# Evaluating Personal Information Management Behaviour using Markov Models

David Elsweiler[1], Stefan Mandl[1], Leif Azzopardi[2]

[1]Dept. Computer Science 8 (AI), Univeristy of Erlangen-Nuremberg, Germany
[2]Dept. Computing Science, University of Glasgow, UK
{david.elsweiler,stefan.mandl}@i8.informatik.uni-erlangen.de
leif@dcs.gla.ac.uk

**Abstract:**
Personal Knowledge Management (PKM) and Personal Information Management (PIM) are two related research problems currently being examined by different, although potentially overlapping communities of researchers. We believe that the cross-fertilisation of ideas between these communities could lead to progress in both fields. With this aim in mind, in this position statement we outline our research project, which will use mathematical modelling techniques to help understand and evaluate Personal Information Management behaviour. Evaluation is a major challenge in PIM research and the methods currently available to researchers each have associated problems. In this work, we will discuss the feasibility of using Markov Models to overcome some of these difficulties. Here, we outline some of the challenges involved in PIM evaluation, describe the theory behind the approach and explain how to validate the approach through experimentation in a practical setting.

## 1 Introduction

People now have unparalleled access to digital information. The information comes from different sources, is in different formats and can be stored in different locations, including distributed devices such as mobile phones and Mp3 players. The wealth and diversity of digital information makes it incredibly difficult to organise information in such a way that it can be re-accessed and re-used when it is needed in the future. This situation, termed information overload [Wur89], is a growing problem both in industrial and personal settings [1]. Previous research has shown that people require to re-find and re-use information regularly [DCC+03, CRDS06] and the evidence suggests that this is a task that they find both frustrating and difficult to perform [AJK05, BS04, BJD04, TAAK04]. Personal information management (PIM) as a research field covers attempts to understand PIM behaviour and develop systems to help people manage and re-find their information effectively, without frustration. Researchers in the separate, but related research field of Personal Knowl-

[1]As an example of the cost information overload, a recent study showed that in the UK, IT managers spend 5 million hours per year searching for lost email messages. This equates to 140M in staff costs. Computing Magazine, 24/9/07 (http://www.computing.co.uk/computing/news/2199363/five-million-hours-wasted) last accessed on 13/10/07

edge Management (PKM) have been looking at the problem of how employees can use their personal information collections as knowledge stores, creating in essence a repository of their knowledge for use by themselves or others. The problem spaces for these two fields are very similar, differing perhaps only in focus and approach [VA08]. While PIM focuses on the information within an individual's collection, PKM focuses on the knowledge cued by the information. PKM may even be considered as a specialised area of PIM. Therefore, understanding PIM behaviour and advances in the methods available to achieve this can only benefit both the PIM and PKM research communities.

While PIM focuses on the information within an individual's collection, PKM focuses on the knowledge cued by the information. Therefore, understanding PIM behaviour and advances in the methods available to achieve this can also benefit PKM research, and vice versa.

A key challenge for PIM researchers lies with evaluation. Few techniques exist to help understand how people use PIM tools and consequently, very few of the many prototypes that have been designed have actually been evaluated. This lack of tool evaluation has been repeatedly identified as a factor restricting progress in the field [BS04, ER07, KT07]. By studying how people use PIM tools we can understand what interactive support people need when re-finding, evaluate the effectiveness of existing tools and inform the design of more useful tools for both the management of information and knowledge.

Two methods that have been used to study users' behaviour with PIM tools are log-file analysis (LFA) and laboratory-based user studies (US). LFA is based on recorded user interactions with a PIM tool, normally in naturalistic conditions out-with the control of the experimenter [DCC+03, CRDS06, TAJP07]. LFA examines the quantitative aspects of user behaviour, including the nature of submitted queries and the properties of items accessed. This is an important technique as it allows the capture of a large quantity of data relating to how users behave with systems without the expense and distracting influence of an observer. However, as the captured data show nothing about the goals and intentions of the user, it is difficult to make any concrete statements about the reasons for the behaviour depicted in the logs. The alternative method, user studies, involves observing users in controlled environments where experimenters know about the users' goals because they provide the participants with specific tasks to complete. These investigations can provide an understanding of participants' re-finding strategies, such as the identification of the teleporting and orienteering strategies observed in [TAAK04] and the discovery that people generally have a preference for spatial browsing over keyword search [BN95]. Laboratory-based user studies can also be used to verify the benefits of particular tools, as has been demonstrated in works such as [RCL+98, RCDH03, HNS+04]. However, there are also limitations to laboratory-based studies. Firstly, they are performed in artificially created environments with the presence of an experimenter, both of which are likely to impact on the participants' behaviour. Secondly, when any more than a handful of participants are observed over long time periods it becomes difficult to establish fine grain patterns in behaviour. Further, as user studies rely heavily on experimenter observations, the findings are often criticised for being anecdotal and open to subjective bias.

In our project[2] we will investigate a technique for formally triangulating the findings of the naturalistic log-file analyses and laboratory-based user studies. We will explore how patterns of interaction recorded in naturalistic studies can be modelled mathematically and mapped to the behaviour observed in user studies. This will offer several advantages to the both the PIM and PKM communities, as well as to the larger human-computer interaction and information seeking, information science and information management communities, where LFA and US are used in a similar manner to investigate different problems. First, it will allow the quantitative data captured in interaction logs to be used to confirm and corroborate observations, leaving findings less open to criticism. Second, it will allow a more robust analysis of interaction logs by providing a means to compare behaviour statistically to that exhibited in user studies, as well as providing a means to test different hypotheses regarding the behaviour of users. A further possible benefit of the proposed approach is that the behaviourial information encoded within the models could be used in simulations to provide low cost, repeatable experiments to test algorithms for PIM tools.

## 2 Approach

The basis of our approach will be to statistically model recorded interaction data using processes such as Markov chains. These tools have been applied successfully to numerous problems such as speech [LM98], handwriting [NWF86] and gesture recognition [YX94], etc. A Markov chain is a discrete-time stochastic process which describes the state of a system at successive points in time. By mapping the possible ways that a user can interact with a PIM tool to a set of possible states, we can model the user's interaction as a Markov chain. This is achieved by using the logged data to count the number of times a user moves from one state to another and using the counts to calculate the probability of moving between states. A possible mapping for a typical folder-based email tool, for example, could be at the level of clicks [see Table 1].

| State | Mapping |
|---|---|
| S0 | Clicking on a folder to view the contained messages |
| S1 | Submitting a search query |
| S2 | Clicking on an individual message to read the contents |
| S3 | Sorting the emails by sender |
| S4 | Sorting the emails by date |
| S5 | Sorting the emails by subject |

Table 1: An example mapping scheme from user interaction to Markov chain state

Using Markov chains as a modelling technique offers several advantages in the context of PIM evaluations. First, it provides an intuitive means to visualise complicated sets of interaction data (i.e. the state transition diagrams provide a simple and intuitive representation of the behaviour of users). Second, it offers a set of well established tools to classify

[2]We have an 18 month long project sponsored by the Alexander von Humboldt foundation in Germany

and compare models of behaviour based on the state transitional probabilities. For example the state transitions of novice users and expert users can be compared statistically to determine whether the different users behave differently. Third, the models can be used to simulate behaviour, as has been used in applications such as musical score following [OD01]. However, before these benefits can be realised a number of challenges must be faced.

In the following sections we describe these advantages in greater detail, explain the challenges to be faced before they can be realised, and outline our thoughts on meeting these challenges.

## 3 Markov Models as a Visualisation Technique

One potential use of this approach is the ability to visualise user behaviour. Models could be constructed based on captured interaction data offering, in essence, a picture of exhibited user behaviour. Such models could be created based on different subsets of the captured data. For example, models could be built that characterise the behaviour of different types of user (e.g. experienced vs. novice users, filers vs. pilers, searchers vs. browsers, older participants vs. younger participants etc.), behaviour for different types of task (e.g. looking for older or newer information), or for different types of systems (e.g. browse-based vs. search-based etc.) This would allow researchers to visually examine how behaviour changes in different situations and offer the possibility to spot patterns in user behaviour not possible via observation alone. Currently, little is known about email re-finding strategies, how re-finding strategies generally change in different scenarios, nor the factors that influence these changes. The observations made from analysing the models visually could then be corroborated and analysed further using robust statistical techniques such as linear regression modelling.

The resulting models could also be used to verify and support any observations made in laboratory studies. For example, in our previous work we performed a controlled user study investigating the email re-finding behaviour of 21 participants, which we examined exhaustively using traditional methods [Els07, EBR08a, EBR08b]. One of the outcomes of our analyses was the observation that when completing certain experimental tasks, the participants applied particular re-finding strategies, including the orienteering and teleporting strategies observed previously in [OJ93, CPQ03, TAAK04]. Although such behaviours have been observed often in previous studies, they are open to criticism because they are based on experimenter observations alone. By building Markov models based on the interaction data from our previous user study, we would be able to evidence such observations. A teleporting strategy would be evidenced by a model featuring few states and high probabilities for transitions between the states and an orienteering strategy would be evidenced by a model featuring many states and low probabilities for the transitions.

Thus, modelling interaction data as Markov models and using them as a means to visualise interaction behaviour promises to be a powerful research tool for researchers. Nevertheless, before these visualisation facilities can be realised, we must discover the best ways to map user interactions to models. In a theoretical stage of the project, we plan on analysing potential mapping strategies from interaction data to models in order to establish the benefits that different mappings and models[3] would provide. A number of different solutions will be analysed to determine whether one solution is best suited to all of the project's requirements (visualising, comparing and categorising) or whether different mappings are required for different purposes. We intend to iteratively refine and test the theory throughout the project on both our existing lab-based dataset and a naturalistic dataset (see below) as more is learned about the benefits and potential applications of the approach.

## 4 Triangulating the Findings of Lab-based and Naturalistic Studies

Triangulation of findings from lab-based and naturalistic user-studies is another feature of the Modelling approach that offers potential benefit to researchers. We are currently planning a naturalistic study of email usage patterns in order to evaluate the usefulness of this feature, to learn about the re-finding strategies people use, as well as the problems they have when implementing these strategies. We are interested, for example, in discovering how often people utilise the re-finding strategies that we observed in our lab-based study, "in the wild". To achieve this, we will build Markov models using relevant portions of the lab-based interaction data and use these as behaviour signatures. We will then treat the data captured from the naturalistic study as a stream of interactions and identify the signature interactions from the stream. The process is very similar to that used in speech recognition and will provide us not only with the ability to determine how often these behaviours are exhibited, but to discover which type of users exhibit such behaviours and for which tasks etc.

Nevertheless, again, before we can achieve this, we will need to overcome the difficulties involved in analysing naturalistic email-client interaction logs. One of the challenges here will be filtering out the "noise" in the logs. For example, we must be able to distinguish between the interactions associated with re-finding tasks and other interactions with the email-client, such as reading and browsing emails. We also have to determine when sequences of interactions associated with re-finding behaviour begin and end.

One potential solution is to ask the participants to keep a diary of their email re-finding activities. This would provide time-points of interest in the logs, as well as offering an insight into the motivation behind the behaviour. We have used this technique successfully in the past to examine email re-finding behaviour [ER07]. However, using diaries would not only mean extra work for the participants, possibly discouraging participation, but may also impact on the quality of the collected data, as the diary entries may serve as a reminder that their behaviour is being logged. Another possible way of reducing the noise in the data would be to learn from methods often applied when analysing logs to learn about web searching behaviour. There is a large body of literature in this domain, analysing both search-engine logs [JS06, TAJP07] and client-side logs of web-browser

[3]To incorporate some features, such as temporal patterns and sequential interactions, it may be necessary to evaluate alternative models, such as higher-order Markov chains, or to evaluate other factors not pre-thought of by researchers it may be useful to build Hidden Markov Models

interactions [CP95, WD07], where particular features of interaction are used to identify finding patterns. We believe that analogous techniques may be derived for email-client interaction logs. Nevertheless, such an approach has not been used to examine email re-finding behaviour in the past. Therefore, we plan to pilot both methods – with and without the diary – studying a small number of participants for a short period of time. This will enable us to determine the best approach to take.

## 5 Simulating User Behaviour

The third benefit that we envisage our modelling approach to offer is the possible move towards effective user simulations for PIM. Simulated user studies are very important in fields such as information retrieval [Har00, WJRR04, AdRB07], offering a a low cost and repeatable means to test the utility and effectiveness of search algorithms. A similar approach has been proposed for PIM [KT07]. However, until now, several difficulties have prevented the creation of such a collection for PIM [ER07]. We plan to investigate how our Markov models can be used to represent the behaviour of individual users in known-item enterprise search context e.g. [BCdVS07] such as searching for emails. Given the user behavior as encoded by the Markov models, we aim to combine this information within the development of known-item topics. Already proposed in [AdRB07] is a generative probabilistic model which simulates the generation of known items. We aim to extend this model to incorporate the re-finding behaviour in order evaluate the impact of the user behaviour on the effectiveness of the algorithms and tools used. It is envisioned that this will lead to a realistic PIM simulator that is seeded by the lab-based and naturalistic studies. Before this can be achieved, however, we must find ways to transform system specific interactions into independent behaviour - this relates to the mapping issue referred to above. We must also find ways to test the usefulness of our approach and determine how different factors impact on performance. Once a fully probabilistic model has been developed, we hope to run experiments which will take the input from real interactions to seed the simulations and the process repeated to refine and improve the simulations based these interactions.

## 6 Summary and Conclusions

In this position statement we have described our research project, which aims to explore new methods of learning about and evaluating PIM behaviour. Our approach will involve mathematically modelling participants' interactions with a PIM system and we envisage using the models to visualise exhibited behaviour, relate the findings of laboratory-based and naturalistic studies and also to form the basis of simulated user experiments. We have outlined the benefits these tools will bring and also the challenges we face before they are realised. We believe that the project and the insights it will provide will be of benefit to both the PIM and PKM research communities. We also believe the PKM community will be able to give useful comments and ideas that will improve the project.

## References


[AdRB07] L. Azzopardi, M. de Rijke, and K. Balog. Building simulated queries for known-item topics: an analysis using six european languages. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 455–462, New York, NY, USA, 2007. ACM.

[AJK05] A. Aula, N. Jhaveri, and M. Käki. Information Search and Reaccess Strategies of Experienced Web Users. In *Proc. 14th Intl Conf. World Wide Web*, pages 583–592. ACM Press, 2005.

[BCdVS07] P. Bailey, N Craswell, A. P de Vries, and I. Soboroff. Overview of the TREC 2007 Enterprise Track. In *Proc. of TREC 2007 Notebook.*, 2007.

[BJD04] H. Bruce, W. Jones, and S. Dumais. Keeping and Re-Finding Information on the Web: What Do People Do And What Do They Need? In *ASIST 2004: Proceedings of the 67th ASIST annual meeting*, October 2004.

[BN95] D. K. Barreau and B. Nardi. Finding and reminding: File organization from the desktop. *ACM SIGCHI Bulletin*, 27(3):39–43, 1995.

[BS04] R. Boardman and M. A. Sasse. "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In E. Dykstra-Erickson and M. Tscheligi, editors, *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, New York, NY, USA, 2004. ACM Press.

[CP95] L. Catledge and J. Pitkow. Characterizing browsing strategies in the World Wide Web. *Computer Networks and ISDN Systems*, 27(6):1065–1073, 1995.

[CPQ03] R. G. Capra and M. A. Perez-Quinones. Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information. Technical report, Computer Science Dept., Virginia Tech, 2003.

[CRDS06] E. Cutrell, D. Robbins, S. Dumais, and R. Sarin. Fast, flexible filtering with phlat. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 261–270, New York, NY, USA, 2006. ACM Press.

[DCC$^+$03] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D.C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In Mark Sanderson, editor, *SIGIR '03: Proc. 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 72–79, New York, NY, USA, 2003. ACM Press.

[EBR08a] D. Elsweiler, M. Baillie, and I. Ruthven. Exploring memory in email refinding. *ACM Trans. Inf. Syst.*, 26(4):1–36, 2008.

[EBR08b] D. Elsweiler, M. Baillie, and I. Ruthven. On Understanding the Relationship Between Recollection and Refinding. *Submitted to Journal of Digital Information sp. ed. Psychology and PIM*, 2008.

[Els07] D. Elsweiler. *Supporting Human Memory in Personal Information Management*. PhD thesis, The University of Strathclyde, 2007.

[ER07] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 23–30, New York, NY, USA, 2007. ACM Press.

[Har00] D. Harman. What we have learned, and not learned, from TREC. In *Proceedings of the BCS IRSG 22nd Annual Colloquium on Information Retrieval Research*, 2000.

[HNS+04] S. Harada, M. Naaman, Y. J. Song, Q.-Y Wang, and A. Paepcke. Lost in memories: Interacting with photo collections on PDAs. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 04)*, pages 325–333, 2004.

[JS06] B. J. Jansen and A. Spink. How are we searching the world wide web?: a comparison of nine search engine transaction logs. *Inf. Process. Manage.*, 42(1):248–263, 2006.

[KT07] D. Kelly and J. Teevan. *Personal Information Management*, chapter Understanding what works: Evaluating personal information management tools, pages 190–204. Seattle: University of Washington Press., 2007.

[LM98] B. Logan and P. Moreno. Factorial HMMs for acoustic modeling. In *ICASSP*, pages 813–816, 1998.

[NWF86] R. Nag, K.H. Wong, and F. Fallside. Script Recognition Using Hidden Markov Models. In *ICASSP86*, pages 2071–2074, 1986.

[OD01] N. Orio and F. Dechelle. Score Following Using Spectral Analysis and Hidden Markov Models. In *ICMC*, 2001.

[OJ93] V. L. O'Day and R. Jeffries. Orienteering in an information landscape: How information seekers get from here to there. In *Proceedings of INTERCHI '93*, pages 438–445, 1993.

[RCDH03] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz. Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores. In *Proc. INTERACT 2003*, pages 184–191, 2003.

[RCL+98] G. Robertson, M. Czerwinski, K. Larson, D. C. Robbins, D. Thiel, and M. van Dantzich. Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 153–162, New York, NY, USA, 1998. ACM Press.

[TAAK04] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In E. Dykstra-Erickson and M. Tscheligi, editors, *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, New York, NY, USA, 2004. ACM Press.

[TAJP07] J. Teevan, E. Adar, R. Jones, and M. Potts. Information Re-Retrieval: Repeat Queries in Yahoo's Logs. In *Proceedings of the 30th Annual ACM Conference on Research and Development in Information Retrieval (SIGIR '07)*, 2007.

[VA08] M. Völkel and A. Abecker. Cost-Benefit Analysis for the Design of Personal Knowledge Management Systems. In *Proc. of 10th International Conference on Enterprise Information Systems*, JUN 2008.

[WD07] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 21–30, New York, NY, USA, 2007. ACM.

[WJRR04] R. W. White, J. M. Jose, C. J. Van Rijsbergen, and I. Ruthven. A simulated study of implicit feedback models. In *In Proceedings of ECIR 2004*, pages 311–326, 2004.

[Wur89] R. S. Wurman. *Information Anxiety*. New York: Doubleday, 1989.

[YX94] J. Yang and Y. Xu. Hidden Markov Model for Gesture Recognition. Technical Report CMU-RI-TR-94-10, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 1994.