

Report for the Short Term Scientific Mission

Dr. Elton Domnori

Purpose of the STSM

The scope of the STSM was to analyze possible solution to the assignment problem presented in the work [1,2]. Current, the solution has been adopted by providing a modified version of the Munkres Algorithm (a.k.a. Hungarian Algorithm) but this solution requires optimization as its computation grows exponentially with the increase of the number of result generated.

Description of the work carried out during the STSM

The mission has been conducted at the premises of University of Zaragoza, Spain. The duration has been of two weeks, from February 26 to March 11. The persons involved in the collaboration are prof. Raquel Trillo Lado from the department of Informatics and System Engineering, prof. María Carmen Calvo Yanguas and prof. Carmen Elvira Donazar from the department of Economic Analysis (Faculty of Economy and Business).

The first phase of the work was focused on the analysis of the assignment problem, the issues presented in the current approach and the constraints presented by the application (such as one user keywords must be mapped with only one schema element and generation of top-k results with monotonic function).

The second phase of the work was the feasibility analysis of the use of “Symmetric Groups” from the “Group theory” [3]. Considering that the problem could be represented as a permutation of a set of K elements (representing the number of keywords) in a group of N elements ((representing the number schema elements), it was decided to use the symmetric groups representing the permutation of at most $K+1$ elements (+1 element for the group closure). The use of symmetric groups comply to the constraint that one user keyword is mapped to only one schema element.

The third phase of the work consisted in the analysis of the different variants of the symmetric group representation that might better fit the generation of permutation (using generator function) that represent solution in monotonic descending order.

Description of the main results obtained

The representations that can better fit the problem criteria of reducing the permutation space of generation are: (a) the Young Subgroups representation that aims to create partitions reducing the space of permutation to generate by defining groups with max $k+1$ elements (at most $k+1$ elements will permute and all other conserve the previous state); and (b) the Standard Tableaux that, starting from a permutation representing the best solution, defines a tabloids that is representative for not only the permutation but as well represent permutations that are have a close distance from the main one.

Both representations are suitable to be processed by Combinatorial Algorithms in order to produce further permutations that produces results close to the best one. The algorithms that we are testing are the “Robinson-Schensted Algorithm” and the “Viennot's Geometric Construction” aim to obtain a sequence of permutations that represent solutions

close to the best one. The evaluation is still in progress. The aim of the future work is to define the permutation generation function that better fits the monotony constraint of the generation of the top-k results.

Future collaboration with the host institution

This mission has started a collaboration between academic staff between of two different universities and in particular the persons directly involved in this mission. The collaboration is extended as well to two other search projects I have been introduced during my work at the university. In particular, the "Recommendation Systems" project led by Dr. Sergio Ilario and the "Process Mining to improve the security of Web Information Systems" project led by Dr. Simona Bernardi. Both of such projects give space to future collaboration due to the common interest in the research field. Moreover, I had the possibility to attend CONPilar2017 conference (<http://conpilar.es/>) on data and service security.

Foreseen publications/articles resulting from the STSM;

At the current stage we have not decided yet about the publication as the work is still in progress and the results have not been finalized yet but we foreseen to have a few conference publication and at least one journal article.

Other comments.

This experience was very fruitful for me and I had the chance to continue a project that has already started and produced good qualitative publications (see [1,2]). I found a very collaborative environment and people and I am confident that this collaboration will give good results.

References

- [1] Sonia Bergamaschi, Elton Domnori, Francesco Guerra, Raquel Trillo Lado, Yannis Velegarakis: "Keyword search over relational databases: a metadata approach.", SIGMOD2011;
- [2] Sonia Bergamaschi, Elton Domnori, Francesco Guerra, Mirko Orsini, Raquel Trillo Lado, Yannis Velegarakis: "Keymantic: Semantic Keyword-based Searching in Data Integration Systems", PVLDB 3(2): 1637-1640 (2010);
- [3] "The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions", Bruce E. Sagan, 2E, Springer