ERCIM "Alain Bensoussan" Fellowship Scientific Report

Fellow: Ioanna Lykourentzou
Visited Location : INRIA Nancy Grand-Est
Duration of Visit: 05/07/2011 - 04/04/2012

I - Scientific activity

(1 page at maximum)

My scientific activity during the second period of my fellowship at INRIA Nancy Grand-Est, focused on two main axes: i) semantic information indexing and retrieval using the method of Formal Concept Analysis and ii) collective intelligence algorithms to improve crowd collaboration in wikis. For these I collaborated with Dr. Amedeo Napoli and the Orpailleur team. The above research activities are briefly analyzed in the following:

Semantic information indexing and retrieval

I worked using the method of Formal Concept Analysis on semantic information indexing and retrieval tasks. The initial application domain was song indexing and retrieval, which then we expanded to other types of document datasets.

Specifically, as the amount of information grows, the ability to retrieve documents relevant to the user increasingly becomes a necessity. Regarding this task, several approaches have been proposed in the field of Information Retrieval (IR). However, as the information becomes more complex (including not only text but also multimedia documents) and specific (e.g. domain-oriented), the ability to organize it is as important as the capacity to retrieve it. Formal Concept Analysis (FCA) is a robust framework to organize objects in a concept lattice according to the attributes that they share. Indeed, concept lattices have been used in the past to support IR, with better or comparable performance to traditional approaches, such as Hierarchical Clustering and Best Match Ranking. Together with the rest of the team, I worked on the enhancement of its performance by considering features such as structural and semantic concept similarity metrics and by proposing new lattice navigation techniques.

Collective intelligence algorithms

I worked on the field of collective intelligence-based algorithms that use resource allocation techniques in order to coordinate a user crowd towards optimized collaboration results. The application field of the algorithms was the domain of corporate wikis.

In particular, a typical problem that enterprises face nowadays is how to effectively harness the tacit knowledge of their staff members. Tacit knowledge harnessing and its codification into usable document formats are vital for the enterprise since they directly affect its ability to innovate and solve complex new problems. Wikis are typically used for this purpose, especially when the corporation wishes to involve many of its staff members into the tacit knowledge harnessing process. However, typical wikis work on a self-coordination basis, where users self-appoint themselves to the wiki articles that they will contribute to. Due to this coordination pattern, current wikis cannot guarantee the quality of the produced wiki articles, or the time that these will reach acceptable quality levels. Given this, I proposed that the above problem can be formulated as a resource allocation problem, where the resources of the system are the users and their expertise, the tasks are the articles that need enhancement and the objective is to match people to articles so that overall the quality and timeliness of the produced knowledge is

increased. In this context, I worked on designing a new crowd coordination mechanism that combines machine learning, to predict user expertise, with resource allocation techniques, to solve the aforementioned people-to-task matching problem.

II- Publication(s) during your fellowship

Please insert the title(s), author(s) and abstract(s) of the published paper(s). You may also mention the paper(s) which were prepared during your fellowship period and are under reviewing.

Title: Collective Intelligence based Mechanism for the Coordination and Harnessing of the Corporate Crowd's Innovation Potential
Authors: Lykourentzou I., Vergados D. J., Napoli A. (2011)
Venue: Collaborative Innovation Networks Conference (COINs2011), Basel, Switzerland, September 8-10 COINs
Abstract : In this paper we propose a novel, Collective Intelligence–based mechanism to coordinate the collaboration of the enterprise community. Experimental results show that, compared to the typically observed self-coordination pattern, the proposed mechanism can help the community better allocate its intangible skill resources and thus achieve better knowledge production and innovation results, in terms of quality and timelines.

An extension of the above work was invited and accepted for publication in the special issue on COINs at the journal International Journal of Organisational Design and Engineering, as follows:

Title: Collective intelligence-based resource allocation to optimise knowledge and innovation harnessing in corporate environments
Authors: Lykourentzou I., Vergados D. J., Napoli A. (2011)
Venue: Int. J. Organisational Design and Engineering (accepted for publication)
Abstract: A typical problem that large enterprises face is how to effectively harness the intangible knowledge, expertise, skills and lessons learnt of their staff members. Intangible knowledge harnessing and its codification into usable document formats are vital for the enterprise since they directly affect its ability to innovate and solve complex new problems. However, the identification of the individuals whose contribution can solve each knowledge-demanding problem may be difficult. In this paper, we propose a novel mechanism that uses the collective intelligence of the corporate crowd to identify the tacit knowledge competencies of each employee and coordinate their contributions, inside a wiki-like system, so that each individual may contribute in the most efficient way. Experimental results show that, compared to the fully self-coordinated pattern used by current collaborative knowledge harnessing approaches, the proposed mechanism can help the corporate community allocate its intangible skill resources more efficiently, and thus produce more qualitative knowledge in a timelier manner.

Authors: Codocedo V., Lykourentzou I., Napoli A.
Title: Semantic querying of data guided by Formal Concept Analysis
Venue: FCA4AI Workshop at ECAI European Conference on Artificial Intelligence (under review)
Abstract. In this paper we present a novel approach to handle querying over a concept lattice of documents and annotations. We focus on the problem of “nonmatching documents”, which are those that, despite being semantically relevant to the user query, do not contain the query’s elements and hence cannot be retrieved by typical string matching approaches. In order to find

these documents, we modify the initial user query using the concept lattice as a guide. We achieve this by identifying in the lattice a formal concept that represents the user query and then by finding potentially relevant concepts, identified as such through the proposed notion of close concepts. Finally, we use a concept semantic similarity metric to order and present retrieved documents. The main contribution of this paper is the introduction of the notion of close concepts of a given formal concept followed by a discussion on how this notion is useful for lattice-based information indexing and retrieval.

Authors: Codocedo V., Lykourentzou I., Napoli A.
Title: Semantic indexing and retrieval based on FCA: Application to song datasets
Venue: CLA2012 - The Ninth International Conference on Concept Lattices and Their Applications (under review)
Abstract: Semantic indexing and retrieval has become an important research area, as the available amount of information on the Web is growing more and more. In this paper, we introduce an original approach to semantic indexing and retrieval based on Formal Concept Analysis. The concept lattice is used as a semantic index and we propose an original algorithm for traversing the lattice and answering user queries. This framework has been used and evaluated on song datasets.

III -Attended Seminars, Workshops, and Conferences

Please identify the name(s), date(s) and place(s) of the events in which you participated during your fellowship period.

Event 1
Name: Collaborative Innovation Networks Conference (COINs2011)
Dates: September 8-10, 2011
Place: Basel, Switzerland

Event 2
Name: Invited talk. Subject: "Tacit knowledge harnessing in corporate environments through Collective Intelligence, Machine Learning and Resource Allocation Techniques"
Date: February 16, 2012
Place: Xerox Research centre Europe, Grenoble, France

IV – Research Exchange Programme (12 month scheme)

Please identify the name(s), date(s) and place(s) of your Research Exchanges during your fellowship period and detail them.

Not applicable. The fellowship was a 9+9 month scheme.