $T = \{t_1(s_{11}, \cdots, s_{1i}), t_{21}(s_{21}, \cdots, s_{2i}), \cdots, t_q(s_{q1}, \cdots, s_{qi})\}$, set that contains q expressions that indicate the theoretical development of a method or tool and their respective synonyms ($s_i$).

Each expression of domain is composed by a term of M matched to a term of A or T. The expressions of domain are searched in each sentence of the documents. The frequency of each term in each document is calculated and the BoED is built for the text collection. In Figure 4 is shown the BoED for a collection of n documents represented from expressions generated by k methods, p application expressions and q expressions of theoretical development.

| | $m_1$+$a_1$ | ... | $m_1$+$a_p$ | ... | $m_k$+$a_1$ | ... | $m_k$+$a_p$ | $m_1$+$t_1$ | ... | $m_1$+$t_q$ | ... | $m_k$+$t_1$ | ... | $m_k$+$t_q$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_1$ | | | | | | | | | | | | | | |
| $d_2$ | | | | | | | | | | | | | | |
| $d_3$ | | | | | | | | | | | | | | |
| ⋮ | ⋮ | | ⋮ | | ⋮ | | ⋮ | ⋮ | | ⋮ | | ⋮ | | ⋮ |
| $d_n$ | | | | | | | | | | | | | | |

*Figure 4. Bag of Expressions of the Domain (BoED) Schema*

## 3.3 Pattern extraction

In patter extraction step, clustering methods are used when the objective of TM process is to organize the collection of documents and labelled documents are not available. The clustering of texts aims to organize a collection of text documents into clusters (groups), so that documents of the same cluster are highly similar and dissimilar to documents from other clusters. Our proposal is to use the BoED as input for clustering algorithms to organize the collection of articles into two clusters: documents relating to applications of methods and documents relating to theoretical development of methods.

## 3.4 Post-processing

The post-processing step refers to the evaluation of the knowledge (patterns) extracted in the previous step. That analysis will determine whether the pattern extraction step must be repeated or not. The TM internal cycle will end when the post-processing activities concludes that the TM goal (stated in the problem identification step) was achieved.

## 3.5 Knowledge Usage

This last step comprises the use of the knowledge acquired by the TM process. In this case, the knowledge of interest is the automatically identified methods and tools as well as their applicability mentioned in the scientific publications considered. This knowledge will be incorporated to the Portal of Knowledge in an ongoing process as described in section 1 and will be available to the design community.

# 4 PROOF OF CONCEPT

## 4.1 Problem Identification

This Proof of Concept (PoC) was performed with three sets of scientific articles' abstracts on methods and tools for PSS design but also containing generic design methods and tools as mentioned in section 1. The first set is about methods and tools for product design (set A), the second is related to product and service design (set B) and the third is for PSS design (set C).

A base of abstracts was built up to each set by searching (search strings presented in Table 1) on the fields Article Title, Abstract and Keywords at Scopus database (www.scopus.com). The searches were performed in October 29th 2014 and were limited to articles in English, constrained by Document Type (Conference Paper; Article; Article in Press) and Subject Area (Engineering; Computer Science; Business, Management and Accounting; Decision Sciences; Environmental Science). The results were exported to comma-separated values (CSV) files and the duplications were removed at each base. The bases are characterized in Table 1.

*Table 1. Abstracts bases*

| Base | Search String | Abstracts found | Abstracts used |
|---|---|---|---|
| A | TITLE-ABS-KEY (("product development" OR "product design" OR "systems engineering") AND (method OR tool )) | 38.947 | 38.494 |
| B | TITLE-ABS-KEY (("product development" OR "product design" OR "systems engineering") AND (service) AND (method OR tool )) | 2.839 | 2.819 |
| C | TITLE-ABS-KEY (("product service system" OR "industrial product service system") AND (method OR tool )) | 234 | 232 |

## 4.2 Preprocessing

Initially the abstracts in each base were converted to TXT files. Secondly, the construction of the three lists of terms (methods and tools; application terms; theory terms) was performed by using two approaches: manual and supported by TM.

It was adopted the following flow to the manual identification of the terms: selection of a set of most cited articles; reading of these articles in order to identify methods and tools, terms used to indicate application and terms used to express theoretical development; search for variations of the terms, such as synonymous and verbal tenses; building of the three lists bellow.

- M={Analytic Hierarchy Process (AHP), Brainstorming, Computer Aided Design (CAD), Conjoint Analysis, Delphi, Design for Assembly (DFA), Design for Disassembly (DFD), Design Structure Matrix (DSM), Eco-costs/Value Ratio Model (EVR Model), Failure Mode and Effects Analysis (FMEA), Focus Group, Kansei Engineering, Life Cycle Assessment (LCA), Product-Service Blueprint, Quality Function Deployment (QFD), Technology Roadmap (TRM), Theory of Inventive Problem Solving (TRIZ)};
- A={use (uses, using, used, usage), apply (applies, applying, applied, application), validate (validates, validating, validated), case study (case research, action research, cases, real case, practical case)};
- T={develop (develops, developing, developed), propose (proposes, proposing, proposed), introduce (introduces, introducing, introduced), suggest (suggests, suggesting, suggested), provide (provides, providing, provided)}.

In this Proof of Concept NER was carried out with the OpenCalais Service (www.opencalais.com). This tool uses natural language processing and other methods of machine learning to find named entities, facts and events in documents. Entities recognized at the categories Industry Term or Technology were analyzed by a domain expert.

Through this approach, 11 methods were identified (DSM, LCA, QFD, AHP, Focus Group, Conjoint Analysis, Kansei Engineering, TRM, Brainstorming, CAD, FMEA) from the total of the 17 methods and tools that have been raised manually. Interestingly, the technique was more efficient in the base A, where there were abstracts dealing with methods and tools for product design, that is, it was easier to identify more traditional terms of the area.

To find the possible words that represent terms that may indicate the theoretical or practical use of a method or tool were generated association rules with Apriori algorithm (Agrawal et al., 1993). From the text collections were generated transactions through sentences and sliding windows. Only the transactions that contained the words "method", "methods", "tool" or "tools" were used to generate the association rules. Association rules whose consequent was one of those words were selected for review. We tried to identify words representing terms of application or theoretical development in the antecedent of those rules.

It was observed that the rules have adequate list of issues addressed in papers, identifying relations as: {product}⇒{tools}, {design}⇒{method}, {service}⇒{tools}, {pss}⇒{method} and others. The only rule identified indicating theoretical use was {proposed}⇒{method}.

After the term extraction, a BoED was built for each base. For this construction, we used the lists manually generated.

### 4.3 Pattern Extraction and Post-processing

The patterns were extracted using the traditional K-means algorithm (Jain and Dubes, 1988; Wu and Kumar, 2009). We used the K-means available in Torch tool (sites.labic.icmc.usp.br/torch). The algorithm was run on the BoED representation to generate two clusters. Our goal was organize the documents into two clusters: application reports and theoretical development reports. The same processing was applied for the bases A, B and C.

The analysis of the clusters extracted in addition to the clusters descriptors (the most frequent terms of each cluster) showed limitations due to some characteristics of our text collections and the text representation adopted (BoED), but is promising. As we are working with article abstracts in this PoC, we identify a low frequency of domain expressions due to the following facts: i) as abstracts are short texts, they may not have enough information for the traditional algorithms organize the documents in applications and theoretical developments; ii) in abstracts, the main method that the scientific article deals with (the method that was applied or proposed) will not be cited with high frequency to produce a high number of domain expressions; iii) both main methods and secondary methods (those used in a comparison basis, for example) will normally be cited in almost the same frequency in abstracts; iv) the same facts stated previously (itens ii and iii) apply to terms that indicate applications and terms that indicate theoretical development.

To illustrate those facts, in the following we present a brief analysis of the BoED generated in this PoC. Table 2 presents a summarization of the BoED generated for each base considering only Life Cycle Assessment (LCA) and Technology Roadmap (TRM) methods.

*Table 2. Summarization of BoED for bases A, B and C, related to LCA and TRM methods*

| **Base** | **Method** | **Distinct Expressions** | | **Domain Expressions Frequency** | | **Documents with Domain Expressions** | |
|---|---|---|---|---|---|---|---|
| | | Application Terms | Theory Terms | Application Terms | Theory Terms | Application Terms | Theory Terms |
| A | LCA | 7 | 5 | 337 | 242 | 178 | 146 |
| | TRM | 4 | 4 | 25 | 29 | 20 | 21 |
| B | LCA | 5 | 5 | 42 | 35 | 22 | 26 |
| | TRM | 0 | 2 | 0 | 3 | 0 | 2 |
| C | LCA | 2 | 2 | 2 | 2 | 1 | 3 |
| | TRM | 0 | 0 | 0 | 0 | 0 | 0 |

There is no expressive difference between the frequencies of domain expressions of application and the theoretical development ones. In the case of LCA in base A, the application expressions appeared 39,3% more than theory expressions what should mean that LCA is a frequently applied method according to the literature of product design. To the bases B and C, which are smaller sets dealing with areas newer than product design, expressions related to the application of methods appear about the same frequency than expressions related to theoretical development. To TRM in base A, the theory expressions surpass the application expressions in 16,0%, what should suggest that TRM is more frequently approached by researchers in theoretical than practical situations. At base B just theoretical reports were found and to base C there were any expressions found.

## 5 CONCLUSIONS

This exploratory research has reached its main purpose of developing an initial procedure based on Text Mining techniques to automatically identify methods and tools for PSS design which have been already applied in practical real cases. Additionally it was developed the BoED, which is a new approach that uses selected domain expressions instead of just words as in the traditional BoW. It was possible to recognize some methods and tools in an automatic manner, to identify some relations and to verify, yet inconclusively, the applicability of some of them for PSS design.

Besides its limitations, the results indicate the feasibility of the procedure. Some methods and tools were found using a generic NER tool, which recognized the more traditional and well-known methods. On the other hand, this technique did not perform well on the task of identifying methods and tools less known. Thus, there is a need to adapt the NER technique to perform better in this domain. It is understood that it is also possible to improve the use of Association Rules technique in order to capture more terms which express application or theoretical development of a method/tool. Some issues should be considered as identifying only rules in which the method or tool is the predicate. It is important to avoid cases as "QFD applies matrixes…" which would be captured as an application expression instead being a theoretical development one. Thus, a deeper investigation of the techniques we have adopted (NER, Association Rules and BoED) is being performed.

As stated in section 2.2, TM is an iterative process. In this PoC it was performed the first preprocessing-pattern extraction-post-processing cycle. In the second cycle, the following changes should be considered: i) use of the full article to enrich the data available for automated analysis; ii) incorporate new terms to the lists of methods and tools, application terms and theoretical development terms to complement the BoED; iii) use the BoED as privileged information, being a complement to the textual representation based on the traditional BoW (Marcacini and Rezende, 2013; Sinoara et al., 2014).

For future work, it is expected the elaboration of a procedure to construct the lists of application and theoretical development terms in a more systematic manner. It is also expected the upgrade of the domain expressions base, incorporating terms utilized by the authors but that are still out of the BoED. Additionally, it is intended to explore the use of other text mining techniques, such as the bag-of-related-words text representation (Rossi and Rezende, 2011) and other methods for clustering documents, such as Latent Dirichlet Allocation (Blei et al., 2003).

As an ongoing research, there are still open issues. Additional tests, as well as improvements, are being made in order to turn the initial proposal a more robust procedure. Then it will be possible to: i) assess and to conclude with more certainty about the relation applicability/theoretical development of the methods and tools for PSS design; ii) understand if the origin of each one is at product, service or PSS theory and; iii) provide the Portal of Knowledge with this information.

## ACKNOWLEDGMENTS

The authors would like to thank Capes (Higher Education Personnel Training Coordination) and grant #2014/08996-0 and #2013/14757-6, São Paulo Research Foundation (FAPESP) for financial support.

## REFERENCES

Agrawal, R., Imielinsk, T. and Swami, A. N. (1993). Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia (Eds.), Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216.

Aggarwal, C. C. and Zhai, C., editors (2012). Mining Text Data. Springer.

Baines, T. S., Lightfoot, H. W., Evans, S., Neely, A., Greenough, R., Peppard, J., Roy, R., Shehab, E., Branganza, A., Tiwari, A., Alcock, J. R., Angus, J. P., Bastl, M., Cousens, A., Irving, P., Johnson, M., Kingston, J., Lockett, H., Martinez, V., Michele, P., Tranfield, D., Walton, I. M., Wilson, H (2007). State-of-the-art in product-service systems. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, Vol. 221, No. 10, pp.1543–1552.

Barquet, A.P.B., Amigo, C. R., Oliveira, M. G., Cunha, V. P. and Rozenfeld, H. (2013). Employing the business model concept to support the adoption of product–service systems (PSS). Industrial Marketing Management, Vol. 42, No. 5, pp.693–704.

Beuren, F.H., Gomes Ferreira, M.G. and Cauchick Miguel, P. A. (2013). Product-service systems: a literature review on integrated products and services. Journal of Cleaner Production, Vol. 47, pp.222–231.

Blei, D. M., Ng A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation. Ed. John Lafferty. Machine Learn. Res.3, pp.993-1022.

Cavalieri, S. and Pezzotta, G. (2012). Product–Service Systems Engineering: State of the art and research challenges. Computers in Industry, Vol. 63, No. 4, pp.278–288.

Eppinger, S., Torlind, P., Eckert, C. and Howard, T. (2014). This house believes that in its present state design research spends too much effort on explaining what is done by designers instead of developing new ways to design. International Design Conference-DESIGN 2014, www.designconference.org accessed on 14 October 2014.

Feldman, R. and Sanger, J. (2007). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Grishman, R. and Sundeheim, B. (1996). Message understanding conference-6: A brief history. The 16th International Conference on Computational Linguistics, Proceedings COLING-96, volume 1, pp. 466-471

Gupta, V., and Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, Vol. 1., No. 1, pp. 60-76.

Han, J., Kamber, M. and Pei, J. (2011). Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems Elsevier.

Jain, A. K. and Dubes, R. C.(1988). Algorithms for Clustering Data, Prentice Hall.

Kozareva, Z. (2006). Bootstrapping named entity recognition with automatically generated gazetteer lists. In: Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, pp. 15–21.

Manzini, E. and Vezzoli, C. (2003) A strategic design approach to develop sustainable product service systems: examples taken from the "environmentally friendly innovation" Italian prize. Journal of Cleaner Production, Vol. 11, No. 8, pp. 851–857.

Marcacini, R. M. and Rezende, S. O.(2013). Incremental hierarchical text clustering with privileged information. In Proceedings of the 2013 ACM Symposium on Document Engineering, pp 231-232.

McMahon, C.A. (2012). Reflections on diversity in design. Journal of Engineering Design, Vol. 23, pp.563–576.

Meier, H., Roy, R. and Seliger, G. (2010). Industrial Product-Service Systems—IPS2. CIRP Annals - Manufacturing Technology, Vol. 59, No. 2, pp.607–627.

Meier, H., Bossla, M. (2012) Design and Engineering of Dynamic Business Models for Industrial Product-Service Systems. 4th CIRP Conference on Industrial Product Service Systems, Tokyo, p.179–184.

Mont, O. (2002). Clarifying the concept of product–service system. Journal of Cleaner Production, Vol. 10, No. 3, pp.237–245.

Mont, O., Dalhammar, C. and Jacobsson, N. (2006). A new business model for baby prams based on leasing and product remanufacturing. Journal of Cleaner Production, Vol. 14, No. 17, pp.1509–1518.

Morelli, N. (2006). Developing new product service systems (PSS): methodologies and operational tools. Journal of Cleaner Production, Vol. 14, No. 17, pp.1495–1501.

Peruzzini, M., Marilungo, E. and Germani, M. (2014). Sustainable Product-Service Design in Manufacturing Industry. International Design Conference - DESIGN 2014. Dubrovnik, pp. 955–964.

Reim, W., Parid, V. and Örtqvist, D. (2014) Product Service Systems (PSS) business models and tactics e a systematic literature review. Journal of Cleaner Production, In Press.

Rezende, S. O. (2003). Sistemas Inteligentes: fundamentos e aplicações. Editora Manole.

Rossi, R. G. and Rezende, S. O. (2011). Building a Topic Hierarchy Using the Bag-of-related-words Representation. Proceedings of the 11th ACM Symposium on Document Engineering , pp.195-204.

Rozenfeld, H., Amaral, C. S.T., Costa, J.M.H., Jubileu, A.P. (2009). Knowledge-oriented process portal with BPM approach to leverage NPD management. Knowledge and Process Management, Vol. 16, pp. 134-145.

Shimomura, Y., Hara, T. and Arai, T. (2009) A Unified Representation Scheme for Effective PSS Development, CIRP Annals – Manufacturing Technology, Vol. 58, No. 1, pp. 379-382.

Sinoara, R. A., Sundermann, C. V., Marcacini, R. M., Domingues, M. A. and Rezende, S. O. (2014). Named entities as privileged information for hierarchical text clustering. In Proceedings of the 18th International Database Engineering & Applications Symposium, IDEAS '14, ACM, pp. 57-66.

Tan, A.R. (2010). Service-oriented product development strategies. PhD Thesis. Technical University of Denmark.

Tan, A.R., Mcaloone, T.C. and Gall, C. (2007). Product/Service-System Development – An explorative case study in a manufacturing company. International Conference of Engineering Design-ICED07. Paris, pp. 1–12.

Tukker, A. (2004). Eight types of product-service system: Eight ways for sustainability? Experiences from SUSPRONET. Business Strategy and the Environment, Vol. 13, pp.246–260.

United Nations Environment Programme – UNEP, (2002). Product-Service Systems and Sustainability: Opportunities for sustainable solutions, www.unep.org accessed on 21 November 2014.

Wu, X. and Kumar, V.(2009). The Top Ten Algorithms in Data Mining, Chapman & Hall/CRC.