

Data management in social networks

Mots clés :

- **Directeur de thèse** : JEAN-LOUIS DESSALLES
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire Traitement et Communication de l'Information
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

Résumé du projet de recherche (Langue 1)

Advisors: **[[<http://www.infres.enst.fr/~talel> | Talel Abdesslem]] and [[<http://perso.telecom-paristech.fr/~cautis> | Bogdan Cautis]].** (Assistant Professors, Telecom ParisTech) === Description: We are witnessing today the rapid emergence of large online communities for sharing content on the Web. Among them, social tagging platforms (such as Flickr, Del.icio.us, CiteUlike) and collaborative environments (Wikipedia, Knol) offer simple yet highly effective mechanisms for sharing data and contributing meta-data. They are essentially oriented towards building social networks and large repositories of quality user-generated content, where connectivity of information and structured knowledge representation are keys to simplifying access to data. With their rapid growth in data volumes and user base, social platforms raise new challenges in many aspects of data management such as search or controlling access to resources. They call for new approaches, that are user-centred, socially driven and have as key enablers models of trust/distrust, reputation and connectivity. The main goal of this thesis is to study the interplay of data management and connectivity/trust in open contexts such as collaborative environments and social tagging systems. We detail next two main facets of this project. === Access control in open communities. Spam and adversarial interactions are increasingly affecting social media. Moreover, privacy issues are often raised, e.g., in photo-sharing or blogging. Beyond whitelists / blacklists, we believe that a different approach for access control is necessary, one in which rights are related to user inter-connections - be it explicit (the social network) or implicit (via relations revealed by related activities, resources, tags - user reputation and utility. In such open communities, this naturally translates into measures of trust and distrust. We believe that in many social applications (e.g., tagging systems) these notions should replace the traditional access-control lists as the main criteria for granting or revoking access. Towards this goal, we will investigate mechanisms by which trust (or distrust) relations between users can be established based on user activities, thematic proximity, reputation and peer evaluation, social links, and so on. === Search. Social applications are the fastest growing segment of the web yet, to date, social media has been primarily popular for connecting people, not for finding information. In social applications, users can not only access content but they can also generate, share and modify content. Different notions of user information, such as preferences (community memberships, social interactions), context information (e.g., user's social network, location, time) and other features of a user's environment, are thus of paramount importance in order to improve and personalize user experience in accessing information. We intend to investigate models that can capture the user needs and are specific to social media search, based on confidence, authority and accuracy. Traditional one-size-fits-all techniques for information retrieval cannot scale when shifting to a social/user centric approach. This is why efficiency and scalability become key challenges, requiring carefully designed, highly adaptable data structures and algorithms. We see the topic of the thesis as being at the intersection of data management and information retrieval, with a strong data mining component. The thesis requires a strong theoretical background (algorithm design, data structures, complexity) and strong programming skills (especially Java).