# Semantic (Group Formation)
*PhD Research Proposal*[*]

Asma Ounnas[†]

School of Electronics and Computer Science
University of Southampton, UK
ao05r@ecs.soton.ac.uk

## 1    Motivation

For decades, group formation has been a subject of study in many domains. In learning, teachers form groups of students for different types of collaborative activities. For the formation to be efficient, teachers need take into account any constraints that can influence the performance of the group as a whole and that of the individuals within the group, such as students' previous experience, gender, nationality, and interests. The formation of groups in this context involves the creation of balanced groups in terms of expected performance in addition to maximizing each individual's goal from the collaboration. As the number of formation's constraints grows, forming groups that satisfy these constraints increases in complexity. We know that the Semantic Web (SW) aims at providing a promising foundation for enriching resources with well defined meanings and making them understandable for programs and applications. The potential of the SW in this context has allowed the semantic formation of social networks to be successful [1]. From this point, we trust that the problem of constraint group formation can as well be solved using SW technologies. The question is how to apply the SW vision to the problem, and take the most of its potential to apply it in real life applications such as e-learning. In particular, the problem can be formulated as how can we generate optimal groups by reasoning over possibly incomplete data about the students.

## 2    Research Overview and Essential Questions

Since forming groups of students with attention to constraint satisfaction is not a simple task for the teacher to do manually, especially for a large number of students, the proposed research is intended to investigate the automation of constrained group formation. In order to cover different types of collaborative activities, we consider the formation of different types of groups including: Teams, Communities of Practice (CoPs), and Social Networks (SNs). We believe that by reasoning on learners' profiles and the teacher's constraints, we can achieve a powerful foundation for automated group formation. With respect to SW concepts, our present and future work intends to give appropriate answers to the following questions: What do we model for the formations of different types of groups? How do we enable the teacher to get the group formation they want? How do we achieve that formation? And how effectively we achieved it? Due to their self-organized nature, for formation of CoPs and SNs to be effective, the instructor has to provide a degree of dynamic self organization within these groups. In this research we address the question of how do we enable the dynamic formation of instructor-initiated CoPs and SNs? If we do not have all the required information about the users, how do we process the formation with incomplete data? Can we find this data or similar data and substitute it to maintain the robustness of the grouping? Where can we get this data, and what type of data should it be? If we substitute the data, how significant is the measurement of the correlations between the required data and the alternative one?

## 3    Research Methodology

---

[*] Supervisors: Dr. Hugh C Davis and Dr. Dave E Millard, School of Electronics and Computer Science, University, Southampton, Emails: {hcd, dem}@ecs.soton.ac.uk

[†] Second Year of PhD research, approximate defense time is January 2009.

To answer the research questions and examine the soundness of the assumed hypotheses, we aim at building a Semantic Web based system that allows the instructor to automatically form different types of groups. The formation of the groups generated by the system will then be evaluated based on the quality of the generated groups, and the robustness of the formation in case of incomplete data.

**1. Research Implementation:** The system will have three main components:

*1. The Ontology*: called Semantic Learner Profile (SLP), the ontology is an extension of the FOAF vocabulary that aims at providing semantic data about the learner [2] for the formation of all types of groups. Each student has an extended foaf file that can be updated at any time. This allows them to publish data about themselves using a URI, which enables the data to be referred to from any dataset. An interface based on foaf-a-matic (http://www.ldodds.com/foaf/foaf-a-matic.html) will be provided to facilitate the creation of these profiles. Since FAOF allows the users to define their friends, social connections can be made for CoPs and SNs formation. As the students can modify their friends' list at any time, the relationships links between them allow a dynamic formation which provides the groups with a degree of self-organisation.

*2. The Instructor Interface:* The teachers will be allowed to choose the constraints they want to base the formation on. They will be provided by an option that enables them to set constraints on those values and the relationship between those values. The interface will also enable the instructor to rank the importance of these constraints to enable the system to manage compromises based on these priorities.

*3. The group generator:* The group generator will be supported by a set of rules that represent different formation algorithms that allows reasoning on the data provided by the learners and the teacher in order to generate effective groups. The system will be empowered by Jena inference engine and SAPRQL for querying over the data. To allow an effective grouping, students are to be encouraged to create meaningful descriptions of themselves with as much details as they can. In case they do not provide all required data for a formation, the instructor will be supported by an option that enables the system to use Semantic Web mining techniques to look for the missing data in the web and form correlations to the required data. Moreover, we need to address the data provenance, especially if it is extracted from blogs and web pages.

**2. Research Evaluation**: To evaluate the system and hence the research hypotheses, we intend to test the system on real life data by forming groups of students taking a software engineering course (SEG) in the University of Southampton. To ensure the system is tested for different groupings we also use randomly generated data, and a simulated population of students. For this, a person generator is created. The efficiency of the system will be measured based on the quality of the formation provided by the system which involves: to what degree did each generated group meet satisfied the constraints, how many groups satisfied the constraints, and what is the systems confidence in generating successful grouping. The same measures will be applied to evaluate the system's capability to form groups with incomplete data.

## 4    Current Status

So far, we implemented the SLP ontology, and the random person generator. Both the student interface and simulated data are currently under development. To support the creation of the simulated data and prepare for the evaluation of the semantic formation on the real life data next year, we are currently running an observational study based on two questionnaires one to get information about the student, and the other to evaluate the group formation. The questionnaires are given to the students taking the SEG course this year who have already been grouped manually by the teacher based on their previous grades and gender. This observational pre-study will enable us to compare the results of this manual formation with the automated semantic formation, which is intended to run as a controlled study on the same course next year. Moreover, the pre-study will help in getting information about the students' population for the creation of the simulated data. For our future work, the core components of the semantic formation system are to be implemented so that the hypothesis of the research can be evaluated. Future work will include more research on managing group formation with incomplete data.

## 5    References

1. Golbeck, J., Parsia, B. & Hendler, J, Trust Networks on the Semantic Web, Proc. of CIA. Helsinki, Finland, 2003.
2. Ounnas, A., Davis, H. C. and Millard, D. E. (2007) Semantic Modeling for Group Formation. In Proceedings of PING workshop at the UM2007, Corfu, Greece.