The University Of Sheffield.

Extracting Semantic User Networks From Informal Communication Exchanges

A.L Gentile V.Lanfranchi S.Mazumdar F.Ciravegna

OAK Group
Department of Computer Science
University of Sheffield

Organisations, Information and Knowledge

The University Of Sheffield.

Introduction

- Exploit Organisational Knowledge that is often **buried**
- Generate Semantic Profiles
- Application

  Organisational Knowledge Management Context

Organisations, Information and Knowledge

Informal Communication

- Emails
- Meeting Requests
- Meeting Records
- Chats

Organisations, Information and Knowledge

Informal Communication

From:13
To:18

I thought the first prize would have been a copy of the Guardian or a pair of sandals...

From:13
To:3

Potential Xmas buffet?

Interesting blog post from Mischa Tufffield (of Garlik) here about tracking web users passively through Facebook's like button:

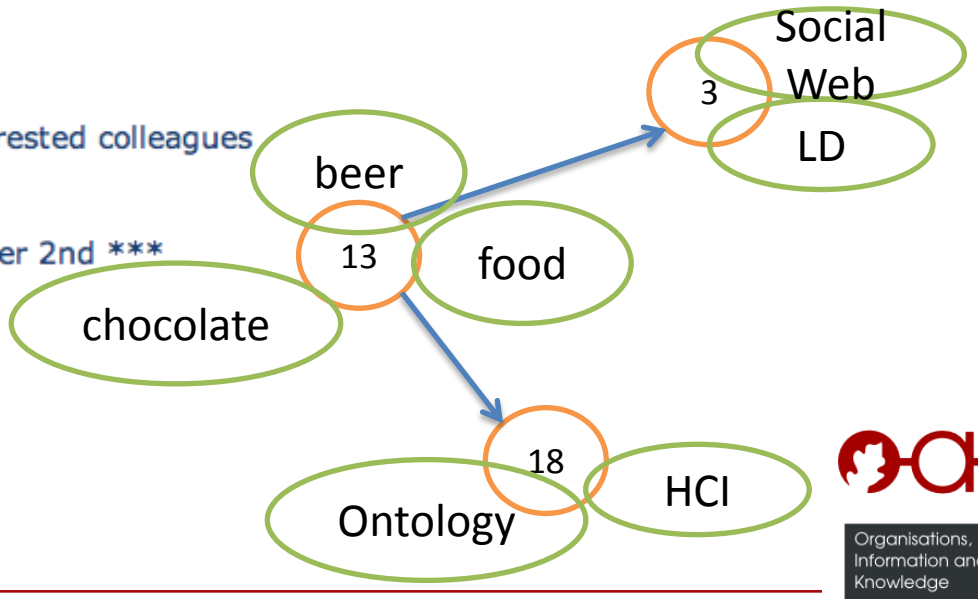
http://mmt.me.uk/blog/2010/07/30/the-facebook-like-button/

From:13
To:18

Apologies for cross-postings. Please forward to interested colleagues and mailing lists.

*** Submissions deadline extended until September 2nd ***

CALL FOR PAPERS: SDoW2010



Organisations, Information and Knowledge

The University Of Sheffield.

Approach

Collect Features

Generate User Profiles

User Profile Usage

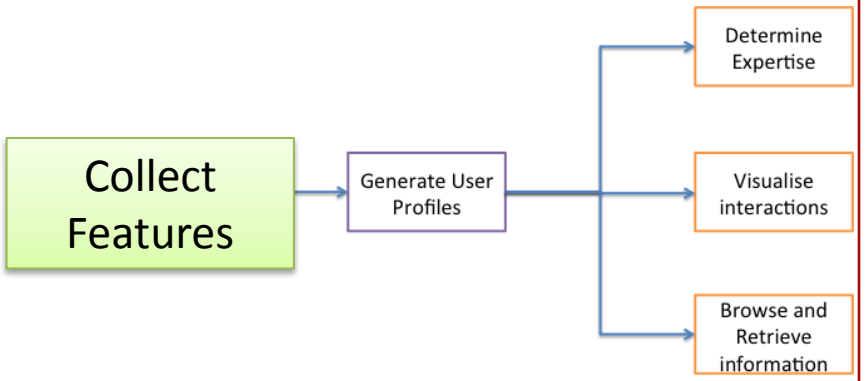
- Determine Expertise
- Visualise interactions
- Browse and Retrieve information

Organisations, Information and Knowledge

The University Of Sheffield.

State of the Art

Collect Features from Emails



Exchange Frequency

- Absolute frequency thresholds (Tyler et al. 2005)
- Time-dependent thresholds (Cortes et al. 2003)

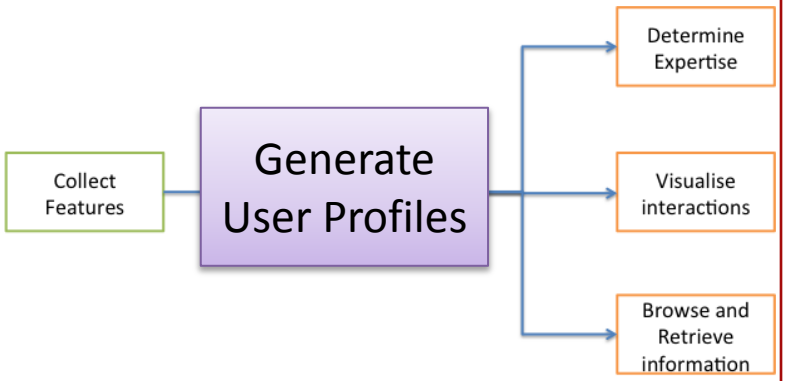
Content-Based Analysis

- Determine expertise (Schwartz and Wood, 1993)
- Analyse relations between content and people (Campbell et. al., 2003)
- Extract personal information (names, addresses, contacts) (Laclavik et. al., 2011)

oak Organisations, Information and Knowledge

State of the Art

Generate User Profiles



Monitoring User activities on the web

(Kramar,2011)

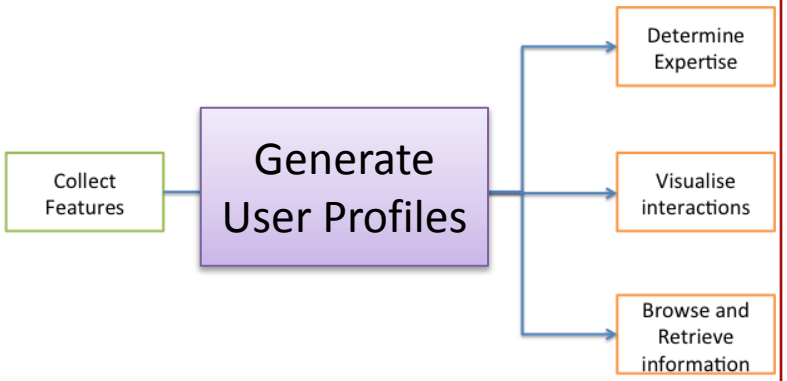
Analysing user generated content (Tweets)

(Abel et. al., 2011a)

The University Of Sheffield.

State of the Art

Measures for User Similarity



Binary Function (are the two users connected?)

Non Binary Function (how strong is their connection?)

Features typically exploited

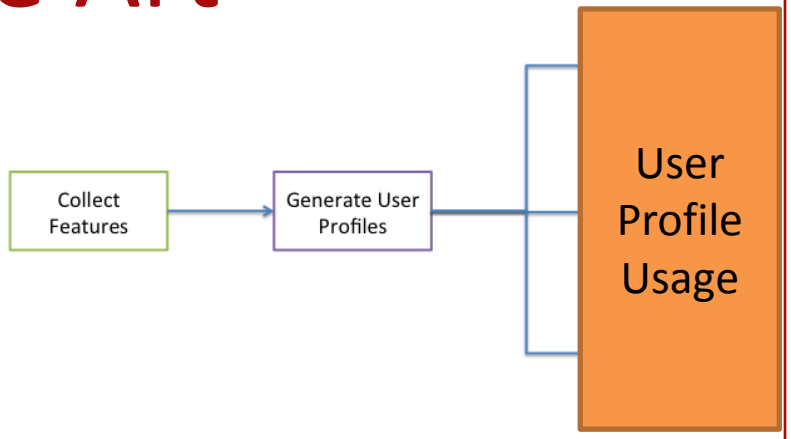
- geographical location, age, interests, social connections
- Facebook friends, interactions, pictures

oak
Organisations, Information and Knowledge

The University Of Sheffield.

State of the Art

User Profile Usage



Information Retrieval – Customised search results

(Daoud et. al., 2010)

Recommender Systems - Effective customised suggestions

(Abel et. al., 2011b)

Organisations, Information and Knowledge

The University Of Sheffield.

Research Question

Does increasing the level of semantics in user profiles outperform current methods?

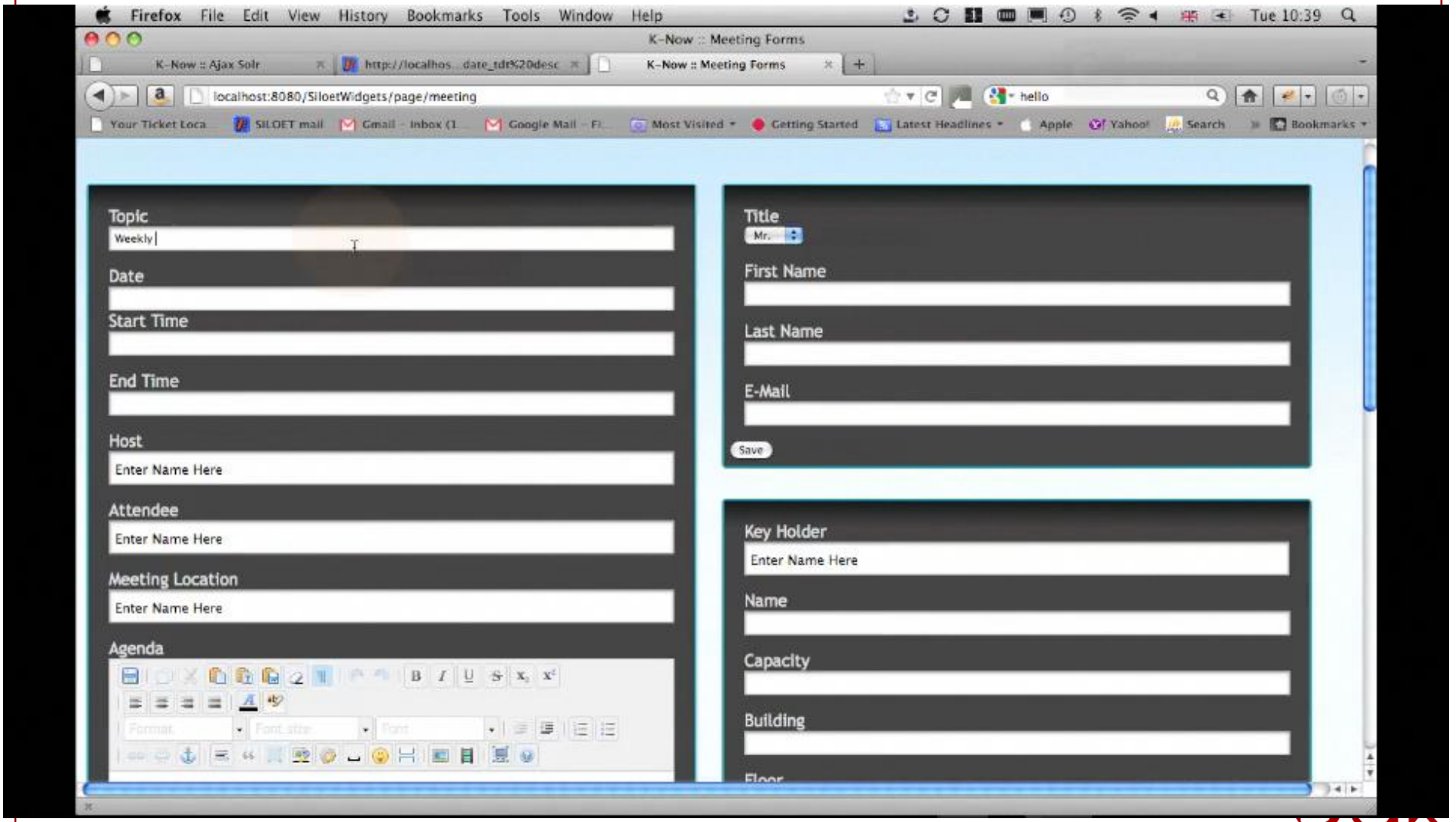
Task: Inferring similarity among users

Assessment: Correlation with human judgement

Organisations, Information and Knowledge

The University Of Sheffield.

Capture Information



Organisations, Information and Knowledge

The University Of Sheffield.

Experiment Settings

Corpus

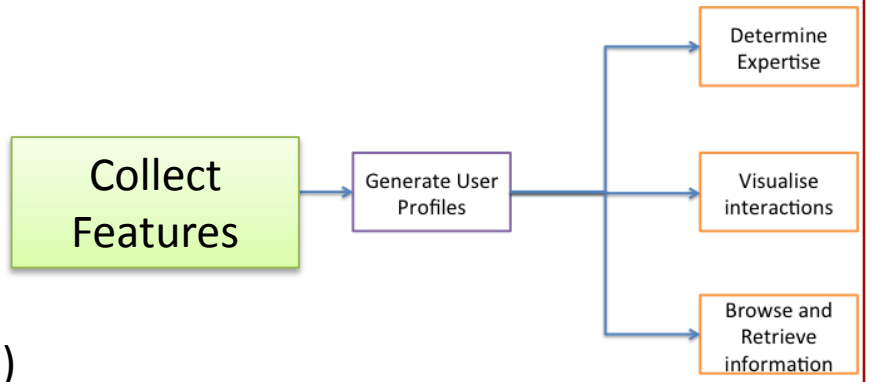
Internal mailing list of the OAK group in the Computer Science Department of the University of Sheffield

- 1001 emails
- Users in mailing list : 40
- Active users (sending emails to list) : 25
- Users participating in the evaluation: 15

Organisations, Information and Knowledge

Collect Features

For each email e_i in the collection E



Keywords (Java Automatic Term Recognition)

Bag of keywords representation: $e_i = \{k_1,\ldots,k_n\}$

Named Entities (Open Calais web service)

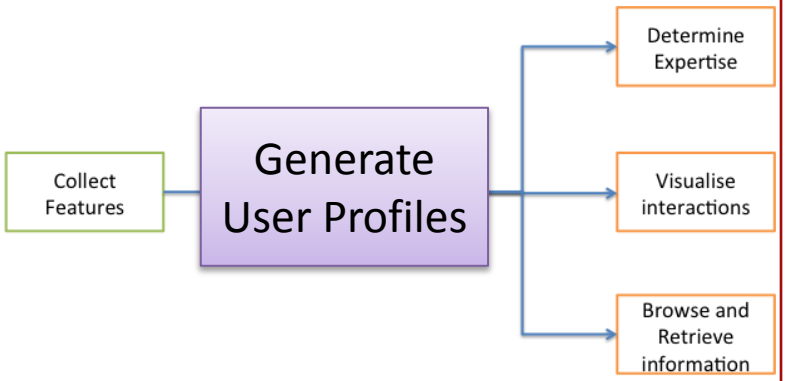
Bag of Entities representation: $e_i = \{ne_1,\ldots,ne_n\}$

Concepts (Wikify, Milne and Witten, 2008)

Bag of Concepts representation: $e_i = \{c_1,\ldots,c_n\}$

The University Of Sheffield.

Generate User Profiles



Amount of knowledge shared among individuals

(Keywords, Entities, Concepts)

Similarity strength on a [0,1] range

$$Jaccard(P_1, P_2) = \frac{P_1 \cap P_2}{P_1 \cup P_2}$$

Sample sets for P_1, P_2: Keywords, Named entities or Concepts

Organisations, Information and Knowledge

The University Of Sheffield.

Evaluation

- Participants were asked their perceived similarity with colleagues
  - Professional and social point of view
  - Topics of interest
- Similarity on a scale of 1 to 10

  1 – Not similar at all

  10 – Very similar

Organisations, Information and Knowledge

The University Of Sheffield.

Evaluation

- Compare user's perceived similarity with achieved similarity

Pearson's correlation

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

σ_{xy} - Covariance of X and Y (how much they change together)

σ_x, σ_y - Standard deviation for X and Y (how much variation from the average)

Organisations, Information and Knowledge

The University Of Sheffield.

Results

User ID	Correlation Keyword	Correlation Entity	Correlation Concept	Inter-Annotator Agreement
14	0.55	0.41	0.68	0.91
7	0.48	0.39	0.58	0.87
28	0.5	0.41	0.57	0.89
10	0.47	0.39	0.57	0.94
27	0.32	0.29	0.48	0.92
21	0.34	0.42	0.42	0.91
1	0.35	0.32	0.42	0.94
3	0.3	0.31	0.38	0.86
9	0.28	0.36	0.38	0.9
18	0.5	0.5	0.36	0.87
8	0.17	0.19	0.35	0.82
11	0.59	0.42	0.34	0.83
25	0.25	0.33	0.3	0.73
23	0.21	0.33	0.19	0.86

Organisations, Information and Knowledge

The University Of Sheffield.

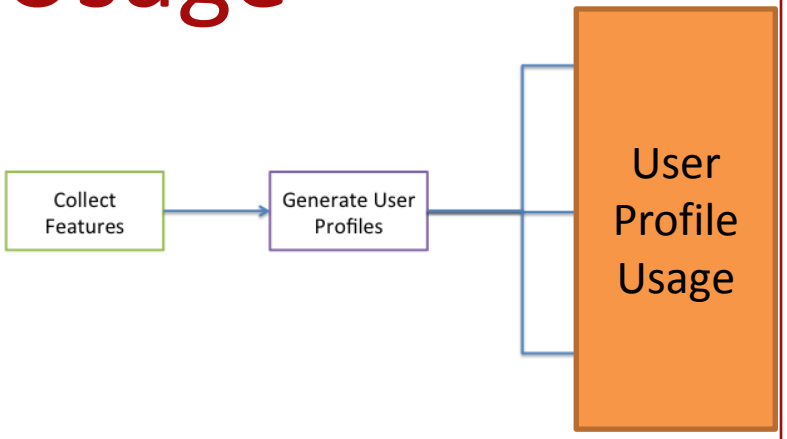
Results

Keyword (Avg)	Named Entities (Avg)	Concepts (Avg)
0.379	0.362	**0.430**

Organisations, Information and Knowledge

The University Of Sheffield.

User Profile Usage



- Email Browsing

  Topics of communication

  User expertise

- Email Retrieval

  Perform specific queries

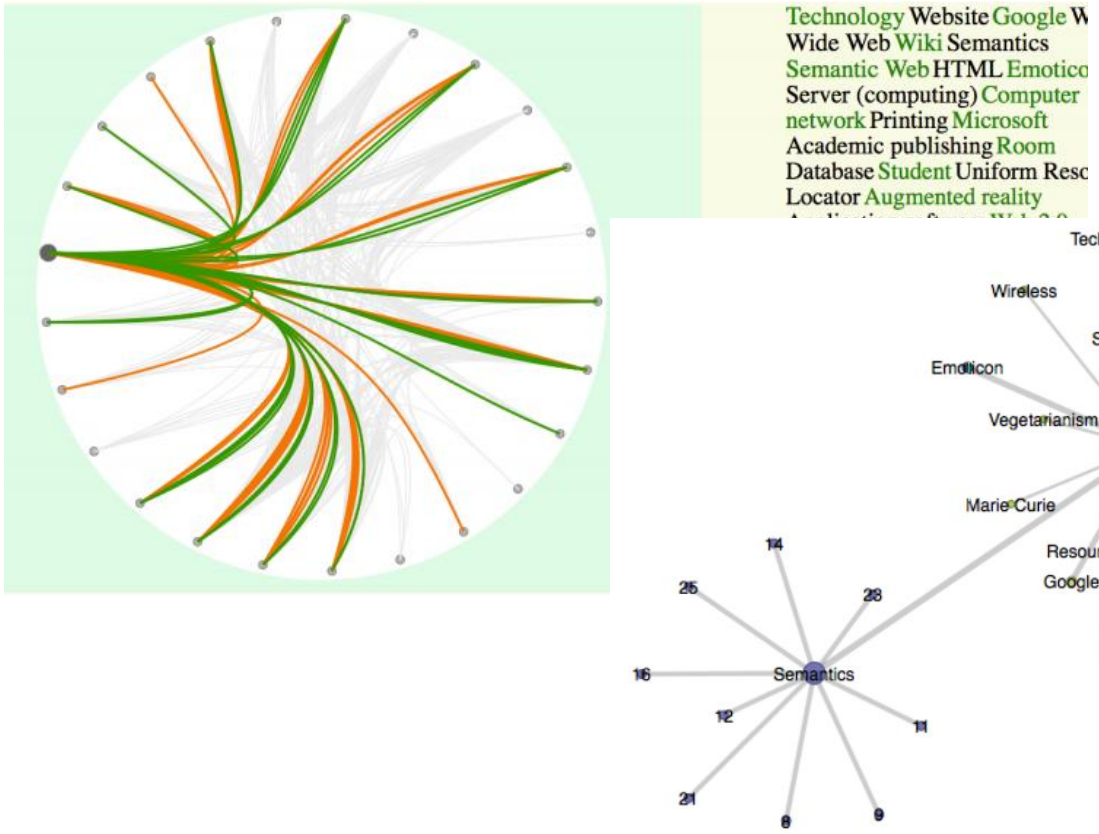
  Selecting individuals

- Email Visualisations

  Investigate interaction networks

Organisations, Information and Knowledge

SimNET – Exploring Interaction Networks



Organisations, Information and Knowledge

The University Of Sheffield.

SimNET

The University Of Sheffield.

Conclusions

- Dynamically model user expertise from informal communication exchanges
- Generate semantic user profiles from textual content, generated by users
- Making use of buried knowledge within an organisation

Organisations, Information and Knowledge

The University Of Sheffield.

Future Directions

- Long term trials of the system in an organisation with ‘knowledge workers’
- Explore new visualisations to facilitate real time visualisation of dynamic networks and profiles
- Connect user profiles to Linked Open Data
  - Investigate how profiles can be further enriched using Linked Data

Organisations, Information and Knowledge

The University Of Sheffield.

Reference

- Abel, F., Gao, Q., Houben, G.-J. and Tao, K. (2011a). Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. In ESWC (2), (Antoniou, G., Grobelnik, M., Simperl, E. P. B., Parsia, B., Plexousakis, D., Leenheer, P. D. and Pan, J. Z., eds), vol. 6644, of Lecture Notes in Computer Science pp. 375–389, Springer.
- Abel, F., Gao, Q., Houben, G.-J. and Tao, K. (2011b). Analyzing User Modeling on Twitter for Personalized News Recommendations. In User Modeling, Adaption and Personalization, (Konstan, J., Conejo, R., Marzo, J. and Oliver, N., eds), vol. 6787, of Lecture Notes in Computer Science pp. 1–12. Springer.
- Adamic, L. and Adar, E. (2005). How to search a social network. Social Networks 27, 187–203.
- Campbell, C. S., Maglio, P. P., Cozzi, A. and Dom, B. (2003). Expertise identification using email communications. In Proceedings of the twelfth international conference on Information and knowledge management CIKM '03 pp. 528–531, ACM, New York, NY, USA.
- Cortes, C., Pregibon, D. and Volinsky, C. (2003). Computational methods for dynamic graphs. Journal Of Computational And Graphical Statistics 12, 950–970.
- Daoud, M., Tamine, L. and Boughanem, M. (2010). A Personalized Graph-Based Document Ranking Model Using a Semantic User Profile. In User Modeling, Adaptation, and Personalization, (De Bra, P., Kobsa, A. and Chin, D., eds), vol. 6075, of Lecture Notes in Computer Science chapter 17, pp. 171–182. Springer.
- De Choudhury, M., Mason, W. A., Hofman, J. M. and Watts, D. J. (2010). Inferring relevant social networks from interpersonal communication. In Proceedings of the 19th international conference on World wide web WWW '10 pp. 301–310, ACM, New York, NY, USA.
- Eckmann, J., Moses, E. and Sergi, D. (2004). Entropy of dialogues creates coherent structures in e-mail traffic. Proceedings of the National Academy of Sciences of the United States of America 101, 14333–14337.
- Keila, P. S. and Skillicorn, D. B. (2005). Structure in the Enron Email Dataset. Computational & Mathematical Organization Theory 11, 183–199.
- Kossinets, G. and Watts, D. J. (2006). Empirical Analysis of an Evolving Social Network. Science 311, 88–90.
- Kramar, T. (2011). Towards Contextual Search: Social Networks, Short Contexts and Multiple Personas. In User Modeling, Adaption and Personalization, (Konstan, J., Conejo, R., Marzo, J. and Oliver, N., eds), vol. 6787, of Lecture Notes in Computer Science pp. 434–437. Springer.
- Laclavik, M., Dlugolinsky, S., Seleng, M., Kvassay, M., Gatial, E., Balogh, Z. and Hluchy, L. (2011). Email analysis and Information Extraction for Enterprise benefit. Computing and Informatics, Special Issue on Business Collaboration Support for micro, small, and medium-sized Enterprises 30, 57–87.
- McCallum, A., Wang, X. and Corrada-Emmanuel, A. (2007). Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email. Journal of Artificial Intelligence Research 30, 249–272. Milne, D. and Witten, I. H. (2008)

Organisations, Information and Knowledge

The University Of Sheffield.

Reference

- Milne, D. and Witten, I. H. (2008). Learning to link with wikipedia. In Proceeding of the 17th ACM conference on Information and knowledge management CIKM ’08 pp. 509–518, ACM, New York, NY, USA.
- Schwartz, M. F. and Wood, D. C. M. (1993). Discovering shared interests using graph analysis. Communications of the ACM 36, 78–89.
- Tyler, J., Wilkinson, D. and Huberman, B. (2005). E-Mail as Spectroscopy: Automated Discovery of Community Structure within Organizations. The Information Society 21, 143–153.
- Zhou, Y., Fleischmann, K. R. and Wallace, W. A. (2010). Automatic Text Analysis of Values in the Enron Email Dataset: Clustering a Social Network Using the Value Patterns of Actors. In HICSS 2010: Proc., 43rd Annual Hawaii International Conference on System Sciences pp. 1–10,.

Organisations, Information and Knowledge

The University Of Sheffield.

Acknowledgements

A.L Gentile and V. Lanfranchi are funded by SILOET (Strategic Investment in Low Carbon Engine Technology), a TSB-funded project. S. Mazumdar is funded by Samulet (Strategic Affordable Manufacturing in the UK through Leading Environmental Technologies), a project partially supported by TSB and from the Engineering and Physical Sciences Research Council

Organisations, Information and Knowledge