**Postdocs in the area of Social Computing**

**Location:** QCRI, a newly open research institute in Doha, Qatar. QCRI employs researchers interested in different areas of computing and who want to have global impact through publications in top conferences and a local impact by working closely with businesses in Qatar. In particular, QCRI has a tight relationship with Al Jazeera to work on its online data.

**Contacts:** Sihem Amer-Yahia (syahia@qf.org.qa), Principal Research Scientist at QCRI and Marie-Christine Rousset (marie-christine.rousset@imag.fr), Professor at LIG, Grenoble, France.

**Length:** 1 year renewable another year

**Compensation:** Highly competitive with free housing

The development of Web 2.0, that is, the evolution of the Web from a technology platform to a social milieu, has resulted in an unprecedented reliance of users on the Web to assist them in a variety of tasks. Information on the Social Web is massive and is characterized by a combination of factual and opinion data in the form of user-generated tags, ratings and reviews. Users are both providers and consumers of content. Social streams such as Twitter are fundamentally changing the ways content is produced and consumed. Social media are seeing a shift from broadcast news to a network of information and people. There is no one source of information anymore, and devising new methods to determine content relevance based on popularity and reputation, and effectively exploring that content is key to the survival of the Social Web.

As a result, the traditional data management architecture, physical access algorithms, logical algebra and query optimization layer, and user-facing SQL layer, needs to be revisited. We propose a new architecture for gathering, processing and querying social content. The bottom layer is responsible for integrating social data from multiple sources into a virtual social graph of users and content. The middle layer provides advanced social data analytics primitives such as finding similar users based on shared behavior, extracting entities from Facebook posts, or finding hidden topics in user-provided tags. The top layer is a set of user-facing primitives that enable social content exploration.

**Postdoc1: Algebra for Querying Social Analytics**

This postdoc position focuses on the top and middle layers of the proposed social content architecture. The candidate will be responsible of the following tasks:

- Develop a data model for querying the results of social analytics

- Design a set of composable primitives to query social data (e.g., most influential users, most prominent discussion topic in a given community, etc)

- Devise novel algebraic optimizations

The candidate must have strong modeling skills and a good knowledge of the relational algebra and calculus and relational query optimization.

Knowledge of PigLatin as well as Perl/Python are a plus.

**Postdoc2: Social Exploration on MapReduce**

This postdoc position focuses on the middle and bottom layers of the proposed social content architecture. The candidate will be responsible of the following tasks:

- Develop a data model for social analytics

- Design a set of composable primitives for social analytics (e.g., similar users, hidden topics, community opinion on a topic, etc)

- Devise parallel algorithms for the implementation of social analytics primitives on top of MapReduce

The candidate must have strong modeling and programming skills.

Knowledge of Hadoop/MapReduce and Pig Latin as well as Perl/Python are a plus.

**References:**

1. Sihem Amer-Yahia, Laks V. S. Lakshmanan, Cong Yu: SocialScope: Enabling Information Discovery on Social Content Sites. CIDR 2009

2. Alan Gates, Olga Natkovich, Shubham Chopra, Pradeep Kamath, Shravan Narayanam, Christopher Olston, Benjamin Reed, Santhosh Srinivasan, Utkarsh Srivastava: Building a HighLevel Dataflow System on top of MapReduce: The Pig Experience. PVLDB 2(2): 1414-1425 (2009)

3. Tomasz Nykiel, Michalis Potamias, Chaitanya Mishra, George Kollios, Nick Koudas: MRShare: Sharing Across Multiple Queries in MapReduce. PVLDB 3(1): 494-505 (2010)

4. Alexander J. Smola, Shravan Narayanamurthy: An Architecture for Parallel Topic Models. PVLDB 3(1): 703-710 (2010)