

Automatic association of personal documents on the cloud

Jimmy Mian-Guan Lim and Frank Wang

Future Computing Research Group,

School of Computing, University of Kent, Kent CT2 7NT, UK.

Tel: +44 (0)1227 82 3819, email: m.g.lim@kent.ac.uk, f.z.wang@kent.ac.uk

The management of the data collected over a lifetime is a growing problem. This problem is outlined by the growth of files in our personal inventory, which causes increasing difficulty to find, retrieve and organise files. Current research is focusing on cloud storage and how to tap into the benefits of having unlimited space [1]. The ever-expanding storage availability allows individuals to store more data and consider less on which data is important. This also encourages less important personal data to be kept and degrades the performance on retrieval of this data.

There have been efforts to speed up retrieval by placing text search on desktops [2][3], however the technology has not been able to deviate from text matching methods. Text matching methods are part of content-based classification methods, which uses word matching. The alternative method, context-based classification uses surrounding information to identify a single object or file. This can range from time, place or even access of the file by individuals.

As part of an EPSRC¹ sponsored project, we move into investigate the automatic association of files by user behaviour. User behaviour is defined by how users interact with desktop applications and files through mouse motions.

To evaluate the relationships built on by mouse actions and time, we constructed a windows based prototype tracking system and relationship builder FRUBO (File Relationship User Behaviour Ontology) that tracks the mouse actions passively. The mouse actions are then translated into series of time events and positions to build relationship R between a file or application measured when the user switches between them. The results show that the system is capable of grouping files with an increment of up to 5 documents per searched document based on user mouse motion over a period of time without using text-matching methods. Fig.1 shows the new relationships of files generated. Future work will be directed toward finding the optimal time for building relationships of files, decreasing error rate of relationships and building relationships of files on distributed networks.

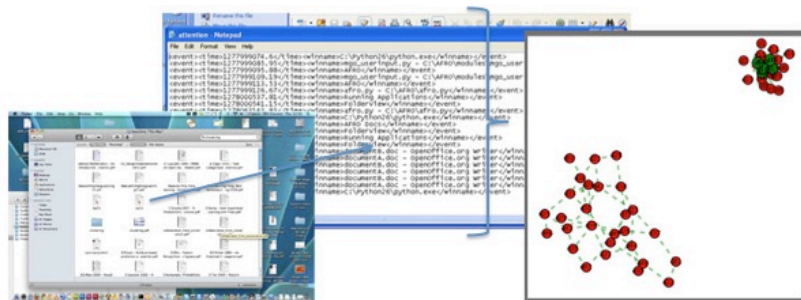


Fig.1 Linking of files based on user behaviour

¹ This work is supported in part by the European Commission under EPSRC Grant EP/E064930/1

References

- [1] M.G Lim, S.Wu, S.Tomasz, Md Rashid, N.Helian, "Personal Storage Grid Architecture: Consuming Cloud Data Space Resources." International Journal of Grid and High Performance Computing, Vol 2, No. 3, 2010
- [2] <http://www.google.com/quicksearchbox> (Accessed Oct 2011)
- [3] <http://developer.apple.com/technologies/mac/> (Accessed Oct 2011)