



**HAL**  
open science

# A Recursive K-means towards trade-off between PLM Competences, Positions, and Offers

Mourad Messaadia, Samir Ouchani, Anne Louis

► **To cite this version:**

Mourad Messaadia, Samir Ouchani, Anne Louis. A Recursive K-means towards trade-off between PLM Competences, Positions, and Offers. The 15th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2018) / The 13th International Conference on Future Networks and Communications (FNC-2018), Aug 2018, Gran Canaria, Spain. pp.297 - 302, 10.1016/j.procs.2018.07.174 . hal-04415778

**HAL Id: hal-04415778**

**<https://hal.science/hal-04415778>**

Submitted on 24 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain



The 13th International Conference on Future Networks and Communications  
(FNC-2018)

## A Recursive K-means towards trade-off between PLM Competences, Positions, and Offers

Mourad Messaadia<sup>a</sup>, Samir Ouchani<sup>b,\*</sup>, Anne Louis<sup>a</sup>

<sup>a</sup>LINEACT, Laboratoire d'Innovation Numérique  
École d'Ingénieur; CESI EI, Rouen, France

<sup>b</sup>LINEACT, Laboratoire d'Innovation Numérique  
École d'Ingénieur en Informatique, CESI eXia, Aix-en-Provence, France

### Abstract

Last decade recognizes a high job demand, more specialized trainings with very oriented jobs offers. This situation makes hiring and recruitments officers in the difficulty to select and find easily the appropriate candidate as well for candidates to choose the best practices and trainings to find later a respectable position. This work aims to help all actors in the job sector by modeling the Product Lifecycle Management (PLM) competencies and analyzing the demands especially in industry 4.0. First, the enterprises needs, in terms of skills, are identified through various job offers distributed on online media. Job offers are structured according to profile, geolocation and required competencies, etc. Then, the analysis is based on information retrieval and text mining through a statistical measure used to evaluate how important a competence to a job offer in a given collection. This contribution applies the Term Frequency Inverse Document Frequency (TF-IDF) to determine what skills in a corpus of job offers is the most requested in PLM jobs. This contribution addresses more than 1300 job offers, written in French, and posted in France. The offers cover more than 388 K words, from which 20 types of PLM job titles and 106 terms are related to the job competencies. The obtained results allow us to identify the most requested jobs, skills and classifying jobs and competencies for a better guidance of PLM job actors.

© 2018 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>)

Peer-review under responsibility of the scientific committee of the 13th International Conference on Future Networks and Communications, FNC-2018 and the 15th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2018.

**Keywords:** PLM Competencies, PLM job offers, TF-IDF, Recursive M-Means, Clustering;

### 1. Introduction

*Motivation.* The industry sector, which is part of a globalized economy with high competitiveness requirements, is currently undergoing deep structural changes in its business model that is related principally to scientific, economic, demographic and societal factors. This change will, in the short or medium term requires new skills able to understand and to face the jobs searching and hiring challenges.

\* Corresponding author. Tel.: +33-442-976-324; fax: +33-442-971-440.

E-mail address: [souchani@cesi.fr](mailto:souchani@cesi.fr)

The management of skills, through short and long term requirements, instructs a relevant modeling of competencies [3] that took many forms and preoccupies in both academia as well industry, especially it is a hot topic in scientific research era. Further in industry, engineering competences are integrated as a primordial point of view for an enterprise modeling and strategies [2].

*Problems.* The notion of competence is multidimensional, and its meaning depends on the contexts and sought of use, especially when a competence covers several disciplinary fields. Ley Tobias and Albert Dietrich [6] identified a competence as a set of personal peculiarity (knowledge, know-how, know-how-to be) that are relatively stable across different situations. Furthermore, it has been seen as the integration and coordination of a set of aptitudes [8], which can themselves be combined or recombined [4]. The notion of aptitude refers here to the ability of an individual (or an enterprise) to mobilize and exploit resources.

*Contributions.* Our intention is to help all PLM market actors by modeling the Product Lifecycle Management (PLM) competencies and analyzing the demands especially in industry 4.0. First, the enterprises needs, in terms of skills, are identified through a data base of job offers distributed on on-line media. Then we propose to format data for a specific structure according to technological norms. Further, the analysis is based on information retrieval and text mining through a statistical measure used to evaluate how important a competence to a job offer in a given collection. This contribution applies the Term Frequency Inverse Document Frequency (TF-IDF) to determine what skills in a corpus of job offers is the most requested in PLM jobs. Also it adopt a recursive K-means clustering to measure the tradeoff between job offers, candidates profiles and the needed competences. The approach is successfully applied on more than 1300 job offers, written in French, and posted in France. The offers cover more than 388 K words, from which 20 types of PLM job titles and 106 terms are related to the job competencies. The obtained results allow us to identify the most requested jobs, skills and classifying jobs and competencies for a better guidance of PLM job actors.

*Related Work.* Few contributions are dedicated to the competence modeling and analysis, for example Gilbert Paquette [7] proposed an ontology that describes the concept of a competence in order to create an online e-learning systems. The ontology provides the relationship between generic skills applied to knowledge at a certain level of abstraction. As context, it uses prerequisite, actual, and target competences. Further, other typologies are proposed [9] to describe technical, management, and personal abilities knowledge. Girodon Julien [5] addressed competencies and skills modeling by proposing a knowledge reuse organizational model that covers the organizational dimension of knowledge and competence in relationship with professional situations (concept of activity). Also he proposed a unified enterprise modeling language that helps to describe an organization activities, the role of resources through the qualification and capabilities aspects. The competency resource aspect individual is to model the internal aspects of competence such (to-know-how, to-know and to-know-whom). UECML (Unified Enterprise Competence Modeling Language) covers the aspect of individual and collective competencies with promoting their assessments.

## 2. PLM Competencies analysis

Industry 4.0 plays a key role on integrating tasks between different departments of many institutions by implying more collaborators. For a better guidance to handle different tasks, we propose an automatic approach that analyses the PLM market needs, categorizes the proposed jobs offers and the candidate profiles in the purpose to produce the best orientations for an appropriate selection that matches the best candidate to the open position.

The main steps for a best guidance described as follows.

- Jobs offers are textually to enclose: title, position description, etc in different forms initially: web, document, email, etc.
- Unified representation restructures the job offers in a unique formatting which produce a central access and a unified structure respecting the PLM norms.
- Job competencies step extracts the main keywords of the existing offers formatted earlier for a field and/or its classes.
- Bases on the job competencies, the dependencies measure how much a competence is dependent to an offer and to a given knowledge.
- The results groups knowledge needed for a set of position and helps recruiters to orient better candidates.

### 2.1. PLM Competencies Model

To model the PLM competencies and their system management, we align within the existing modeling standards, especially UML.

The competence class is modeled to be composed of three classes:

- *Knowledge* describes all what has been learned through a training including: theoretical knowledge, methodologies, and tools, etc. In the PLM context, it represents the product life-cycle, work-flows, data management, and standard modeling languages, etc.
- *KnowExperience* is related to the experience and the operational capabilities (techniques, hands abilities, expertise, etc). it means I know how to do it, I can do it and I can teach someone else how to do it. In PLM, it represents CAD, PDM, PLM tools, and programming, etc.
- *KnowBehave* represents the individual characteristics that allow adopting a particular behavior in a given situation. They combine relational skills (associated with cooperative behaviors), cognitive skills (intellectual operations of problem solving) and behaviors (personal qualities). In the PLM context, it includes at least communication, autonomy, rigor, and teamwork.

Each candidate, targets a job offer, follows training composed of one or more modules, and has a status describing his profile and study level.

- *JobOffer* describes a finite set of tasks by covering: a job title that details tasks in few words, Geolocation, duration, salary and the activity area. The field *JobDescription* defines textually the job's boundaries, the needed competencies, role, function with specific tasks, etc. Further it presents the enterprise details if exist.
- *Training* class describes the teaching and learning followed by a candidate (person) to achieve a certain competence. *Training* class is defined by a title, duration, contents, and composed by one or more modules. A *module* class describes the field, the content, and the duration needed to train a candidate to have a competence.
- *Status* class describes the profile of a candidate according to his competencies and training. This class reflects the curriculum vitae orientation, where it contains his professional experience, academic level, etc.

### 2.2. Job Offers Representation

The Job offer of a vacant position described textually and posted on recruitment websites (APEC, LinkedIn, Indeed, Monster) in order to attract the appropriate candidate needs to be reformatted and restructured for analysis. We reformat a job offer according two types of fields: non-structured field that contains the job offer description which is a pure text, and a structured field that is represented by well structured data such as experience and postal code.

- Structured fields might includes title, location, type of contract, salary, status, company description, experience, sector, and job category, the enterprise field, year of job offer apparition, the position of the candidate, the occupation and the job location.
- non-Structured fields is written in a natural language and it includes the job offer description and sometimes the map of a location, for example.

### 2.3. Competencies Extraction

To extract the competencies from a set of formatted documents, we follow a three steps based approach and it is described as follows.

1. *Tokenisation* step breaks up the text of a document into individual pieces, we called tokens, such as words, keywords, symbols, etc. At this step, the special characters (punctuations, accents,...) are excluded.
2. *Filtering* step removes tokens that are not conveying any information with respect to a pre-known dictionary , such as determinants, or if they belong to a beforehand settled list.

3. *Processing* step reduces the word's list by excluding tokens having the same or a close meaning (e.g. competence/competencies) with respect to the synonym tree of the dictionary.

#### 2.4. Competencies Evaluation

We express a job offer as a set of terms 'bag of words' denoted by

$$\mathbf{T} = \{\dots, t, \dots : t \text{ is a word}\} \quad (1)$$

For a data set of  $n$  documents with  $m$  distinct terms, we consider the matrix  $W$  of size  $n \times m$  that shows how much important a term  $t$  is to a document  $d \in \mathbf{D}$ . To evaluate  $W$ , we adopt the term frequency–inverse document frequency approach. It is the product of term frequency  $\text{tf}$  and inverse document frequency  $\text{idf}$ .

For a document  $d_i$  and a term  $t_j$ ,  $\text{tf}_{j,i}$  is adjusted to be the frequency of  $t_j$  (denoted by  $f_{j,i}$ ) with respect to the term-size of  $d_i$  denoted by  $\|d_i\|$  (i.e. number of terms/words) such that  $\text{tf}_{j,i} = \frac{f_{j,i}}{\|d_i\|}$ . To measure of how much information a term  $t_j$  provides  $\text{idf}_j$  is the logarithmically scaled inverse fraction of all documents that contain the term  $t_j$  and given by  $\text{idf}_j = \log \frac{|\mathbf{D}|}{\|d \in \mathbf{D} : t_j \in d\|}$ . Equation 2 evaluates  $W$  as follows.

$$w_{i,j} = \frac{f_{j,i}}{\|d_i\|} \times \log \frac{|\mathbf{D}|}{\|d \in \mathbf{D} : t_j \in d\|} \quad (2)$$

#### 2.5. Analysis and Interpretation

For each  $d \in \mathbf{D}$ , we have  $w_d$  is a  $|\mathbf{T}|$ -dimensional real vector and we aim to partition the  $|\mathbf{D}|$  documents with respect to the frequencies evaluated for terms  $\mathbf{T}$  representing the competences. Lets consider  $\mathbf{P}$  the set of possible partitions where  $|\mathbf{P}| < |\mathbf{D}|$ . For a clear partitioning, we consider to minimize initially the variance of  $\mathbf{P}$  expressed by Equation 3.

$$\underset{\mathbf{P}}{\operatorname{argmin}} \sum_{i=1}^{|\mathbf{P}|} |p_i| \times \operatorname{Var}(p_i) = \underset{\mathbf{P}}{\operatorname{argmin}} \sum_{i=1}^{|\mathbf{P}|} \frac{1}{2 \times |p_i|} \sum_{d_i, d \in p_i} \|w_{d_i}, w_d\| \quad (3)$$

To find the best separated partitions, Algorithm 1 develops a recursive based K-means algorithm (RBKM) [1] that controls the trade-off between the number of distance computations and the quality of the obtained partitions.

---

**Algorithm 1** RBKM: Recursive Based K-Means Algorithm.

---

**Require:**  $W$  is a  $|\mathbf{D}| \times |\mathbf{T}|$  -matrix of reals and  $|\mathbf{P}|$  is an integer.

**Ensure:**  $\mathbf{P}$  is a set of partitions.

Initialize all  $p_k \subset \mathbf{P}$ ;

Compute the centroids  $c_k$  of  $p_k$ ;

**for all**  $p_k \subset \mathbf{P} \wedge d \in p_k$  **do**

**for all**  $p_{k'} \subset \mathbf{P} \wedge p_{k'} \neq p_k$  **do**

**while**  $\|w_d, c_k\| > \|w_d, c_{k'}\|$  **do**

$p_k = p_k \setminus \{d\}$ ;

$p_{k'} = p_{k'} \cup \{d\}$ ;

**end while**

**end for**

**end for**

---

### 3. Experimental Results

Based on (1300) job offers collected from APEC database, we extract PLM related jobs from the offers titles. The result finds 20 categories as shown in Fig. 1. It is well clear that 8 PLM jobs account for 80% of the required job: PLM developer, PLM project manager, PLM consultant, Design and R&D engineer, industrialization engineer, PLM functional consultant, and PLM solution architect. This result gives the initial statistic regarding the PLM job market for the job seekers. Further, schools and companies looks toward how to perform new training and courses to fulfill the PLM market.

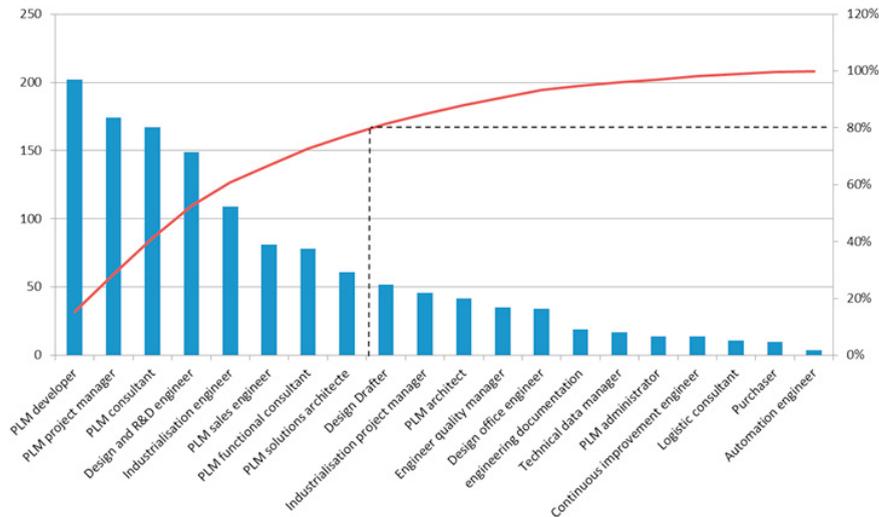


Fig. 1. Types of PLM job offer.

Further, the goal of the approach looks to identify the main competencies (Knowledge, Know experience and know behave) required by recruiters. First, it measures TF-IDF of a specific job offer and Fig. 2 gives an overview of the frequently required competencies for PLM functional consultant (as example).

For analysis, the proposed approach partitions the competencies regarding the PLM job offers with respect to the extracted competences. The recursive K-means groups height partitions enclosing the job offers presented as follows: 167, 12, 164, 16, 1, 33, 9, and 631.

### 4. Conclusion

To help job actors in Industry 4.0, we developed a PLM modeling and analysis framework. The latter model describes enterprises needs where job offers are structured according to profile, geolocation and required competencies, etc. Further, the analysis of the model relies to the information retrieval and text mining to evaluate how important a competence to a job offer in a given collection. The framework enhances TF-IDF to determine what skills in a corpus of job offers. The effectiveness of the framework has been shown on more than 1300 job offer. The obtained results allowed us to identify the most requested jobs, skills and classifying jobs and competencies for a better guidance of PLM job actors.

We would like to extend our work in different directions. First, we will include the level of experience needed in order to guide the level of deepening of knowledge during training. Then, we will evaluate distances between different types of PLM jobs across skill types. Further, experiment the framework with different sectors and datasets.

### Acknowledgment

Acknowledgement is made to APEC for the job offers database.

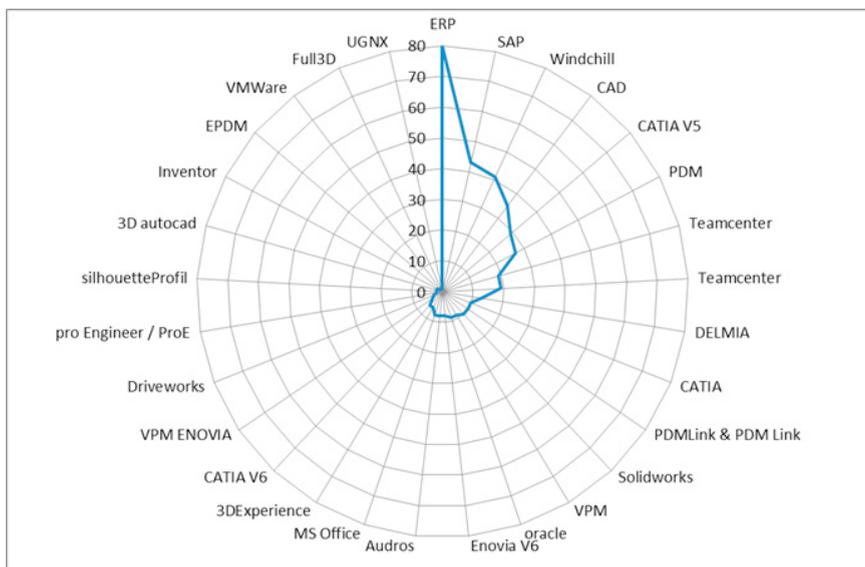


Fig. 2. Knowledge competency for PLM project head job

## References

- [1] andAritz Pérez andJosé Antonio Lozano, M. C., 2017. An Efficient Approximation to the K-means Clustering for Massive Data. *Knowl.-Based Syst.* 117, 56–69.
- [2] Boucher, X., Bonjour, E., Grabot, B., 2007. Formalisation and use of competencies for industrial performance optimisation: A survey. *Computers in industry* 58 (2), 98–117.
- [3] Boumane, A., Talbi, A., Tahon, C., Bouami, D., 2006. Contribution à la modélisation de la compétence.
- [4] Denford, J. S., 2013. Building knowledge: developing a knowledge-based dynamic capabilities typology. *Journal of Knowledge Management* 17 (2), 175–194.
- [5] Girodon, J., 2015. Proposition d'une approche d'amélioration des performances des organisations par le management opérationnel de leurs connaissances et compétences. Ph.D. thesis, Université de Lorraine.
- [6] Ley, T., Albert, D., 2003. Identifying employee competencies in dynamic work domains: methodological considerations and a case study. *J. UCS* 9 (12), 1500–1518.
- [7] Paquette, G., 2007. An ontology and a software framework for competency modeling and management. *Journal of Educational Technology & Society* 10 (3).
- [8] Torckeli, M., Tuominen, M., 2002. The contribution of technology selection to core competencies. *International journal of production economics* 77 (3), 271–284.
- [9] Vila, C., Ugarte, D., Ríos, J., Abellán, J., 2017. Project-based collaborative engineering learning to develop industry 4.0 skills within a plm framework. *Procedia Manufacturing* 13, 1269–1276.