Systematic community of Practice activities evaluation through Natural Language Processing: application to research projects

Virginie Goepp, Nada Matta, Emmanuel Caillaud, Françoise Feugeas

Systematic community of Practice activities evaluation through Natural Language Processing: application to research projects

Virginie Goepp[1], Nada Matta[2], Emmanuel Caillaud[3], Françoise Feugeas[1]
[1]INSA de Strasbourg - ICube
24, Bld de la Victoire - 67084 Strasbourg Cedex France
[2]UTT – TechCICO
12 Rue Marie Curie CS.42060 - 10004 Troyes Cedex France
[3]Unistra, ICube
3, rue de l'université - 67084 Strasbourg Cedex France
e-mails: {virginie.goepp , francoise.feugeas} @insa-strasbourg.fr ; nada.matta@utt.fr; emmanuel.caillaud@unistra.fr

Phone numbers [1] (0033) (0)3 88 14 47 00, [2](0033) (0)6 07 34 03 89, [3](0033) (0)3 68 85 49 66
Underlined : corresponding author

Short Title: NLP for CoP activity evaluation

Manuscript 39 pages with Title Page, Tables and Figures

7 figures

10 tables

Systematic community of Practice activities evaluation through Natural Language Processing: application to research projects

Abstract

Community of Practice (CoP) efficiency evaluation is a great deal in research. Indeed, having the possibility to know if a given CoP is successful or not is essential to better manage it over time. The existing approaches for efficiency evaluation are difficult and time-consuming to put into action on real CoPs. They require either to evaluate subjective constructs making the analysis unreliable, either to work out a knowledge interaction matrix that is difficult to set up. However, these approaches build their evaluation on the fact that a CoP is successful if knowledge is exchanged between the members. It is the case if there is some interactions between the actors involved in the CoP. Therefore, we propose to analyse these interactions through the exchanges of emails thanks to Natural Language Processing. Our approach is systematic and semi-automated. It requires the e-mails exchanged and the definition of the speech-acts that will be retrieved. We apply it on a real project based CoP: the SEPOLBE research project that involves different expertise fields. It allows us to identify the CoP core group and to emphasize learning processes between members with different backgrounds (Microbiology, Electrochemistry and Civil engineering).

Highlights:

- We define the concept of Community of Practice (CoP) and its efficiency evaluation.
- We make a state of the art of the existing CoP diagnosis approaches
- We propose an approach to evaluate CoPs based on the e-mails exchanged.
- We apply the approach proposed to a real study case

Keywords: Community of practice, Performance Evaluation, Pragmatics, Natural Language Processing, Research Project

1. Introduction

Knowledge is one of the most valuable resources for modern organizations (Zack, 1999). The efficiency of organizations is strongly linked to the way the knowledge is created, shared and managed (Chu & Khosla, 2009; Nonaka & Takeuchi, 1995). After (Guptill, 2005) and (Kim, Hong, & Suh, 2012), Communities of Practice (CoPs) are particularly effective to consider the whole knowledge lifecycle. CoPs are "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis." (Wenger, McDermott, & Snyder, 2002).

In this boarder, the efficiency evaluation of these groups becomes crucial. Several propositions are worked out in the literature generally focused on the knowledge created and exchanged between the CoPs members. These approaches, as stated in (Kim, et al., 2012), can be grouped into: (i) subjective methods; (ii) objective methods. Generally, these are difficult and time consuming to apply on real CoPs. For example, the interaction matrix required for SNA (Social Network Analysis) as proposed in (Kim, et al., 2012) takes a lot of time to be built. In this paper, we exploit the analysis of electronic messages exchanged between the members of a CoP to refine the CoPs activities performance evaluation by going beyond the classical evaluation of knowledge sharing level largely studied in knowledge management (Zack, 1999). More precisely, we look for speech acts (Austin, 1975) in the message content to identify, among other, the role of the people involved in the CoP, the knowledge shared and the learning processes that take place between the CoP members. Pragma-linguistics techniques are then applied for this purpose (Levinson, 1983). Another advantage of our approach is that it is systematic and semi-automated.

The paper is structured as follows. The second section deals with the state of the art, describing the current approaches for CoPs performance evaluation. The third section presents the text analysis approach we propose for evaluating CoPs activities. In section

four the approach we propose is applied on a project based CoP the SEPOLBE scientific project dedicated to develop bioadmixtures for concrete. Section five concludes and gives further research directions.

2. State of the Art

In this state of the art, we will analyse the different definitions of the CoPs and how the efficiency of CoPs is evaluated.

2.1.CoPs definition

In (Wenger, 1998) CoPs are treated as the informal relations and understandings developed in mutual engagement on an appropriated joint enterprise. In other words, a community of practice is defined as a group that coheres through "mutual engagement" on an "appropriated enterprise", and creates a common "repertoire".

In (Wenger, et al., 2002) the concept is redefined towards a more managerial stance, making the concept more popular and simple. Here CoPs are "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis." According to (Cox, 2005), this definition is much vaguer than the previous one. The definition is of a group that is somehow interested in the same thing, not closely tied together in accomplishing a common enterprise. The purpose is specifically to learn and share knowledge, not to get the job done. From this point of view a CoP has three structural features:

(i) Domain: it "*… creates common ground and a sense of common identity. A well-defined domain legitimises the community by affirming its purpose and value to members and other stakeholders. The domain inspires members to contribute and participate, guides their learning, and gives meaning to their actions"* (Wenger, et al., 2002).

(ii) Community: it *"…creates the social fabric of learning. A strong community fosters interactions and relationships based on mutual respect and trust. It encourages a willingness to share ideas, expose one's ignorance, ask difficult questions, and listen carefully. Community is an important element because learning is a matter of belonging as well as an intellectual process."* (Wenger, et al., 2002).

(iii) Practice: it *"…is the specific knowledge the community develops, shares and maintains''* (Wenger, et al., 2002).

Considering our objective of a systematic approach based on the e-mail exchanged, we are interested in the practice feature of a CoP that is to say the way knowledge is created, shared and managed.

Different types of CoPs are identified in the literature. McDermott (2000) indicates four types of community:

(i) Communities which are linked to a strategic objective;

(ii) Communities which focus on tactical processes, process optimisation and sharing of best practice;

(iii) Project-based communities

(iv) Communities developing a particular body of knowledge.

We are interested in project based communities. This kind of CoP has several interesting features. First of all, it is confined in time with fix start and end, enabling to study all its lifecycle phases as detailed in (Lee, Suh, & Hong, 2010) from the building stage corresponding to the initial state of a CoP, where people come together and develop a detailed plan of structure of community to the close stage, during which a CoP declines or transforms into several other communities.

Secondly the people involved in the CoP are project members and therefore act actively in the achievement of the project. Their participation degree, as described in Wenger, et al.

(2002), is as core group. A core group is the group of people that carries out the work in the community. It actively participates in discussions and identifies the topics to be addressed within the community. The coordinator takes place in this group. He is the one who organises events, connects communities and generally keeps the community alive. Moreover, project members are known at the outset enabling to study their mutual interactions easily.

2.2.CoP performance evaluation

A CoP is defined as successful when its members exchange specific knowledge, practices and/or experiences that contribute to developing a practice (know-how) in a specific field (McDermott, 2004). The research around the performance evaluation and management of CoPs gains gradually importance. Indeed, in the nineties CoPs were presented as spontaneous, self-organizing, and fluid processes that management cannot intentionally establish (Brown & Duguid, 1991; Lave & Wenger, 1991). In other words, at that time it was considered that the performance of a CoP cannot be measured and managed to improve its efficiency. Later some works suggest that CoPs are amenable to manipulation and can be managed (Lesser & Everest, 2001; Wenger, 2000; Wenger & Snyder, 2000). In turn as a result of these works diagnosis frameworks for CoP efficiency management were proposed. These aim at assessing the knowledge creation or sharing level inside a given CoP. To do so, as stated in (Kim, et al., 2012), these frameworks are either based on subjective methods or on objective one. Subjective methods are methods that use qualitative constructs to assess the performance level whereas objective one assess this performance level through a quantitiative indicators. In the next sub-sections we detail two diagnosis frameworks one based on a subjective method (Borzillo & Kaminska-Labbe, 2011) and the second on an objective one (Kim, et al., 2012). For each category we analyze its advantages and drawbacks in order to set our research objectives.

2.2.1 CoP diagnosis framework based on subjective methods

In (Borzillo & Kaminska-Labbe, 2011) the underlying hypothesis of diagnosis is that organizations need to guide CoPs to generate usable knowledge sustainability. As a result, the authors aim to elucidate the related knowledge creation dymanics. Therefore, they go beyond the mere evaluation of the four factors – knowledge objectives, leadership, collaboration and boundary spanning – that are generally associated with knowledge creation in CoPs. Contrary to this "fragmented" approach, they propose an "integrated" one enabling to understand how the interactions between these factors lead to dynamic knowledge creation processes. To do so, they exploit the complex adaptative system (CAS) theory that provides an integrative and dynamic framework to understand the interaction patterns in networks of interdependent agents who interact and are bound by their common needs or objectives.As a consequence, the four factors of knowledge creation are translated according to the CAS theory lenses into four constructs.

First, the knowledge objectives factor becomes *adaptative tension* that drives self-organization and emergence. It emerges from external constraints and corresponds to the energy diffrential between the system and its environment. The focus is therefore on the definition of the upper and lower bounds of adaptative tension that will define the "region of complexity" inside which a system is able to create new order (self-organization) and producing new knowledge (emergence). Second, the leadership factor becomes *enabling leadership*. In this perspective, the managers' role is to design systems in which distributed intelligence can easily emerge. For CoPs, enabling leadership means enhancing the socialization between individuals. Third, collaboration becomes *enhanced cooperation*. Indeed, CAS theory emphasizes that knowledge creation depends on the nature of the interactions or connections between agents implying for CoPs regular meetings, workshops and the enabling information technology for interaction. Fourth, the *boundary spanning*

factor remains unchanged. It highlights cognitive diversity for knowledge creation by interacting with actors external to a given CoP.

Then, during 4 years, 5 CoPs from an international-operating industrial group were studied via a longitudinal exploratory study (six series of interviews) in order to uncover if the four constructs of CAS theory impact positively or negatively the dynamics of knowledge creation. The data gathered was analysed qualitatively (preliminary analysis for data categorization, within-in community analyses to search linkages between first order concepts, croos-community analysis to reveal consistencies and contradictions between the CoPs) and quantitatively (average ranking caculation of each constructs). As a result, two modes of CoPs are proposed. Each mode couples two of the four studied constructs (see Figure 1). When adaptive tension and enabling leadership are prevalent, the CoP is in a "guided" mode. When enhancing cooperation with boundary spanning are prevalent the CoP is in a "self-directed" mode.

---Insert the Figure 1 here---

These modes are then coupled with knowledge creation processes. The authors argue that a "guided-mode" supports knowledge expansion, while a "self-directed mode" simulates knowledge probing. The guided mode is used to improve the existing product offerings, which requires creating improved knowledge. During the "self-directed" mode, communities explore radically new knowledge.

This study is interesting as it exploits complexity theory to focus on the interaction of the factors leading to knowledge creation in CoP. Even if this link is not formally described, there is generally acknowledged that there is a positive correlation between performance and knowledge creation. However, it seems difficult, on a given CoP, to evaluate the four constructs: adaptive tension, enabling leadership, enhanced cooperation and boundary spanning. It requires time-consuming data collection (interviews from

different informants (community sponsors, leaders and members); attending to CoP workshops) and data exploitation. Moreover, only two modes are described in this paper the other possible configurations of the constructs are not considered. That is why another research stream explores the use of objective methods.

2.2.2 CoP diagnosis framework based on objective methods

In (Kim, et al., 2012) a framework diagnosis for CoPs is proposed based on SNA (Social Network Analysis). SNA is a scientific method to analyze a social network by focusing on patterns of relationships between actors and examining the availability of resources and their exchange between actors. Here, performance evaluation of CoPs focuses on knowledge sharing activity by providing a view of the relationship network between the members o f the CoP that is to say of the knowledge receivers and the knowledge knowledge propagators and the related amount of knowledge exchanged as shown in the conceptual framework of Figure 2.

---Insert the Figure 2 here---

The diagnosis methodology proposed by Kim, et al. (2012) is the following (see Figure 3):

1. Pre-process: this step enables to understand the methodology of knowledge sharing and to build the knowledge sharing matrix. This matrix records the knowledge propagators in columns, and knowledge receivers in rows. It will be the input for the SNA;
2. Analysis: SNA and development of new indexes for CoP diagnosis;
3. Strategy: suggestion of a strategy for future knowledge sharing activities.

---Insert the Figure 3 here---

The SNA is based on the knowledge sharing matrix set up during the pre-process step. If the CoP has n members the matrix will be an n*n size one with each cell filled. The

input data are generally retrieved from questionnaires, interviews and data log, even if other transaction data could be used (online messenger tools, e-mails, etc.). The data processed have to define if two members exchange data or not. For e-mails it requires to identify the sender and the recipients of a given message. It is stated that the working out of this matrix spends tremendous time. Moreover, the kind of questionnaires to use is not detailed.

By using a knowledge sharing matrix for input data, basic indexes can be generated by SNA. Some of the basic indexes are:

- Node type (transmitter or out-flow only node, receiver or in-flow only node, carrier or node with only one connected in-flow and only one out-flow except from the in-flow node, ordinary or node with a mixed in- and out-flow, isolate or node that is not connected to others);
- Network density that is an indicator for the general level of connectedness of the graph;
- Betweeness centrality: it is the share of times that a node i needs a node k (whose centrality is being measured) in order to reach a node j via the shortest path;
- In and out degree of centrality: it is the proportion of a connected edge to the maximum possible connections.

Based on these indexes the analysis step can be carried out. To do so, the authors provide a member and a CoPs typology. Both typologies are based on knowledge propagating and receiving abilities. There are four kinds of members (see Figure 4):

- Balanced player: a member who propagates knowledge to and receives knowledge from other members. This kind of member corresponds to the ordinary or carrier nodes;
- Egoistic propagator: a member who propagates knowledge to other members, but does not receive knowledge from other members. Such a member corresponds to the transmitter nodes;

- Egoistic receiver: a member who receives knowledge from other members, but does not propagate knowledge to other members. Such a member corresponds to the egoistic receiver nodes;
- Knowledge isolator: a member who does not propagate knowledge to or receive knowledge from other members. These members are the isolator nodes.

---Insert the Figure 4 here---

Based on the member typology the CoP typology is set up (see Figure 5). It is based on the knowledge receiving and knowledge propagating core group ratio that is to say the ratio between the propagators or the receivers and the total number of member in the core group. The core group is identified thank to in- and out-degree centrality. Four communities' types are proposed:

- Active community: the core group has a high ratio of knowledge propagation and receiving;
- Spreading community: the core group has a high ratio of knowledge propagation but a low ratio of knowledge receiving;
- Learning community: the core group has a low ratio of knowledge propagation but a high ratio of knowledge receiving;
- Inactive community: this community has low ratios of knowledge propagation and receiving within the core group.

---Insert the Figure 5 here---

Then, diagnosis of the CoP under study (thrid step of the framework) is made according the CoP typology. For each community type improving strategies like, for example, "create more practical knowledge" or "redefine the knowledge domain" are proposed.

This diagnosis framework is complete as it observes a CoP according knowledge propagating and receiving actions and proposes improving strategies. However, this

approach has two main drawbacks. First, the setting up of the knowledge sharing matrix is complex and time consuming, even if e-mails could be exploited to retrieve the required data automatically. Secondly the diagnosis focuses only on the knowledge exchanges on a binary mode (knowledge receiving yes/no and knowledge propagating yes/no). There is no in-deep analysis of the quality and the type of exchanges that would be interesting to better highlight the CoP performance.

2.2.3 Synthesis

Existing diagnosis methods for CoPs rely efficiency evaluation on the ability of a CoP to exchange and sustain knowledge over time. However, these are difficult and time-consuming to put into action on real CoPs. They require either to assess subjective constructs, either to work out a matrix of knowledge interaction. In both cases the input data required for evaluation stem from questionnaires and interviews of the members of the CoP.

In (Borzillo & Kaminska-Labbe, 2011) four subjective constructs (adaptative tension, enabling leadership, enhanced cooperation and boundary spanning) are evaluated and their interactions studied. This enables to reflect the way knowledge is created inside a given CoP in a “guided” or “self-directed” mode.

In (Kim, et al., 2012) the members who propagate knowledge and these who receive knowledge are identified and their interactions are analysed. This enables to evaluate the level of knowledge sharing in a binary manner “Is knowledge exchanged or not between two members ?”, then to characterize the CoP members and the CoPs according their role in knowledge sharing (propagator and receiver) and to last to provide improving strategies for each kind of CoP identified.

The main advantage of the approach of (Kim, et al., 2012) is to evaluate the CoP efficiency in an objective manner, even if the indicators proposed are binary and do not

reflect as it is the case for (Borzillo & Kaminska-Labbe, 2011) the complexity of knowledge sustainability. In our view, the objective method option such as proposed in (Kim, et al., 2012) is the most promising because it ensures the reliability of the evaluation. However, it could be improved and completed with a semi-automated analysis of the content of the knowledge shared as well as the related intentions of the actors. In this way, the nature of the knowledge shared can be identified so as the contribution of this sharing to organizational learning by indentification of the leraning processes involved and the role of the actors in the CoP. For this purpose, our approach looks for the impacts of interactions on the CoP users and their organization by analysing communication.

There are different kinds of techniques to analyse communication (as TextMining, Natural Language Processing (NLP), etc.). On the one hand, TextMining groups a set of techniques enabling to extract information from documents. On the other hand, NLP is the field of study that focuses on the interactions between human language and computers. Both approaches do not enable to emphasize the interactions between actors that is the main feature of knowledge sharing and learning evaluation in CoPs. Therefore, we use Pragmatics approach because that helps to analyse the content of communication and to identify intention of interactions between participants. We present in the next section this type of analysis.

3. Text analysis for CoP performance evaluation

3.1.Communication analysis

Several approaches study how to analyze communication as a specific discourse. We note for instance, tagging work in (Yelati & Sangal, 2011), in which the authors present techniques that help to identify topics in e-mails. We also note NLP (Natural Language Processing) community on automated speech act identification in emails (Baron, 1998; Carvalho & Cohen, 2006; Corney, De Vel, Anderson, & Mohay, 2002). For instance,

Kalia, Motahari Nezhad, Bartolini, and Singh (2013) use NLP in order to identify messages concerning tasks and commitment. They parse verbs and sentences in order to identify tasks and track messages between senders and receivers. Pragmatics, the study of language in use, is concerned with the intended meaning of speakers beyond what is explicitly stated. It is a branch of linguistics concerned with the use of language in social contexts and the ways in which participants produce and comprehend meanings through language. Pragmatics focuses on aspects of signification that are not only predicted by linguistics knowledge. It is concerned with the analysis of the speaker's meaning rather than on the meaning of words and utterances (semantic or linguistic meaning). Thus Pragmatics takes into account the role of physical and social context (Austin, 1975). Pragmatics analysis of communication using e-mails uses only some of these methods like ngrams analysis by (Carvalho & Cohen, 2006), Verbal Response Mode scheme by Lampert in (Lampert, Dale, & Paris, 2010) or a custom coding scheme like in (Felice & Deane, 2012).

As cited above, we use Pragmatics in order to study communication in community of practices and topic identification. Our aim is first to identify if contributors learns from each other's and on which topics. So, we apply the CaMCa "Context aware Mediated Communication analysis" approach (Rauscher, Matta, & Atifi, 2016), we develop for this aim. CaMCa is based on Pragmatics analysis and context awareness. So, it helps to identify, from one side, intention of communications and from another side, links interactions to the activity context. CoPs activity evaluation needs these dimensions because participants activities and roles are important. As the analysis is context dependent, it makes no sense to study big volume of interactions such as in SNA. The objective differs from SNA for e-mails and that could be used to improve the CoP diagnosis framework of (Kim, et al., 2012). Indeed, as detailed in (Tang, Pei, & Luk, 2014)

SNA for e-mails is generally used for contact identification focusing on the structure of the network built from the email corpus and paying less attention to email contents.

3.2.CaMCa approach

Pragmatics puts on the fact that a dialogue is context and time dependent. Identifying the sense of interaction is related to the conditions, environment and situation of communication. We develop the Context aware Mediated Communication analysis (CaMca) approach that considers from one side the context and the domain of the activity and from the other side the speech acts in mediated communication.

The different phases of this approach are (Rauscher, et al., 2016) (see Figure 5):

- Context identification:
  - Skill and role of actors;
  - Phases of collaboration among time;
  - Goal of collaboration.
- Domain identification
  - Domain topics and subjects.
- Communication analysis
  - Sender/receivers/CC;
  - Date/Hour;
  - Subject;
  - Thread of communication: Reply, comments;
  - Main Speech Acts.

---Insert the Figure 6 here---

Each phase of CaMCa will bring a piece of information about CoP activity. Indeed, the context identification is useful to underline the actors' skills and collaboration roles corresponding to the "Community" feature of a CoP as described in section 2.1.

Thisenables to identify the core group and the coordinator of the CoP under study. The domain identification will be used to identify the topics on which the CoP exchanges, enabling to define the nature of knowledge shared. That is the “Domain” feature of a CoP. Last but not least the Mediated Communication Analysis focuses on the e-mail exchanged enabling to assess the degree of knowledge sharing and interactions between CoP participants (“Practice” feature) and the related knowledge processes that take place.

In fact, in context we try to identify the organization of the communication. It mainly concerns actors and their collaboration goals. The domain analysis puts on, the nature of the activity subject of interactions, on which practices participants discuss. Finally, the analysis of communication helps to emphasis the real effects of interactions: what is happened when actors received and post messages? Is there any learning, coordination, conflict, alliances, etc? and how it is done? CaMCa is used in order to identify the nature of interactions in CoPs. Then the results obtained can be compared to the CoP typology depicted in Figure 4. In this way, performance and diagnosis of studied CoPs can be put on in a systematic way.

So, we propose to apply CaMCa to a project based CoP. Interaction around a project is analysed and first principles of systematic CoPs diagnosis approach is determined. These principles are identified by analysing the communication of the real SEPOLBE project, presented in the next section.

4. Case study

In this section, we will illustrate the use of CaMCa on a project-based CoP. We will analyse the exchanges between the members of the Cop. The roles of the actors, particularly the coordination, will be analysed. The knowledge exchanged and created will also be studied.

4.1.Study context: the SEPOLBE Project

We apply the CaMca approach to the SEPOLBE research project in order to identify the knowledge shared and the learning processes that take place between the members and to analyse how the CoP is coordinated (Who is the coordinator?; Level of coordination). The SEPOLBE project is dedicated to develop bioadmixtures for concrete (Goepp, Munzer, & Feugeas, 2014). These substances are conscientious of the environment and should limit the bio contamination of the concrete surface and improve the resistance to corrosion of its metallic reinforcement. This project implies four research teams and a company. The research teams have different complementary areas of competency: concrete surface analysis, physical chemistry analysis of films on steel and concrete, electro chemistry for steel corrosion inhibition, petro physics for concrete physical chemistry characterization. Competences in micro biology, chemistry and microscopy are also required to develop the substances and to analyse the surface bio contamination. The industrial partner manufactures and markets concrete products such as admixtures. Its product mix already includes protection products but none of them is dedicated to biological contamination. The duration of this project is about four years. Actors are coming from three main domains: Microbiology, Electrochemistry and Civil engineering. To show the applicability of our approach, we focus on two tasks of the project: (i) Project coordination and (ii) Assessment of the cleaning ability of the mortar base surfaces. Here, objective is to evaluate the ability of the bioadmixtures to limit the development of biofilms on the concrete surface and its impact on the cleaning concrete surface (biofilm dropping out). The ICube and B2HM teams are in charge of this task. The ICube team has to provide to the B2HM team “adequate” concrete samples. The B2HM is in charge of the contamination and cleaning tests.

For these tasks we analyze the e-mail exchanges between the people involved in the task. The people involved provided us the e-mails they received and sent to complete a

given task (coordination or cleaning ability evaluation). So, we had access to the e-mails as a whole (sender, recipients, date, content, etc.)

4.2.Learning evaluation using E-mail analysis

The main topics of the task dealing with the assessment of the cleaning ability of the mortar base surfaces concern: Concrete, Mortar, Sample, Bioadmixtures (or BA), Molecular, Bacteria, Essay, Experiment, Polishing, Sample, Ultrasonic, …

Actors communicate together using mainly e-mails but if the content of message is available other sources of electronic messages could be exploited such as forums could be exploited. So, we analyze their communication in order to understand if there was any learning of procedures or concepts. First of all, our expert on Pragmatics identifies a grid of main speech act types concerning learning (see Table 1).

---Insert here the Table 1---

Then, based on different synonyms and sentence forms given by the pragmatics expert and the list of project topics, the NLP algorithm (Lucen) has been used in order to retrieve the corresponding messages. Figure 7 illustrates the global results obtained. The inside wheel shows the topics identified and the outside wheel the corresponding speech acts.

---Insert the Figure 7 here---

In our analysis, we try to identify if there is learning of concepts or procedures. So, we look for some specific speech acts like proposition, explanation, request, verification and information. Then we study this type of interaction among time: at the beginning, on the middle and at the end of task and project. For instance, at **20/12/12** Charlotte “civil engineer” asks about the Bioadmixtures experiment conditions needed. Thierry “Microbiologist” answers her by explaining a procedure (see Table 2).

---Insert here the Table 2 ---

At 21/03/13, some months after this interaction, Charlotte “civil engineer” presents some modification on the procedure (see Table 3).

---Insert here the Table 3 ---

We suppose then that Charlotte learns the procedure, uses it and tries to adapt it to specific conditions.

At 10/04/13, Charlotte asks for more verification about the samples conservation and Bioadmixtures test conditions of sample treatment. Chao and Thierry define more specifications about these conditions (see Table 4).

---Insert here the Table 4---

Charlotte asks then more verification about the procedure she applies (see Table 5).

---Insert here the Table 5---

We suppose that among these interactions, actors learned from each other about bioreceptivity mixture experiment and sample treatment and conservation. Other analysis of messages show also interactions about samples’ name coding and sending modes.

The chronology of interactions proves their dynamicity especially if actors have a timeline to respect, they communicate about the project phases.

4.3.Coordination analysis

Using CaMca approach, e-mails are also analysed in order to analyse how the CoP is coordinated. For this purpose, our analysis is based on coordination intentions (Matta, Atifi, Sediri, & Sadgal, 2011). A specific grid was defined containing speech acts related to coordination and topics concerning the coordination (see Table 6). The main coordination speech acts are about information, proposition and request. Topics are around: meetings, documents and reports. A statistical analysis is also done in order to identify the engagement of actors in interactions. This type of analysis can complete those done by SNA in (Kim, et al., 2012), especially the CoPs users behaviors studied. Adding to

statistical studies of exchanges, the identification of the interaction roles as used in our study help to emphasis the dynamic organization movements.

---Insert here the Table 6---

42 messages have been analysed using this grid and based on the CaMCa approach. These messages correspond to four milestones of the projects: Kick off, 12th, 18th and 30th month. Message date, Senders, Receivers, speech acts, topics are identified. 101 important sentences are identified from these messages. Table 7 illustrates a part of the analysis for the Kick off milestone.

---Insert here the Table 7 ---

First statistics analysis syntheseized in Table 8 shows that Francoise is the main animator of the group. She takes the role of animator of coordination. She is the author of 22 messages. In these messages, Françoise informs about project documents, meeting reports and meetings logistics (19 speech acts), proposes meetings schedules and project presentations (17 speech acts) and asks for modification on Meetings Reports, logistics and project documents. We observe also the dynamic participation of other actors like Bernard (2 speech acts on Information and 1 on Request about Meeting logistics) and Anouk (1 speech act on Information and 1 Request about Meeting logistics) in the organization of meetings.

---Insert here the Table 8 ---

Finally, even messages are addressed to all participants of projects (16 participants), only 7 participate in the discussions. Time interaction analysis shows that messages are close to meetings dates.

As showed in this analysis, CaMca can be used to identify the nature of animation of CoPs and the core group. In SEPOLBE, we can see that some actors are engaged in the animation of CoPs and form the core group that is essential for a successful CoP. The

animation is dynamic around meetings and tasks deadline. Beside that, interactions concern explanations of procedures and precision of techniques.

4.4.Discussion

We focus on this paper on a systematic and semi-automated analysis of CoPs that goes beyong knowledge sharing level assessment. Our analysis technique is based from one side on interaction content analysis and from other side on participants' competencies. This technique is summarized in Table 9.

---Insert here the Table 9---

The analysis of the SEPOLBE project interactions shows that we can answer to some characteristics of CoPs using CaMCa approach especially (see Table 10). So, in the SEPOLBE project there are two interaction sequences about Bioreceptivity Experiment and Samples Conservation that show learning between Charlotte and Thierry. The close dates of interactions can emphasize an active community but only 40% of participants send messages and only 18% are involved on coordination. So, we can note that only 40% of the community are active. The 50% of coordination messages emphasize the cooperation dimension of the CoPs.

Francoise can be identified as the coordinator because she is implied in 22 messages using propose and request speech acts. Finally, we can note that the effective participants are balanced player; all participants are in copy of messages or as receivers, even when there is some dialogue between Charlotte and Thierry (see Table 10).

---Insert here the Table 10 ---

As showed in this paper, combining messages analysis and context awareness can give a technique to go beyong the classical knowledge sharing level assessment of CoPs. We tend in our analysis to identify a systematic methodology that help to diagnosis CoPs. This methodology is based on intention identification principle (Richard, 1990), which shows

that a sense interpretation is linked to an action and environment. We used an approach mixing statistics and content analysis. We succeeded in identifying the real coordinator of the CoP and how the actors learn from each other thanks to the emails exchanges.

5. Conclusion

CoP efficiency evaluation is a great deal in research. Indeed, having the possibility to know if a given CoP is successful or not is essential to better manage it over time. The existing approaches for efficiency evaluation are difficult and time-consuming to put into action on real CoPs. They require either to evaluate subjective constructs making the analysis unreliable, either to work out a knowledge interaction matrix that is time-consuming to set up.

These approaches build their evaluation on the fact that a CoP is successful if knowledge is exchanged between the members but they focus evaluation on the level of knowledge shared and created. This is a useful but partial. Indeed, when knowledge is shared there is some interactions between the actors involved in the CoP. Therefore, we propose to analyse in detail these interactions through the exchanges of emails thanks to NLP. Our approach is easy to put in action as it is systematic and semi-automated. It requires the e-mails exchanged and the definition of the speech-acts that will be retrieved. Our approach allowed us to identify one of the key role of a project-based CoP: the leader of the core group. We also succeeded in identifying the learning process during the project between stakeholders from different domains: civil engineering and biochemistry. These first promising results must be confirmed on other project-based CoPs but also on other types of CoPs for instance to study problem solving exchanges, experts identification, etc.

Last but not least it would be interesting to couple the proposed approach general guidelines of CoP management as they are proposed in (Probst & Borzillo, 2008) or (Jeon, Kim, & Koh, 2011). Indeed, in (Probst & Borzillo, 2008) the most salient reasons for the

success and failure of CoPs are worked out. An investigation on 57 CoPs from major European and US companies led to the discovery of 10 “commandments” that lead to the successful development of CoP. These ten “commandments” describe CoP governance practices and could be linked with CoP diagnosis in order to apply the “commandments” that the best fit to a given CoP context. Jeon, et al. (2011) identify and validate a set of organizational factors that was anticipated to have effects on knowledge sharing by CoP members such as perceived consequences, affect, social factors and facilitating conditions. According to a given CoP diagnosis the corresponding organizational factors could be put into action.

6. References

Austin, J.L. 1975. *How to do things with words* Vol. 367. Oxford university press.

Baron, N.S. 1998. Letters by phone or speech by other means: The linguistics of email. *Language & Communication, 18,* 133-170.

Borzillo, S., & Kaminska-Labbe, R. 2011. Unravelling the dynamics of knowledge creation in communities of practice though complexity theory lenses. *Knowledge Management Research & Practice, 9,* 353-366.

Brown, J.S., & Duguid, P. 1991. Organizational leraning and communities of practice: Toward a unified view of working. *Organization Science, 2,* 40-57.

Carvalho, V.R., & Cohen, W.W., 2006. Improving email speech acts analysis via n-gram selection, HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech, Association for Computational Linguistics. Publishing, Stroudsburg, PA, USA, pp. 35-41.

Chu, M.-T., & Khosla, R. 2009. Index evaluations and business strategies on communities of practice. *Expert Systems with Applications, 36,* 1549-1558.

Corney, M., De Vel, O., Anderson, A., & Mohay, G., 2002. Gender preferential text mining of e-mail discourse, 18th Computer Security Applications Conference. Publishing, Las Vegas, NV, USA.

Cox, A. 2005. What are communities of practice? A comparative review of four seminal works. *Journal of Information Science, 31*, 527-540.

Felice, R.D., & Deane, P., 2012. Identifying speech acts in emails: Toward automated scoring of the TOEIC® Email task. Publishing.

Goepp, V., Munzer, C., & Feugeas, F., 2014. Community of Practice Theory and Process Modelling: Two Tools for Better Collaboration in Research Projects, Advances in Production management systems : Innovative and knowledge-based production management in a global-local world, APMS 2014. Publishing, Ajaccio, France.

Guptill, J. 2005. Knowledge management in health care. *Journal of Health Care Finance, 31*, 10-15.

Jeon, S.-H., Kim, Y.-G., & Koh, J. 2011. Individual, social, and organizational contexts for active knowledge sharing in communities of practice. *Expert Systems with Applications, 38*, 12423-12431.

Kalia, A., Motahari Nezhad, H.R., Bartolini, C., & Singh, M., 2013. Identifying business tasks and commitments from email and chat conversations. Publishing.

Kim, S.J., Hong, J.Y., & Suh, E.H. 2012. A diagnosis framework for identifying the current knowledge sharing activity status in a community of practice. *Expert Systems with Applications, 39*, 13093-13107.

Lampert, A., Dale, R., & Paris, C., 2010. Detecting emails containing requests for action, Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics. Publishing, Los Angeles, USA, pp. 984-992.

Lave, J., & Wenger, E. 1991. *Situated learning. Legitimate Peripheral Participation*. University Press, Cambridge.

Lee, J., Suh, E., & Hong, J. 2010. A maturity model based CoP evaluation framework: A case study of strategic CoPs in a Korean company. *Expert Systems with Applications, 37*, 2670-2681.

Lesser, E., & Everest, K. 2001. Using communities of parctice to manage intellectual capital. *Ivey Business Journal, 65*, 37-41.

Levinson, S.C. 1983. *Pragmatics*. Cambrigde University Press, Cambridge.

Matta, N., Atifi, H., Sediri, M., & Sadgal, M. 2011. Coordination interactions analysis in design projects. *International Journal of Communication and Information Sciences, AICIT, HumanPub, 1*, 22-29.

McDermott, R. 2000. Critical success factors in building communities of practice. *Knowledge Management Review, 3*.

McDermott, R. 2004. How to avoid a mid-life crisis in your COPs: Uncovering six keys to sustaining communities. *Knowledge Management Review, 42*, 10-13.

Nonaka, I., & Takeuchi, H. 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, New-York.

Probst, G., & Borzillo, S. 2008. Why communities of practice succeed and why they fail. *European Management Journal, 26*, 335-347.

Rauscher, F., Matta, N., & Atifi, F., 2016. KTR: an approach that supports Knowledge extraction from design interactions, 8th IFAC Conference on Manufacturing Modelling, Management & Control (MIM). Publishing, Troyes, France.

Richard, J.F. 1990. *Mental activities, Understand, Reason, Find solutions (in French) Les activites mentales, Comprendre, raisonner, trouver des solutions*. Armand Colin, Paris.

Tang, G.T., Pei, J., & Luk, W.S. 2014. Email mining: tasks, common techniques, and tools. *Knowledge and Information Systems, 41*, 1-31.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press, Cambridge.

Wenger, E. 2000. Communities of practice and social leraning systems. *Organization, 7*, 225-246.

Wenger, E., McDermott, R., & Snyder, W.M. 2002. *Cultivating Communities of Practice*. Harvard Business School Press, Boston.

Wenger, E., & Snyder, W.M. 2000. Communities of practice: The organizational frontier. *Harvard Business Review, 78*, 139-145.

Yelati, S., & Sangal, R., 2011. Novel Approach for Tagging of Discourse Segments in Help-Desk E-Mails, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Publishing, Lyon, France, pp. 369-372.

Zack, M.H. 1999. Developing a knowledge strategy. *California Management Review, 41*, 125-145.

7. Tables

Speech acts	Example of verbs
Information	Inform, Send, …
Explication	Explain, Clarify, Precise, Propose
Evaluation	Agree, Refute
Description	Describe, Detail, Can, Propose
Request	Verify, Ask, Wonder, Would you, Could you

Table 1 Main speech acts related to learning communication

Date	Sender	Skill	Reciever	Subject	Speech Act	Sentence
20/12/12 11:44	Charlotte	Civil engineering	To : Thierry	Subject : bioreceptivity Experiment	Request	Can you **explain** the Bioadmixtures experiment you do on samples of concrete: conditions of essai…
11/02/13 15:27	Thierry	Microbiology	To : Charlotte	Subject : bioreceptivity Experiment	Proposition, Explanation	We propose to proceed in two times: Short one depending on Samples recieving time and a long one (after a cure time) in order to detect ageing of samples and biofilms development

Table 2 Communication between actors for procedure clarification

Date	Sender	Skill	Reciever	Subject	Speech Act	Sentence
21/03/13 10:29	Charlotte	Civil engineering	To : Thierry	Subject : bioreceptivity Experiment	Explanation	Otherwise, I **define** a new polishing techniques that allows to have a better planeness of the samples…

Table 3 Modification of the procedure

Date	Sender	Skill	Reciever	Subject	Speech Act	Sentence
10/04/13 14:38	Charlotte	Civil engineering	To : Chao, Thierry	Subject : Sample Conservation	Request Verification	I **need** to **verify** some elements in the procedures of the samples treatment.
15/04/13 14:58	Thierry	Microbiology	To : Charlotte	Subject : Sample Conservation	Explanation	Right, we need to respect the same conditions for sample treatment... below, you find all precisions on sample treatment...

Table 4 Procedure verification about sample conservation

Date	Sender	Skill	Reciever	Subject	Speech Act	Sentence
21/06/13 15:43	Charlotte	Civil engineering	Chao	Samples treatment	Request Verification	**I want to be sure** about the modification of the procedure…
						I turned out the samples, I stored them in filtred water
						Tomorrow I will polish them, treat them with Uvs and I will send them to you.

Table 5 Verification of the procedure for sample shipping

Speech Acts			Topics
Information	Proposition	Request	
Send, Find below, Find, Knowing that, Here is, Remember that, Note that, Take place, Held...	Propose, Will be, Must, Think, Fix, can be, Using of future time	Ask, Wonder, Could you, Would you; Invite, To complete, Please Confirm…	Meeting date, Meeting Place, Presence, Participation, Meeting Start, Meeting Schedule, Meetings Report, Documents, Presentations, Web site, Document Sharing

Table 6 Coordination main speech acts and topics

Date	Sender	Speech Act	Topic	Related Verbs from e-mails	Related Topics from e-mails
Kick off Meeting					
13/07/2012 12:22	Françoise	Information	Document	I Inform you that today I will send the financial document	I Inform you that today I will send the financial document
		Request	Meeting Date	Could you also give me a date between 12 and 14 september for a meeting.	Could you also give me a date between 12 and 14 september for a meeting.
	Bernard	Information	Meeting Date and Place	The first meeting date will be held on Monday 17 september at LISE.	The first meeting date will be held on Monday 17 september.
		Request	Participation	Please confirm your participation	Please confirm your participation
		Proposition	Meeting Schedule	I propose the following Meeting Schedule	I propose the following Meeting Schedule

Table 7 Example of analysis of messages

Sender	Speech act	Topic	Number
Francoise	*Information*	Project Documents	8
		Meeting Report	5
		Project Presentation	2
		Meeting Date, Place	2
		Meeting logistics	2
	Total		**19**
	Proposition	Meeting Date, Place	7
		Project Presentation	3
		Meeting Scheduling	7
	Total		**17**
	Request	Meeting Date, Place	6
		Meeting Report	2
		Meeting Scheduling	1
		Meeting logistics	2
		Project Documents	3
		Project Presentation	3
	Total		**17**
Bernard	*Proposition*	Meeting Date, Place	1
		Meeting logistics	1
	Total		**2**
	Request (Total)	Meeting logistics	**3**
Anouk	*Information (Total)*	Meeting logistics	**1**
	Request (Total)	Meeting logistics	**2**
Thierry	*Request (Total)*	Meeting logistics	**1**
Thierry, Isabelle, Sandrine, Anouk, Sebastien, Virginie	*Information (Total)*	Meeting Participation	**7**

Table 8 Summary of interaction analysis

	Learning	**Coordination/Engagement**
Parameters	Speech Acts: Explanation, Request, Verification, Evaluation	Speech Acts: Inform, Propose, Request
	Participants' Competencies	Animation subjects
	Main Topics of CoPs	Sender- Receivers
	Date of Interaction	Date of Interaction
Method	Identification of Speech Acts on messages related to specific sentences	Identification of Speech Acts on messages related to specific Sentences
	Description of Participants Competencies from participants presentation	Identification of Animation objects from messages
	Identification of main topics of CoP from CoP presentation, linking message to main topics	Identification of Senders/Receivers
	Linking Topic-speech act- competences - Date	Linking sender to Speech Acts and object
		Count messages related to Senders/Speech acts and objects
Analysis	For each topic: - Analyze iteration sequence of : - Explanation - Request - Verification - If different competencies of messages senders then - Suppose learning between participants	For each sender: - Analyze frequencies sequences of: - Object - Proposition/Information - Request - Answer-sender-date - If frequencies of sequences is important then the Sender play animator role for the object - If Answer-sender-date is close to animator messages then the animation is dynamic - Compare senders and message Recievers : - Messages Recievers do not interact then related CoPs participants are not engaged

Table 9 Our analysis technique

		SEPOLBE analysis	
		Subject : bioreceptivity Experiment	**Samples' Conservation**
Learning Community	Repeating Sequence: Explanation, Request, Verification, Evaluation	20 dec. Charlotte: Request / 11 Feb. Thierry: Explanation / 11 Feb. Thierry: Request / 11 Feb. Charlotte : Explanation / 21 Feb. Thierry Explanation / 21 Mar. Charlotte: Explanation	10 Apr. Charlotte: Verification / 10 Apr. Charlotte: request / 15 Apr. Thierry: Explanation / 21 June Charlotte : Verification
	Different competencies of Interactions' participants	Charlotte : Civil engineering / Thierry: Microbiology	
Active Community	Close date of interactions	Discussions around Two or three months	
	Important volume of interactions	about 30 messages per year	
	Answers from the majority of participants	40% active participants	
	Participants involved on coordination	18% participants involved on coordination	
Enhancing cooperation	Important volume of coordination messages	50% coordination messages	
Enhancing Learderships	Sender using a lot of Propose, Request speech acts	Francoise: 22 messages	
Balanced Player	Messages sent to all participants	Messages sent to all participants or CC to all participants	

Table 10 Characterization of CoPs using CaMCa approach

8. Figure captions

Figure 1: Guided and self-directed modes of CoPs (Borzillo & Kaminska-Labbe, 2011)

Figure 2: Conceptual framework of knwoledge sharing activity in a CoP (Kim, et al., 2012)

Figure 3: Diagnosis process of CoPs based on Social Network Analysis (SNA) (Kim, et al., 2012)

Figure 4: CoP member typology according to (Kim, et al., 2012)

Figure 5: CoP typology according to (Kim, et al., 2012)

Figure 6: CaMCa basic approach

Figure 7: Results of NLP analysis

9. Figures

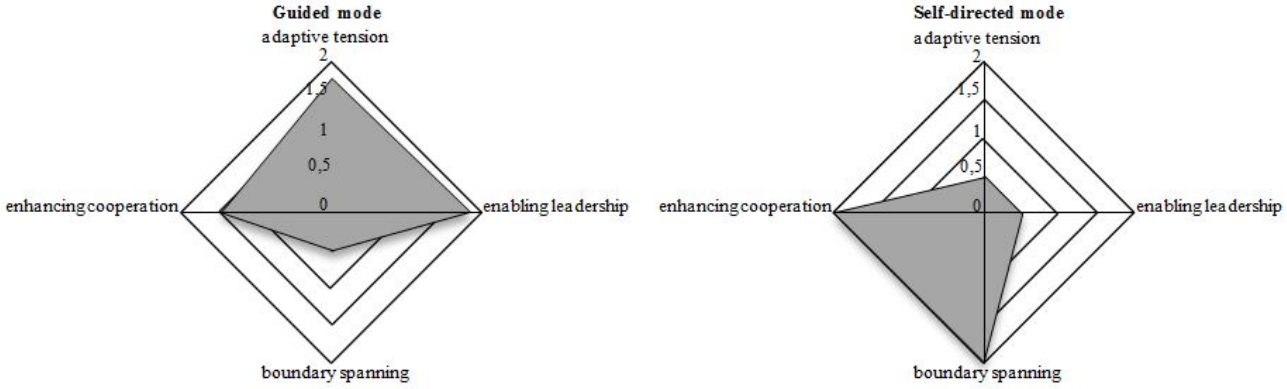

Figure 1: Guided and self-directed modes of CoPs (Borzillo & Kaminska-Labbe, 2011)

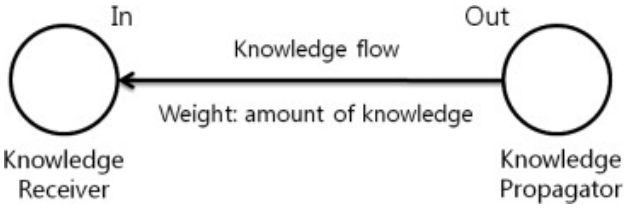

Figure 2: Conceptual framework of knwoledge sharing activity in a CoP (Kim, et al., 2012)

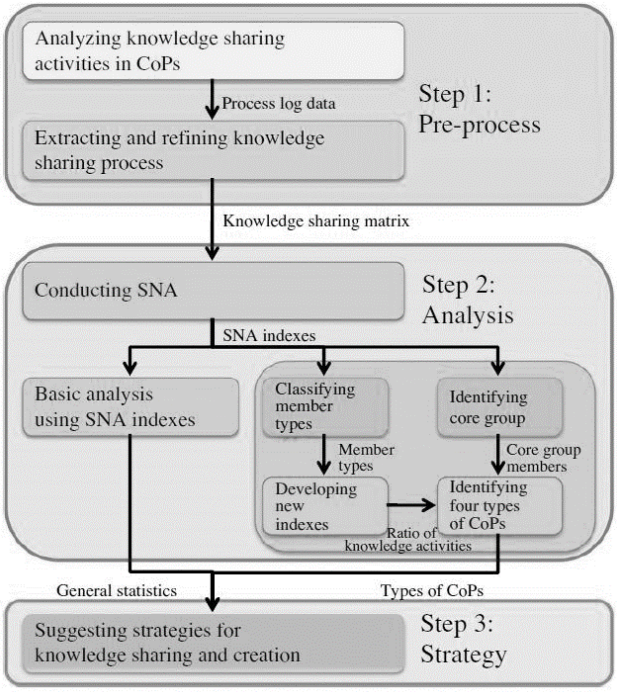

Figure 3: Diagnosis process of CoPs based on Social Network Analysis (SNA) (Kim, et al., 2012)

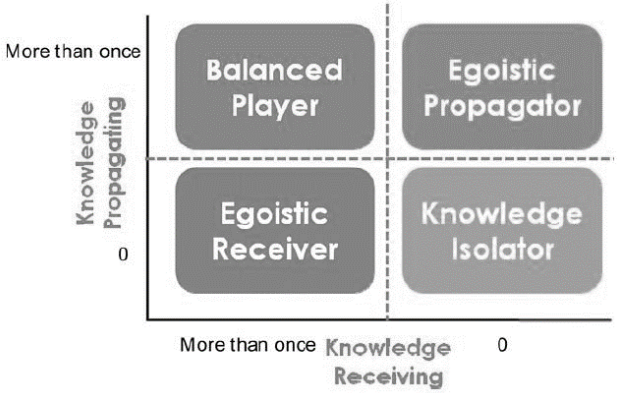

Figure 4: CoP member typology according to (Kim, et al., 2012)

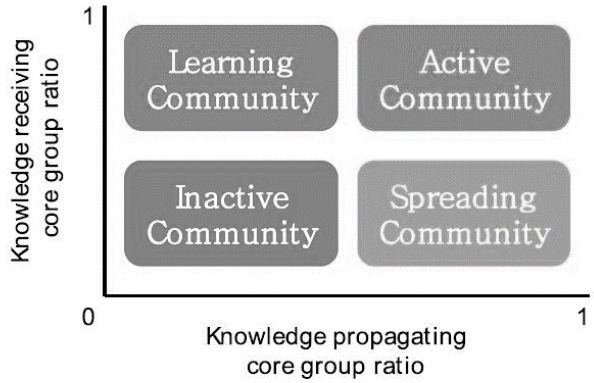

Figure 5: CoP typology according to (Kim, et al., 2012)

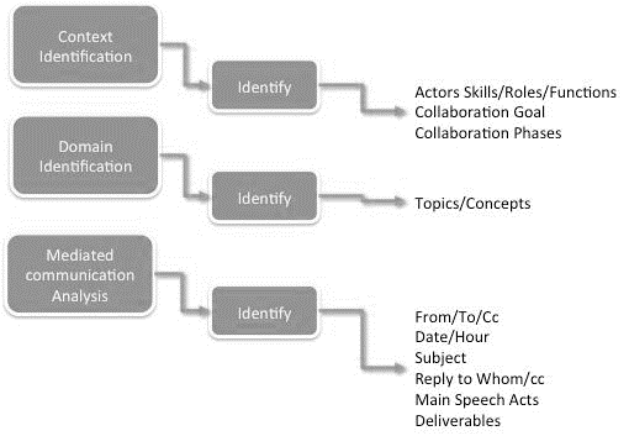

Figure 6: CaMCa basic approach

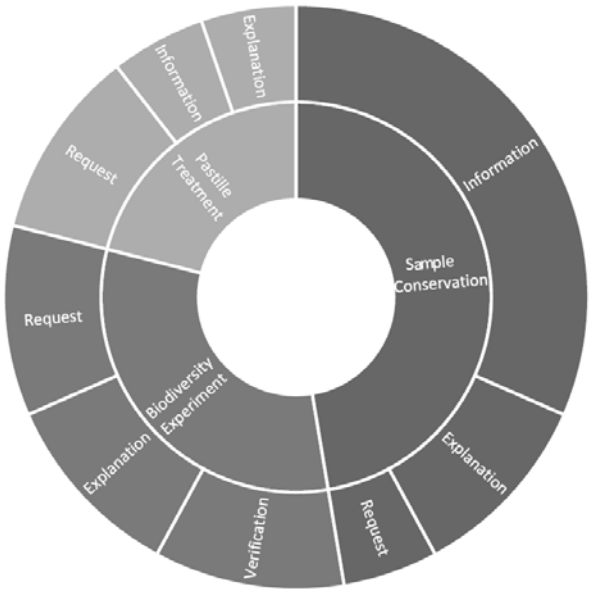

Figure 7: Results of NLP analysis