Doctoral Dissertation
(Overview)
Academic Year 2013

Collective Experiential Computing on the Web via Experience Mapping

（ウェブ上のユーザコンテクストに応じた集合的経験知コンピューティング）

Graduate School of Media and Governance
Keio University

Jeremy Hall

Thesis Abstract

No.__________

Registration Number:	□ “KOU” □ “OTSU” No. *Office use only	Name:	Jeremy Hall

Title of Thesis:

Collective Experiential Computing on the Web via Experience Mapping

Summary of Thesis:

Using the Web to learn about new topics, and discover new information is difficult when searching an unfamiliar topic. Unlike the real-world where we can observe the behavior of others and ask for help if need be, the current Web is a partitioned space with little interaction between people on the Web, and few opportunities to learn from others' experience. This research aims to capture and utilize the experience of people on the Web to allow people to use their built-in human-to-human communications skills. By utilizing this system, people may become familiar with a topic or knowledge-space much more quickly than is otherwise possible, because they can skip the tedious and demanding task of surveying several websites, learning their relative strengths and weaknesses, and then evaluating the merits of each on their own. This research introduces several algorithms dedicated to Mapping, Storing, Comparing, and Visualizing Experiential-Knowledge, which is a complex and generally hidden type of data in the context of the Web. The concept of Experiential-Knowledge is the knowledge that people browsing the Web obtain as a natural course of viewing content, and becoming knowledgeable about the relationships of websites and their reliability, trustworthiness and value. At the heart of the systems described in this research, the ability to perform Experiential-Mapping, converting the experience of one person into another's context while preserving the original point-of-view, is an essential technology which allows for the comparison and understanding of multiple experts' experience simultaneously. We provide several methods of querying, interacting with, and visualizing this Experiential-Knowledge. This paper also describes several architectures to support such Experiential-Knowledge-based systems, as well as discussion of implemented prototypes of these proposed architectures. Finally, we provide an evaluation section which uses volunteer test users and our prototype implementation to calculate the potential for Experiential-Sharing within our system.

Experiential Computing, Data Mining, Data Visualization, Behavior Analysis, Web Architecture

Keio University

Thesis Abstract

No.＿＿＿＿＿

Registration Number:	□ "KOU" □ "OTSU" No. *Office use only	Name:	Jeremy Hall
Title of Thesis: Collective Experiential Computing on the Web via Experience Mapping (ウェブ上のユーザコンテクストに応じた集合的経験知コンピューティング)			

Summary of Thesis:

Web を用いてある新しいトピックについて知識を獲得すること、新しい情報を発見することは、それが精通したトピックでない場合、困難を伴う。実空間では他者の行動を観察し、必要に応じて他者の助けを求めることができるのに対し、現在の Web という情報空間では他者とのインタラクションが制限されており、他者の経験から学習する機会は少ない。本研究の目的は、Web 利用者の経験を獲得・活用し、人間生来の能力である人から人へのコミュニケーション・スキルを活用できるシステムの実現にある。本システムの利用者は、複数の Website を試行錯誤により探索する段階をスキップし、それら Website の相対的な利点や弱点を学習し、各 Website 独自の価値を評価することにより、あるトピックや知識空間により迅速に精通することができるようになる。 本研究は、Web 上の複雑かつ暗黙的な経験的知識（Experiential-Knowledge）を計算可能なデータに変換・マッピング・蓄積・比較・可視化するための複数のアルゴリズムについて述べている。経験的知識という概念は、人々が Website のコンテンツを閲覧する際に獲得する、複数の Website 間の関連性や信頼性、価値についての知識のことを指す。本研究で述べるシステムの本質的な技術は、ある利用者の経験を固有の視点を保持したまま他の利用者の文脈に変換する「経験知マッピング（Experiential-Mapping）」にある。この技術により、利用者は同時に複数の専門家の経験を比較し、知識として獲得することが可能となる。本研究は、経験的知識を図形的に可視化し、共有・比較し、検索条件として設定するための複数の方法について述べる。さらに本論文では、これら経験知ベースシステムのための複数のアーキテクチャについて述べ、それらアーキテクチャに基づいて実装したプロトタイプシステムについて考察する。最後に、実際のテストユーザを対象とした実証実験により、本システムの経験知共有の実現可能性と有効性を確認する。

経験知計算、データマイニング、データ可視化、行動分析、ウェブ・アーキテクチャ

Acknowledgements

First and foremost, I'd like to thank my main research advisor, Professor Yasushi Kiyoki. Professor Kiyoki has worked with me over the course of my Masters and PhD program, and given advice and mentorship on all things related to my academic life. He has also provided many opportunities to enrich my academic life, including working on interesting projects, and attending conferences around the world.

I'd like to thank Shiori Sasaki for the support she has given me not only with reading my papers, watching countless presentations and helping me with my Japanese, but also with everday issues, and the countless, countless numbers of forms I've had to deal with over my time at Keio SFC. Asako Uraki was also an immense help and support during my entire time at SFC, always willing to help, and giving invaluable advice whenever I needed it. and the members of his MDBL laboratory for the constant support I've received. This support wasn't just in the form of helping with paper writing and presentation structure, but also everyday life struggles and too many forms and too much paperwork to count. Professor Kurabayashi was another member of my laboratory who had an immense impact on me in my early academic development, and continuing to lend advice and mentorship even after going on to form his own research laboratory.

I'd also like to thank Professor Kuniaki Mukai from novel computing, for asking a mix of interesting questions and observations, and for giving good ideas regarding presentations that I've given. Dean, Professor, and co-Advisor Hideyuki Tokuda always gave insightful advice, and provided really interesting conversation when I was given a chance to show him a paper or presentation of mine. Professor and co-Advisor Yoshiyasu Takefuji was someone I could trust to pierce through my research and deliver questions on topics at the heart of the matter. I also owe Professor and co-Advisor Ikumi Waragai a big thanks for agreeing to help me mid process. This required her to learn about my research of which there was a lot, in a very short time, and to do so in a topic that is outside of her normal research scope. I enjoyed the alternate perspective her questions and discussion was able to shed light on. Although not an official advisor, I'd like to thank Professor Rodney Van Meter for help and discussion throughout my academic career, and for really pushing me to tighten up my final thesis. He has a keen eye for making sure performance, system architectures, and their overall results add up.

I would also like to thank the members of my MDBL laboratory for their support over the years of my membership to that research lab. They have listened to countless presentations, paper overviews, research ideas... and have done so in a mix of Japanese and English which was less than ideal. Despite this, they have always provided invaluable support, attention, questions, and friendship over the several years I have been here.

I'd like to thank my family and friends for giving me support and advice when needed, especially during times of difficulty. My parent and grandmother live half a world away, but I could always rely on them to help me out. My wife Risa has been a huge support to me, both for bouncing ideas off of and giving advice, and helping me during bleaker times. Her enthusiasm for my research, and encouragement in sharing those ideas with other people has helped me to not only express my ideas more clearly, but also boosted my confidence while doing so.

Finally, thank you to the anonymous volunteers who agreed to let me use their browsing experience to show the value of my research. For that I am extremely grateful...

Contents

Chapter 1

Introduction

1.1 Knowledge Transfer between People

The ability to learn from and utilize the experience of others is an important factor in the advancement of knowledge and technology. The entire education system is built upon this premise, and utilizes direct interactions between human beings to convey knowledge based on teachers' experience.

The Web was created for the purpose of spreading knowledge. In order to disseminate information as efficiently as possible, its architecture was designed with scalability and parallelism in mind. This design emphasizes independence, distributedness, and isolationism. While achieving the goals of scalability and parallelism, the architecture of the Web also effectively partitions people accessing it from each other, preventing our intuitive human-to-human learning capabilities from being effective, as each person is essentially browsing completely alone, in their own instance of the global World Wide Web. While current methods of finding information on the Web focus on ways to link keywords to explanation or discussion, it is still up to the user to find many potentially disparate sources of information, understand how they fit together, draw conclusions about the reliability and utility of various disparate information sources, and to do so with the potential handicap of not knowing the most efficient or proper keywords or wording for finding such sources.

This paper describes ways of utilizing the innate human ability to learn from others, by creating new methods of interacting with other people, and their experience, and by creating new visualizations to represent knowledge on the Web that has been hidden until now. We have designed, modeled, and implemented algorithms and data structures for mapping, storing, comparing, and visualizing human experience, relative experiences, human contexts, human expertise, and knowledge-sources' valuation.

A key technology we use to achieve our goals is Experience-Mapping. We have developed the capability to map a person's Experiential-Data from its original context, to the context of another person. This allows for smooth experience transfer and the ability to direct compare experience from multiple people, as the context is identical.

The majority of focus for knowledge-tools on the Web today focus on keywords, semantics, textual content and manual annotation, and then aggregating such information into single-dimensional ranking. This research focuses on an entirely different approach to sharing knowledge, by automatically capturing and compiling individuals' experience, and then focusing on matching and comparing individuals' experience with other individuals. By utilizing information on an individual-level, we allow more opportunity for each person's built-in human ability to understand, judge, and learn from others actions and explanations to allow efficient knowledge-transfer.

1.2 Overview of Utilizing Peoples' Experience

This section describes conceptually, several approaches to model, calculate, visualize, and the benefits of doing so with regard to experience on the Web. Later sections formally describe the underlying algorithms involved.

1.2.1 Modeling Web Experience

In our model, a person's Web experience is defined as the content that they have seen on the Web, and the methods of retrieving that content. Web Experience is a good candidate for modeling, because the majority of what a person experiences on the Web can be captured in that person's Web browser. Also, the list of all possible experiences dealing with Websites is limited to a few actions in the browser, such as loading a website, switching tabs, moving the mouse, scrolling, or spending time reading Web content. By understanding the context of a person through their experience, we can discover other people who are appropriate knowledge-donors due to their relative high-experience, and their similarity in the Experiential-Patterns described below.

In our research, we specifically utilize the number of times that a person has viewed a website, the action that led to that view such as switching tabs or clicking a link, and the time that the action occurred in order to describe a person's Web Experience. We store the 'relationships' between these sites, which are created by the aforementioned actions that a person can take. For instance, switching tabs between Site A and Site B would create a relationship between those sites in our model.

These data-points allow for the differentiation of many types of users and Experiential-Patterns, for example:

- Low Familiarity: A person just beginning to learn about a topic, characterized by recently browsing websites with very few relationships to other previously browsed websites.
- High Familiarity: A person browsing websites which are very familiar to them, and highly related to what they normally browse, characterized by viewing websites with many recent relationships with each-other.
- Recency: Outdated information, characterized by a cluster of websites with relatively-older relationships.
- Expertise: An expert-user for a given topic, characterized by a person with at least an order of magnitude more utilization of a cluster of websites versus another person using a similar set of websites. The 'expert' moniker is completely context dependent, and is only used when comparing two users for potential knowledge-transfer, and never as a global condition. When comparing expertise, we also factor in experience-decay, which is the phenomenon of older experience being worth less than newer experience.
- Broad Topic: A set of websites characterized by a cluster of highly-related (in terms of actions such as clicking on links or switching tabs) websites, with a medium amount of shared-keywords.
- Focused Topic: A set of websites characterized by a cluster of highly-related websites, each with similar important-keywords.

1.2.2 Calculating People's Experience

By directly modeling people's Web Experience, this research has the ability to perform unique calculations which allows for comparisons between people in terms of their experience. Our data model is fine-grained, so that the calculations can be performed within a very specific topical-context, and thus be sensitive to a person's immediate context including:

- Is this person just learning about a topic, or are they already familiar with it?
- Does a query about cameras refer to camera reviews, camera lenses, or sports photography?

- Is this person looking for general information, or specific and specialized knowledge?
- Is this person's knowledge about their query-subject out of date?

We can thus accept a query from a user in the form of a list of Websites, and extrapolate from their Web Experience the contextually-appropriate meaning of the query. We can utilize the same techniques to find other people with relatively more experience for the same context, and then generate a visualization which teaches the user how a relatively more experienced person values websites within the queried-context, and how that experience compares to their own experience, and how the websites in question relate to each other.

1.3 Practical Applications of this Research

Our research is about enabling people to interact with the Web in a more intuitive way than current Web interfaces. The abstractions are at a higher-level, and are designed to work with natural human instincts and built-in functions by projecting complex and abstract concepts such as 'knowledge-sources', 'experience', 'behavior patterns', and 'experience-decay' into easy-to-understand everyday concepts such as size, shape, color, and two-dimensional movement and euclidean distance.

Our research has created three concrete ways of utilizing experiential-functions to benefit people who use the Web:

- As a way to visualize real-time behavior and movement of friends on the Web.
- As a mechanism to detect experience-imbalance between collaborating users in real-time, and transfer the appropriate knowledge at the appropriate time.
- As a mechanism to enable Query-For-Expertise functionality, whereby a persons experience is automatically and continuously calculated and utilized to get the context of that person, and to find other users called 'experts' who contain relatively-more experience. Relevant experts' experience is then visualized, and compared to a visualized representation of the original queryee's own relevant experience.

1.3.1 Revealing the Hidden Dimensions of the Web

Each person browses the Web alone, without seeing any of the myriad activity that's bound to be occurring around them. Our research described in the paper listed in Appendix A.1.1 "Creating a personal-context oriented real-time dynamic and collaborative space" visualizes this hidden activity, bringing new concepts of distance between websites, real-time popularity of URLs, the locations of friends in relation to oneself, and the browsing patterns of those friends, and how they relate to one's own browsing. By visualizing such information, we allow the built-in human capabilities of learning from one's surroundings, and others' actions to be mapped and utilized within the concept of the Web.

1.3.2 Real-time Dynamic Updates

The Web is a dynamic and constantly evolving body of knowledge. New websites are constantly being added, new information is constantly being added, and everyone has a unique personality and experiential-background which has a correspondingly unique type of desired information. Our research described in the paper listed in Appendix A.1.2 "Creating a personal-context oriented real-time dynamic and collaborative space" allows a distributed network of collaborating users to automatically identify information which other collaborating users would be interested in based on analysis of their Experiential-Knowledge, and then notifies them of the discovery.

1.3.3 Directly Learning from Others

Learning a new topic on the Web is massively inefficient when looking at the big picture. There is undoubtedly someone more qualified and more familiar with the topic than the person researching it for the first time. Our research described in the paper listed in Appendix A.1.3 "Building a Collective-Experience Engine for Experience-Transfer amongst Web Users" takes advantage of that fact by spreading experience amongst people browsing the Web. Now, when faced with a number of potential websites and no way of knowing which to trust, or how to valuate them, instead we can query the collective experience of people browsing the Web. This research provides a unique query, information gathering, and visualization technique for allowing intuitive manipulation of the abstract concepts of experience and expertise.

1.4 Systems and Implementations Created for this Research

In addition to the data models and algorithms created over the course of our research, we have designed two system architectures, and implemented two prototypes based on those designs. By providing such architectures, we hope to enable others to utilize the technologies described in this research in a practical way. We used the implementations in several ways:

- To be able to give practical advice regarding appropriate technologies to utilize.
- To ensure that the algorithms can be implemented in a manner efficient enough to be of practical value to society.
- To reveal bottlenecks, and allow us to offer solutions in the form of optimizations to overcome such bottlenecks.
- To allow us to discuss the practicality of such systems.

Appendix A

Paper List

A.1 Journal Submissions

1. J. Hall and Y. Kiyoki, "Building a Collective-Experience Engine for Experience-Transfer amongst Web Users," International Journal of Data Mining & Knowledge Management Process Vol.4, No.1, pp. 1-18, January 2014.
2. J. Hall and Y. Kiyoki, "Creating a personal-context oriented real-time dynamic and collaborative space," Information Modeling and Knowledge Bases XXIV pp.76-88, 2013.
3. J. Hall and Y. Kiyoki, "Identifying and Propagating Contextually Appropriate Deep-Topics amongst Collaborating Web-Users," Information Modeling and Knowledge Bases XXV, pp. 82-101, 2013.

A.2 International Conference Papers

1. J., Hall, S., Kurabayashi, Y., Kiyoki. "Multimedia Data Analysis on a Massively Distributed Parallelization Network of AnonymousWeb Clients". In Proceedings of IEEE Mining and Web, pp.615-620, Perth, Australia, April 20-23, 2010.

Bibliography

[1] Ronald J. Vetter, Christopher G. Spell, and Charles R. Ward. Mosaic and the world-wide web. *IEEE Computer*, 27(10):49–57, 1994.

[2] Vincent Bouthors and Olivier Dedieu. Pharos, a Collaborative Infrastructure for Web Knowledge Sharing. Research Report RR-3679, INRIA, 1999. URL `http://hal.inria.fr/inria-00072992/en/`. Projet SOR.

[3] Charles A Perfetti, Jean-François Rouet, and M Anne Britt. Toward a theory of documents representation. *The construction of mental representations during reading*, pages 99–122, 1999.

[4] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. Web usage mining: discovery and applications of usage patterns from web data. *SIGKDD Explor. Newsl.*, 1(2):12–23, January 2000. ISSN 1931-0145. doi: 10.1145/846183.846188. URL `http://doi.acm.org/10.1145/846183.846188`.

[5] Christoph Holscher and Gerhard Strube. Web search behavior of internet experts and newbies. *Computer Networks*, 33(1-6):337 – 346, 2000. ISSN 1389-1286. doi: 10.1016/S1389-1286(00)00031-1. URL `http://www.sciencedirect.com/science/article/pii/S1389128600000311`.

[6] Maurice D. Mulvenna, Sarabjot S. Anand, and Alex G. Büchner. Personalization on the net using web mining: introduction. *Commun. ACM*, 43(8):122–125, August 2000. ISSN 0001-0782. doi: 10.1145/345124.345165. URL `http://doi.acm.org/10.1145/345124.345165`.

[7] Michael K Bergman. The deep web: Surfacing hidden value. *Journal of Electronic Publishing*, 7(1):1–17, 2001. URL `http://www.press.umich.edu/jep/07-01/bergman.html`.

[8] Myra Spiliopoulou and Carsten Pohle. Data mining for measuring and improving the success of web sites. *Data Mining and Knowledge Discovery*, 5:85–114, 2001. ISSN 1384-5810. doi: 10.1023/A:1009800113571. URL `http://dx.doi.org/10.1023/A%3A1009800113571`.

[9] M. Anne Britt and Cindy Aglinskas. Improving students'ability to identify and use source information. *COGNITION AND INSTRUCTION*, 20:458–522, 2002.

[10] Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa, and Shunsuke Uemura. Refinement of tf-idf schemes for web pages using their hyperlinked neighboring pages. In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, HYPERTEXT '03, pages 198–207, New York, NY, USA, 2003. ACM. ISBN 1-58113-704-4. doi: 10.1145/900051.900096. URL `http://doi.acm.org/10.1145/900051.900096`.

[11] Lan Yi, Bing Liu, and Xiaoli Li. Eliminating noisy information in web pages for data mining. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 296–305, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0. doi: 10.1145/956750.956785. URL `http://doi.acm.org/10.1145/956750.956785`.

[12] Suresh K Bhavnani, Renju T Jacob, Jennifer Nardine, and Frederick A Peck. Exploring the distribution of online healthcare information. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 816–817. ACM, 2003.

[13] Magdalini Eirinaki and Michalis Vazirgiannis. Web mining for web personalization. *ACM Trans. Internet Technol.*, 3(1):1–27, February 2003. ISSN 1533-5399. doi: 10.1145/643477.643478. URL `http://doi.acm.org/10.1145/643477.643478`.

[14] Peter Biddle, Paul England, Marcus Peinado, and Bryan Willman. The darknet and the future of content protection. In Joan Feigenbaum, editor, *Digital Rights Management*, volume 2696 of *Lecture Notes in Computer Science*, pages 155–176. Springer Berlin / Heidelberg, 2003.

[15] F. Wiesman, H.J. van den Herik, and A. Hasman. Information retrieval by metabrowsing. *Journal of the American Society for Information Science and Technology*, 55(7):565–578, 2004. ISSN 1532-2890. doi: 10.1002/asi.10410. URL `http://dx.doi.org/10.1002/asi.10410`.

[16] Saskia Brand-Gruwel, Iwan Wopereis, and Yvonne Vermetten. Information problem solving by experts and novices: analysis of a complex cognitive skill. *Computers in Human Behavior*, 21(3):487–508, 2005.

[17] Marc Stadtler and Rainer Bromme. Dealing with multiple documents on the www: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, 2(2-3):191–210, 2007.

[18] Yu. Suzuki, M. Mitsukawa, and K. Kawagoe. A image retrieval method using tfidf based weighting scheme. In *Database and Expert Systems Application, 2008.*

DEXA '08. 19th International Workshop on, pages 112 –116, sept. 2008. doi: 10.1109/DEXA.2008.106.

[19] Luis Von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. recaptcha: human-based character recognition via web security measures. *Science*, 321(5895):1465–1468, 2008. URL `http://www.ncbi.nlm.nih.gov/pubmed/18703711`.

[20] Damon McCoy, Kevin Bauer, Dirk Grunwald, Tadayoshi Kohno, and Douglas Sicker. Shining light in dark places: Understanding the tor network. In Nikita Borisov and Ian Goldberg, editors, *Privacy Enhancing Technologies*, volume 5134 of *Lecture Notes in Computer Science*, pages 63–76. Springer Berlin / Heidelberg, 2008.

[21] S. Kurabayashi, T. Ueno, and Y. Kiyoki. A context-based whole video retrieval system with dynamic video stream analysis mechanisms. In *Multimedia, 2009. ISM '09. 11th IEEE International Symposium on*, pages 505 –510, dec. 2009. doi: 10.1109/ISM.2009.85.

[22] Saskia Brand-Gruwel and Marc Stadtler. Solving information-based problems: Evaluating sources and information. *Learning and Instruction*, 21(2):175 – 179, 2011. ISSN 0959-4752. doi: http://dx.doi.org/10.1016/j.learninstruc.2010.02.008. URL `http://www.sciencedirect.com/science/article/pii/S0959475210000228`. Special Section I: Solving information-based problems: Evaluating sources and information, Special Section II: Stretching the limits in help-seeking research: Theoretical, methodological, and technological advances.

[23] Ivar Bråten, Helge I. Strømsø, and Ladislao Salmerón. Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, 21(2):180 – 192, 2011.

[24] Peter Gerjets, Yvonne Kammerer, and Benita Werner. Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, 21(2):220 – 231, 2011. ISSN 0959-4752. doi: http://dx.doi.org/10.1016/j.learninstruc.2010.02.005. URL `http://www.sciencedirect.com/science/article/pii/S0959475210000198`. Special Section I: Solving information-based problems: Evaluating sources and information, Special Section II: Stretching the limits in help-seeking research: Theoretical, methodological, and technological advances.

[25] Susan R. Goldman. Choosing and using multiple information sources: Some new findings and emergent issues. *Learning and Instruction*, 21(2):238 – 242, 2011. ISSN

0959-4752. doi: http://dx.doi.org/10.1016/j.learninstruc.2010.02.006. URL http://www.sciencedirect.com/science/article/pii/S0959475210000204. Special Section I: Solving information-based problems: Evaluating sources and information, Special Section II: Stretching the limits in help-seeking research: Theoretical, methodological, and technological advances.

[26] Erin Catto. Box2d: A 2d physics engine for games. http://http://box2d.org/, 2012.

[27] Google. Google chrome. http://googlechromereleases.blogspot.com/, 2012.

[28] Processingjs. http://processingjs.org/, 2012.

[29] Uli Hecht. box2dweb. http://code.google.com/p/box2dweb/, 2012.

[30] Guillermo Rauch. Socket.io. http://socket.io/, 2012.