

TF15(WG2):Web data coll. from non-reactive sources

Posted by fornaran - 26 Oct 2012 12:18

TF15(WG2):Web data coll. from non-reactive sources

Task Force leader: Nicoletta Fornara

Members: Task force is open to all people interested in automatic extraction of non reactive data.

Aim: Define a case study (for example a source of data and a set of rules) from which we can start to investigate how to design a software tool able to get as input a declarative list of rules that describe how data can be used and how data can be collected. This tool should be able to automatically collect data by taking into account such rules.

Planning: One session (1:45 min.) in Ljubljana.

=====

Re: TF15(WG2):Web data coll. from non-reactive sources

Posted by ulf - 10 Nov 2012 10:37

Thank you, Nicoletta, and congratulations re your project grant!

=====

Re: TF15(WG2):Web data coll. from non-reactive sources

Posted by Ela Oren - 07 Mar 2013 15:43

Hi,

My name is Ela Oren, I am very interested in non reactive data analysis. I research the usage of language in blogs, forums, twitter profiles and videos dedicated to obsessive compulsive disorder (OCD).

My main questions are:

1. What is the efficient way to find and gather trustworthy information?
2. Is the challenge of multi-language research applicable when talking about forums, blogs, twitter profiles etc.?
3. How can I make quantitative and qualitative comparisons between types of speech?
4. Is it legal and ethical to use data taken from these kinds of websites?

I'll be happy to join this TF, if possible. Will you be meeting in the near future?

Thanks,

Ela 

=====

Re: TF15(WG2):Web data coll. from non-reactive sources

Posted by rmtorres - 15 Mar 2013 09:01

Hi,

My name is Rocío Martínez Torres. I am an Associate Professor in the department of Business and Management at the University of Seville (Spain). I am researching on topics related to the extraction of data from the web and their analysis through the use of Social Network Analysis and Semantic Analysis techniques (e-social sciences) for many years. These topics could be summarized in the next research lines:

1. **Virtual communities:** This research line is focused on analyzing the patterns of behaviour of virtual/online community users by modelling their interactions as a social network. The topological features of users participation (degree, centrality, clustering coefficient, ...) are extracted using Social Network Analysis techniques, while the content of the interchanged information is analysed using semantic techniques like Latent Semantic Indexing. The aim consists of obtaining the user profiles attending to their participation features, the structure of the community, the ability of the community to attract new users that can become experts, the main topics of discussions within the community and the ability of the community to create and share knowledge.
2. **Open Source Software communities:** The previous techniques were used in the analysis of open source software communities. They represent a clear example of collective intelligence paradigm. Data are extracted from the information contained in the distribution lists, which are publicly available from the project website. A specific crawler was designed for this task. The features of participation and of shared knowledge are processed through computational intelligence techniques (Genetic Algorithms, Particle

Swarm Optimization) to obtain the structure of the communities and the different users' profiles.

3. Open innovation communities: This is another emerging paradigm in innovation, based on the idea that organizations should also rely on ideas and knowledge developed externally. One of the most popular schemes for open innovation implementation is through open innovation communities, where users can post, share, comment and evaluate ideas using a specific website managed by an organization (i.e. Dell IdeaStorm, My Starbucks Idea,...). Users can interact among them and also with the innovation departments and experts of the organization. Again, shared ideas, interactions among users and the evaluations and scores received by posted ideas can be extracted and processed from the open innovation website. My main words around this topic consisted of identifying lead users within innovation communities as an alternative to collective scoring methods.

4. Website structure analysis through link analysis: the hyperlink structure of websites can be also modelled as a social network, extracting the data through the use of a crawler that explore the complete website looking for links to other internal and external web pages. My work in this research line was focused on analyzing the accessibility and navigability of institutional websites, classifying them according to several base structures. Moreover, we have also developed the concept of local visibility of websites as an alternative to the idea global visibility used by search engines.

5. Knowledge networks: Knowledge networks of Higher Education institutons in the UK were analyzed using the ISI citation index databases. This was a joint work with Prof. Nik Bessis from the University of Derby, UK, as a result of my stay at this University for three months. By analyzing the joint research papers, it is possible to obtain the collaborations among Universities and their internationalization policies.

I think al these research lines are related to the activity of this TF15: Web data collection from non-reactive sources.

I think we could work together in the advance of this TF.

=====

Re: TF15(WG2):Web data coll. from non-reactive sources

Posted by Stephanie - 17 Sep 2013 11:34

Hi Nicoletta,

As discussed please find below the link to the group which could be of interest to you.

In case you would like to come and present and meet them, please let me know. Maybe we could arrange it in the framework of a STSM.

ehumanities.nl/

www.eur.nl/erasmusstudio/

About Ethics (as a starting point as this group is no longer active)

ereseach-ethics.org/resources/

These are resources from the Oxford Internet Institute

www.oii.ox.ac.uk/publications/?type=&keywords=ethics

Best

Stephanie

=====