Revolution of Knowledge Work

A Tekes Large Strategic Opening

First Phase 9/2013−4/2015 Final Report

26.8.2015

Patrik Floréen, Salvatore Andolina, Marco Filetti, Dorota Głowacka,
Mohammad Hoque, Markus Koskela, Kai Kuikkaniemi, Tuukka Ruotsalo
Helsinki Institute for Information Technology HIIT

Benjamin Cowley, Andreas Henelius, Kristian Lukander, Kai Puolamäki, Antti Ukkonen
Finnish Institute of Occupational Health

Abstract

The aim of the Revolution of Knowledge Work (Re:Know) project is to create information-seeking and sense-making methods supporting a revolution of knowledge work. The main idea are the *symbiotic human-information interfaces*, supporting the *combined* potential of *human creativity* and the *capability of computers* to handle data, with the *human in control*.

The project is a Tekes large strategic opening funded with 1.68 million € for the period 1.9.2013−30.4.2015 (20 months). The project continues as the Revolution of Knowledge Work 2 (Re:Know 2) project, which runs for 1.4.2015−31.3.2017 with a planned extension until 31.3.2019. The project research partners are Helsinki Institute for Information Technology HIIT (Aalto University and University of Helsinki) and the Finnish Institute of Occupational Health (FIOH).

The project resulted in research results supporting explorative information seeking and the results were demonstrated with prototypes. A plan for the ecosystem development was made. The original SciNet system ideas were commercialised as Etsimo Oy.

1 Introduction

A huge amount of information is spread out in various data silos. The current systems and search engines have inflexible views to the data and they have only a limited ability to study the large data masses, leaving knowledge workers trapped in individual information bubbles. Current systems constrain the work instead of supporting the combined potential of human creativity and the capability of computers to handle big data.

The project combines the multidisciplinary world-class expertise in machine learning, human-computer interaction, distributed computing, cognitive neuroergonomics and human factors at work, available within Helsinki Institute for Information Technology HIIT and the Finnish Institute of Occupational Health. Our objective is to develop symbiotic human-information interfaces, which pave the way for a revolution of knowledge work.

Symbiotic human-information interfaces combine heterogeneous data sources and utilise the context of use and user actions to jointly with the user determine what information is most likely relevant, and provide the user with a new type of interactive and proactive interface to the data. In the context of knowledge work, we use our know-how on both computational principles and how humans process information to develop a new information management and utilisation paradigm, enabling humans and computers to support each other optimally.

By combining the potential of human creativity and the capability of computers to handle data, we support human cognition, freeing the knowledge worker's time to sense-making, collaboration and creative thinking.

As pre-existing knowledge, HIIT had developed a first version of SciNet with the radar interface, see Figure 1.1. The data used was scientific literature from Thompson Reuters, ACM, IEEE and Springer; in essence almost all scientific literature available online. However, the data did not include the full article texts, but metadata, such as authors, title and abstract. The number of publications included was about 60 million.

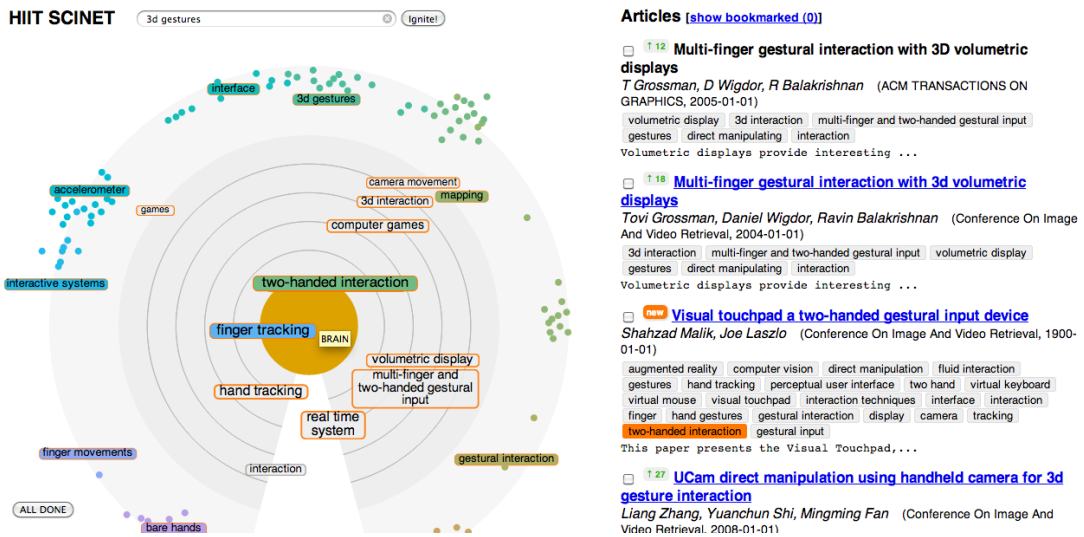

Figure 1.1. SciNet screenshot. Interaction with the data visualisation is undertaken by moving keywords towards the centre (more important) or away from it (less important) and then refreshing the view by pressing the middle. Corresponding articles are shown to the right.

The Tekes decision of 5.7.2013 states the following objectives for the project:

1. Functioning information seeking prototype that takes into account the physiological state of the user and other user context. The prototype adapts on the basis of the user's actions and can search data from different data sources. Results have been published open source.
2. Prototype for information seeking for the use of one company. The prototype has an interactive user interface, is proactive and reacts on user feedback. The prototype can visualise the results in a way that provides added value.
3. The development has started and a plan has been made for the ecosystem that promotes the commercialisation of the tools developed.

Point 1 was realised as so-called Prototype 1 and point 2 as so-called Prototype 2. Point 3 is the open ecosystem plan [KP15]. The results, showing that the project has produced the results required in the Tekes decision, were presented in the Steering Group meeting of 17.11.2014.

2 The Process

A consortium agreement was signed between the project research partners Aalto University, University of Helsinki and FIOH. The formal decision making body is the General Assembly, consisting of one representative per organisation (Samuel Kaski for Aalto, Patrik Floréen for UH, and Kiti Müller until 3.11.2014 and then Kai Puolamäki for FIOH). These representatives are the respective scientist in charge for the project in their own organisation. The General Assembly had 5 meetings: 3.10.2014, 7.11.2014, 8.1.2015 (email meeting), 17.3.2015 and 25.3.2015 (email meeting).

Kiti Müller moved from FIOH to Nokia on 3.11.2014 and Kai Puolamäki took over as Director of the Brain Work Research Centre at FIOH, and also as scientist in charge for the project for the part of FIOH.

A number of companies were invited to an Advisory Board. The Tekes Steering Group consists of the members of the General Assembly and the Advisory Board. The Tekes contact person for the project is Katja Ahola. A non-disclosure agreement was made for the participants in the Advisory Board. Companies with which collaboration was made signed particular collaboration agreements between the research consortium and the company in question.

The companies and respective Advisory Board members were as follows:

- Futurice Oy: Head of UX & Service Design Risto Sarvas
- M-Brain Oy: CTO Kimmo Valtonen
- Nokia Oyj: Director Jyri Huopaniemi. Nokia left 1.11.2014 after the devices and services division had been sold to Microsoft.
- Sanoma Media Finland Oy: Managing Editor of Tiede Magazine Annikka Mutanen
- Wärtsilä Finland Oy: General Manager Tero Hottinen

At the end of the project, the following companies and respective Advisory Board members joined:

- Elisa Oyj: Markus Ahokangas
- IBM Finland Oy: Ville Peltola
- Tieto Oyj: Matti Vakkuri

The Steering Group met in September 2013 (email meeting) and 9.5.2014, 17.11.2014 and 30.4.2015.

The daily work was organised using meetings and two email lists (all Re:Know researchers & the Principal Researchers). The Big Meetings are bi-monthly gatherings for everybody working in the project. The Principal Researchers meetings, held mainly as videoconferences, were targeted to professors and WP coordinators, but also others were welcome to participate. The meetings were used to spread information, manage the project and synchronise the work. In addition there have been many WP-internal and prototype-internal meetings.

Big Meetings were held 2.9.2013 (Kick-off), 11.10.2013 (Re:Know Retreat), 4.2.2014,16.4.2014, 12.6.2014, 27.8.2014, 4.11.2014, 16.1.2015 and 17.3.2015. Principal Researchers meetings were held 20.9., 30.10., 20.11. and 16.12.2013; 20.1., 6.2., 14.2., 28.2., 14.3., 28.3., 11.4., 25.4., 6.6., 19.6., 8.8., 22.8., 5.9., 12.9., 19.9., 26.9., 17.10., 24.10., 31.10., 14.11., 28.11. and 12.12.2014; 9.1., 23.1., 6.2., 20.2., 6.3., 2.4. and 30.4.2015.

The project material is available to the researchers in Google Drive.

The project website www.reknow.fi went operational 21.8.2013. Unigrafia designed the project logo, visible on the front page of this document.

Personnel

Our project personnel is very international. Of the professors, Giulio Jacucci is Italian. We recruited a large number of postdocs for the project, and many were recruited from abroad. In the following is a list of researchers of non-Finnish origin and their total person-months in the project:

At University of Helsinki:
Mohammad Hoque 15.21

Diogo Cabral 9.79
Dorota Głowacka 9.60
Khalil Klouche 9.46
Salvatore Andolina 7.14
Joseph Sakaya 6.81
Oswald Barral 5.68
Yuan Galo 5.42
Han Xiao 3.42
Sayantan Hore 3.21
Haibo Jin 2.39
Maninder Singh 1.79
Total non-Finnish 79.92 out of 106.45 (a share of 75.1%)

At Aalto University:
Xuran Zhao 16.22
Marco Filetti 14.43
Apurva Nandan 6.55
Polina Rozenshtein 4.00
Hanieh Poostchimohammadabadi 3.11
Dee Pedram 2.86
Luana Micallef 1.22
Wu Zhiqian 0.90
Michiel Spapé 0.70
Total non-Finnish 49.99 out of 83.26 (a share of 60.0%)

At FIOH:
Benjamin Crowley 11.15
Total non-Finnish 11.15 out of 56.67 (a share of 19,7%).

The total manpower in the project was 246.38 person-months, out of which 141.06 (57,3%) was by researchers of non-Finnish origin. Note that the person-months noted are effective person-months, i.e., holidays and sick leaves are not included.

Several researchers working in the HIIT Focus Area “Augmented Research,” as well as at the FIOH Brain and Work Research Centre, also in effect contribute to the project, without being paid from the project.

Visits (min. 1 month)

Fulbright visitor Prof. Peter Brusilovsky from University of Pittsburgh was at HIIT and Aalto University 1.8.−2.11.2013 and 2.5.−1.7.2014. At his home institution, an interactive recommender system for academic talks has been developed. By using SciNet, a personal user model is developed and is transferred to their recommender system, to study whether this user model provides advantage for the talk recommendations.

PhD students Eemil Lagerspetz and Ella Peltonen visited UC Berkeley (Prof. Ion Stoica et al.) 26.4.−11.5.2014. Stephan Sigg (Georg-August-Universität Göttingen) was visiting researcher at HIIT with the topic “Sensing technologies and platforms, and user activity and mental state modelling” supporting Re:Know work, 23.4.−22.6.2014 and 7.7.−22.8.2014.

Professor Wray Buntine (Monash University, Melbourne, Australia) visited HIIT (Petri Myllymäki) for a month 22.12.2014−21.1.2015. Research is ongoing with him on using topic models for scientific article data.

Equipment

The most important equipment purchase was a wall of touch-screens. The purchase required a call for tender. As the first call for tender was not successful, a second call for tender with deadline 8.12.2014 finally produced three offers and the best offer was by Multitouch Ltd., which was selected. The price was 50 245 euro + VAT.

3 Content Reports: Introduction

The project was organised into four work packages:

- WP1 Symbiotic Interaction (the user interface part led by Prof. Giulio Jacucci and coordinated by Diogo Cabral and Kai Kuikkaniemi; the psychophysiology part led by Prof. Niklas Ravaja and coordinated by Marco Filetti, Kristian Lukander and Ben Cowley)
- WP2 Intelligent Information Access (led by Prof. Petri Myllymäki and coordinated by initially Kai Puolamäki and later by Markus Koskela)
- WP3 Sensing and Analytics Platforms (led by Prof. Sasu Tarkoma and coordinated by Mohammad Hoque)
- WP4 Applications and Integration (led by Prof. Marko Turpeinen and coordinated by Kai Kuikkaniemi)

The prototypes were instrumental in driving the work and were designed to address the Tekes decision requirements. The main prototypes were the following:

- Prototype 1 displayed the use of psychophysiological measurements to influence the visualisation of search results: Depending on the stress level of the user, different amounts of information was shown. The interface used was SciNet with the scientific articles as data.
- Prototype 2 displayed a new way of visualising the search results. The prototype was first called QueryWall (NewsWall when news data from M-Brain was used) and later developed into a system called IntentStreams. The interface is made for a touch-screen with the intention that it can be used on a large wall display. News and social media data provided by the partner company M-Brain played a prominent role for Prototype 2.

In addition to these, several small project ideas were examined in so-called Fringe Projects, with the idea that these could develop into modules of some forthcoming prototypes. The Fringe Projects were the following:

- refindMail (originally called mailVis) to assist email visualisation and re-finding
- MemEx (Memory Expander) to re-find earlier results
- ImageSearch to identify similar pictures (colour was used as similarity measure)
- Team Building to match interests of students with expertise of professors

As the scientific article database has limited IPR rights, we also downloaded the open arXiv scientific article database, but arXiv is heavily centred on physics.

In the following are reports for each WP and the two prototypes. The Fringe Projects are reported in short as part of the WP reports.

4 Content Report: WP1 Symbiotic Interaction

The objective of WP1 was to implement the front-end of the symbiotic human-information interfaces. On the one hand, this included ways to integrate the physiological indicators of the users' cognitive states into the systems; on the other hand, this included the user interface design. The work included a study on predicting term relevance from brain signals, two particular research studies (attention level and motivational intensity) and software development (MIDAS). Also EDA equipment was tested. WP1 was in central position of realising Prototype 1, which is reported in a separate section.

The support framework for psychophysiological work also includes literature review of the state of the art. Initially in Re:Know, an internal 'primer' document was produced to quickly cover this need. This primer is currently being transformed into a paper for publication.

Predicting Term Relevance from Brain Signals

Relevance prediction is a central challenge of Information Retrieval, as it determines the information presented to the user. We proposed Term-Relevance Prediction from Brain Signals to automatically detect relevance of information directly from brain signals. We were able to show that features automatically extracted from recordings of neural activity across 32 electroencephalography (EEG) channels and a classifier based on a multi-view feature representation showed improvement up to 17% in relevance prediction based on brain signals alone. The implication of this work is that we are a step closer in detecting users' interests and retrieving information in response to detected interests only based on brain measurements. The experiments and results are documented in [ER14].

Attention Level Experiment

The 'level' of attention, broad or narrow, global or local, paid during human-computer interaction (HCI) can be important. Knowledge workers might miss important information if they "can't see the forest for the trees". Thus the processing of global vs. local levels of visual information in HCI is important to understand and measure. The literature suggests that these states can be distinguished by measuring brain activity using EEG.

To induce changes in attention level we developed a novel protocol, based on the so-called Wisconsin Card Sorting Task (WCST) test. The subject has to deduce the rule currently in use based on feedback (right/wrong) given after the selection has been made.

We recorded psychophysiological information from 25 subjects while they performed the task with a full range of sensors: EEG, Electrodermal Activity (EDA, using both wired and wireless methods), Electrocardiogram (ECG) and Electrooculogram (EOG). Behavioural results reinforce the idea that global attention works faster and more accurately than local attention, except when the task is very difficult; also we show that subjective perception of task difficulty can depend on the characteristics of the task performance, which is important for UI design. These results are published in a forthcoming paper.

Motivational experiment

Motivational intensity has been previously linked to information processing. In particular, it has been argued that affects which are high in motivational intensity tend to narrow cognitive scope (e.g., people who are highly motivated pay more attention to central rather than peripheral information), whereas affects low in motivational intensity broaden cognitive scope. We conducted an experiment to verify if this applies to visual search. The aim is to help in the design of future visual interfaces that take into account the motivational state of the user, in order to present information in an "optimal" way.

The experiment was based on a previous study [GH10], in which words were presented either in the centre of periphery of the screen following a manipulation of motivational intensity. In the present experiment, instead, information was organised in relatively more sparse or concentrated approaches, with or without categories pre-identified by the system. More in detail, we displayed keywords in either a list in alphabetical order (concentrated, no identification), a random scatter (sparse, no identification), a random scatter with words coloured according to their category (sparse, with identification), two visual clusters, each containing words belonging to the same category (concentrated, with identification). In addition, we also displayed keyword in two visual clusters identifying category (similarly to the last arrangement that was mentioned) but with an "outlier" word which was swapped between clusters (i.e., a word was assigned to the "wrong" cluster). Hence, in total, we employed five different keyword presentation style conditions, named: *list*, *random*, *colours*, *clusters*, *clusters + outlier*.

One of the main observations of the present experiment was that motivational intensity (in both directions) reduced visual attention during the keyword search task, as predicted. Eye tracking data suggested that participants employed a different approach in visual search when motivation was increased: more attention was allocated to the central areas of the screen and more time to the reading of individual words. Moreover, high motivational intensity increased arousal (measured as an increase in skin conductance and heart rate). These results will be included in an upcoming paper, which will cover additional results not mentioned here.

In the continuation Re:Know 2, we will utilise MIDAS modules developed to assess heart rate and skin conductance in real time. This will enable us to monitor arousal over larger periods of time.

MIDAS platform for online physiological analysis

One of the key contributions was the development by FIOH researchers of the MIDAS framework, short for Modular Integrated Distributed Analysis System. The MIDAS framework provides the building blocks for construction of distributed online data stream processing systems. MIDAS makes it possible to process raw data streams, extract features, perform machine learning tasks and make the results available through a REST JSON API for easy integration with various applications. The MIDAS framework is agnostic with regard to the type of data stream and can be used in multiple domains. MIDAS was released in August 2014 under the MIT open source license and can be freely downloaded from GitHub.

The MIDAS framework was used as part of Prototype 1, allowing psychophysiological signals to be interfaced with the SciNet search engine. The user's level of arousal was determined from the EDA and this signal was used to manipulate the amount of information shown in the search user interface.

A manuscript describing the architecture and general applicability of MIDAS has been prepared [HT154].

EDA Sensor Test Experiment

Electrodermal Activity (EDA) is a well-established recording method and numerous devices exist for performing measurements. With the recent increase in quality and popularity of wearable biosensors, multiple portable EDA devices are now available. However, before such wearable devices can be used for research or clinical purposes, it is important to compare their performance and accuracy with respect to laboratory-grade equipment proven to be suitable for conventional settings.

The Moodmetric EDA Ring (www.moodmetric.com), is a consumer-grade EDA sensor and not strictly intended for laboratory studies. Comparison was performed by recording data from both the Moodmetric EDA Ring and a traditional laboratory EDA sensor during the WCST experiment, and analyzing the similarity of the outputs. We found that the ring sensor was reasonably accurate and robust: similarity analysis between the ring and the lab-grade sensor produced relatively high scores of 83%. This work has been published as [TC15].

Fringe project refindMail

To cope with the increasing number of emails, most users rely on typed queries and intrinsic information provided by email clients that make the inbox more compact (e.g., using email threads). With refindMail (originally called mailVis) we took an alternative design approach and investigated the potential of using visualisation to support email refinding through the design and evaluation of our prototype refindMail. refindMail also features contact and keyword suggestions to act as memory cues and can be used for browsing the email collection. The design of the prototype and a novel interactive feature has been presented in [ST14].

We evaluated RefindMail in a two stage user study in which 11 participants performed realistic refinding tasks, which were gathered through a month-long diary study, using their own email collections. We got encouraging results regarding the adoption and perceived utility of features, with the recommendation and timeline visualisation features being used in 47% of the cases and contact recommendations reported to be to be relevant for 35 search tasks (53%) and supplemented missing information at 13 tasks (20%).

5 Content Report: WP2 Intelligent Information Access

Machine learning is a central component in symbiotic interaction. The main goal of the work package was to convert the existing strong machine learning research expertise into novel algorithms and methods for practical systems and prototypes for symbiotic human-computer interaction.

The results of the work package include novel contributions in modelling dynamic user models, measurement data, and context. In *dynamic user modelling*, we apply *reinforcement learning* methods to use information about the user and her actions (search interests, behaviour, background, context, input from physiology etc.) to find the best displays of information to the user, in a way that the user is able to see and manipulate the model parameters. In particular, new methods for enabling the user to interact with the user model in huge and complex

information spaces were introduced in [KG14a, KG15]. These methods improved the controllability and predictability of the search system, leading to increased task performance and user satisfaction.

Measurement data modelling involves taking in the sensory measurements, possibly from several channels and types of devices to produce usable information for the rest of the project (e.g., “user is under high cognitive load”). Related machine learning research topics include *time series analysis* and *data fusion*. The time series related topics consist of modelling, exploration, and evaluation of time series data, and combining such data from several sources. The calculation of confidence bands for time series data is an important problem, as this allows, e.g., detection of outlying time series. The calculation of confidence bands for time series is a nontrivial problem due to autocorrelation. A novel algorithm solving this problem in a principled way was developed [KP14]. The methods for data fusion include randomisation-based methods which allow for use of efficient algorithms and definition of useful patterns. The prediction of relevance judgements for text stimuli from brain signals was studied in [ER14]. The results showed significantly better performances than the random baseline. As a practical application, the proposed high-precision relevance predictor can construct meaningful sets of terms for unknown topics and unseen participants.

Several topics in *data mining* were studied in the work package. We developed the *GoldenEye* algorithm [HP14] for investigating how a black box classifier utilises the structure of a dataset when making predictions. This allows groups of interacting attributes to be found. The method was successfully utilised in the analysis of a real medical dataset on adverse drug events [HP15b]. Clustering is a widely used unsupervised learning technique. However, clustering results are often presented without any guarantees of the robustness. The clustering result may change due to systematic errors (due to the stochastic nature of the clustering algorithm) and due to random errors (variation in the dataset). We presented in [HP15a] a novel method termed *Core Clustering* that provides statistical guarantees on the clustering result in terms of co-occurrence probabilities. The method makes it possible to find statistically robust clusters using both unsupervised and supervised learning methods.

In biomedical datasets it is of interest to consider the joint information carried by multiple signals. Furthermore, it is often also often of interest to consider the shared variation between different individuals in some experimental condition, e.g., what are the psychophysiological responses to some event on the group level. Many existing techniques are limited to only two datasets and to linear relationships between the datasets for which the joint information is to be extracted. We developed the novel *CoCoReg* algorithm (Common Components by Regression) for the extraction of shared variation using chains of regression functions [KH15]. This is a computationally efficient method for the extraction of shared variation fast, with a high level of adjustability, as the used regression functions can be freely chosen.

In *context modelling*, we use the user’s history as context information by allowing the user to use efficiently any point in his or her history as a context. The first part of the problem is how to use the heterogeneous context data (user history, physiology, etc.) in predicting the suitable navigation choices for the user; the other part is how to implement this idea in a way that the user can fluently navigate in the history timeline. User’s history is used as a relevance feedback source in the MemEx system described below.

Recommender systems face difficulty in cold-start scenarios where a new user has provided only few ratings. Improving cold-start performance is of great interest. At the same time, the

growing number of adaptive systems makes it ever more likely that a new user in one system has already been a user in another system in related domains. To what extent can a user model built by one adaptive system help address a cold start problem in another system? We compared methods of cross-system user model transfer across two large real-life systems. Our results show strong improvement in cold-start recommendation by transferring open user models, and also reveal why explicit open models work better in cross-domain context than traditional hidden implicit models. The work was conducted in cooperation with Prof. Peter Brusilovsky's team at University of Pittsburgh, USA. The experiments and results are documented in [WP15].

The results of WP2 also include a common system architecture consisting of modular components that can be reused in several pilots and applications. Furthermore, a major effort in the work package involved processing of data sets into a form that would be re-usable in the common architecture. The data sets include about 60 million scientific articles, and news and social media data from M-Brain.

Fringe projects

The first versions (v1.0 and v1.1) of *MemEx* was implemented allowing relevant feedback on historical search sessions to retrieve new documents. The current retrieval engine is currently based on search history only. The use of physiological and gaze signals is under development.

- V1.0 Enables users to update results by given relevance feedback to historical search sessions and documents viewed during previous search. See Figure 5.1.
- V1.1 has a new timeline interface which can potentially take user's zooming actions as implicit feedback. See Figure 5.2.
- A feasibility study has also been conducted to test the backend machine-learning algorithm, using v1.0 as a tool to collect ground truth data.

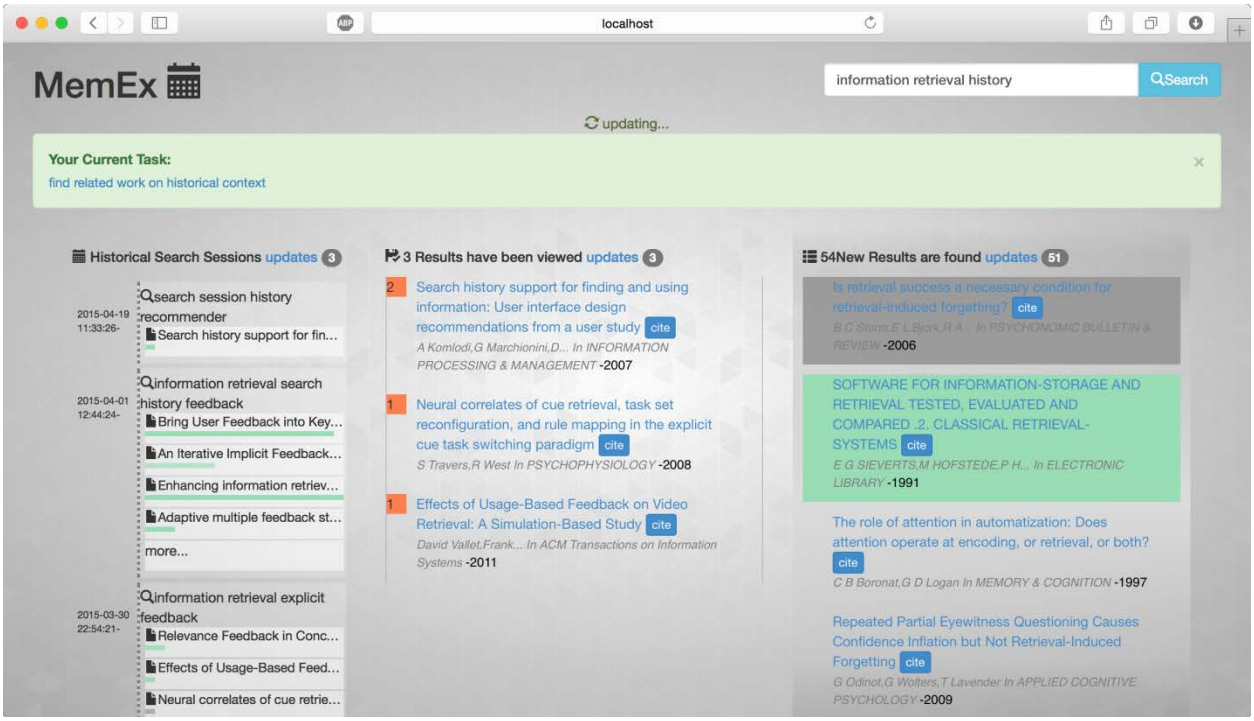

Figure 5.1. A screenshot of MemEx v1.0.

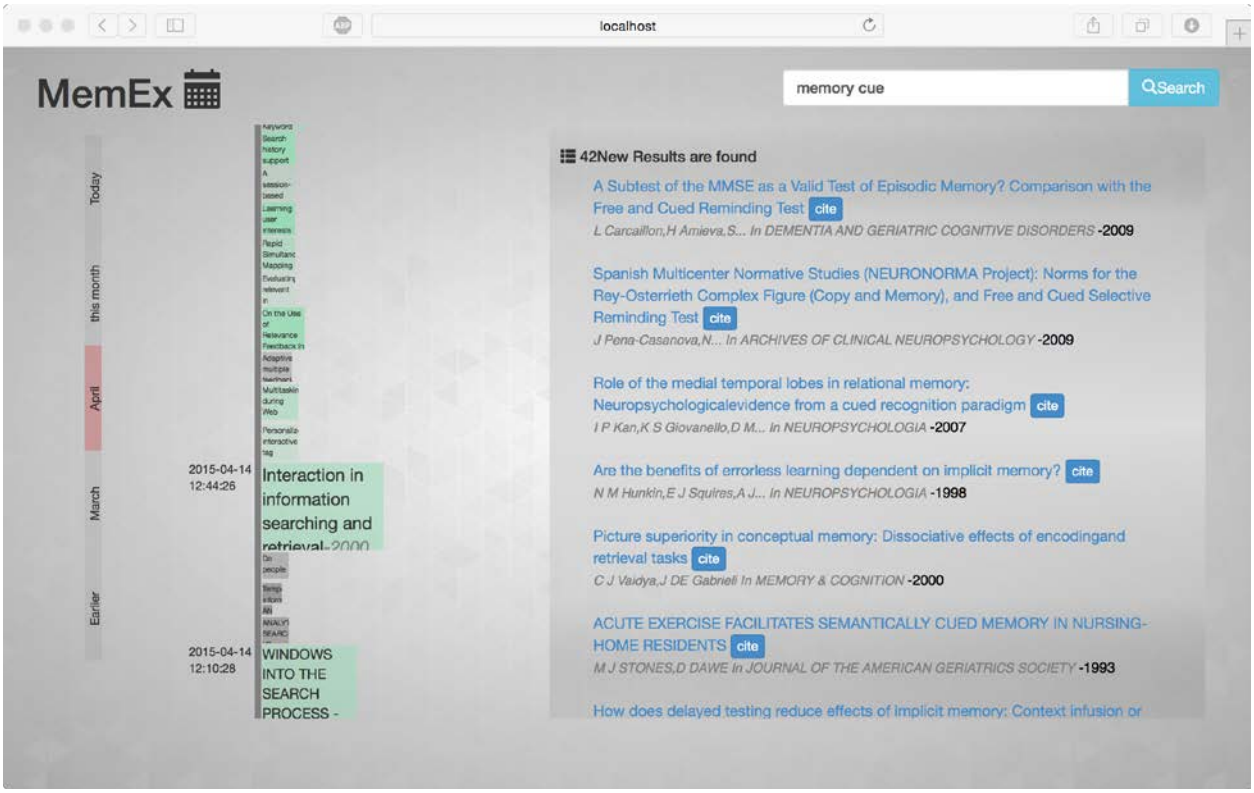

Figure 5.2. A screenshot of MemEx v1.1.

The first version of the *Image Search* has been implemented that allows the system to provide relevant images to the user through a combination of relevance feedback and content based image retrieval. The current retrieval engine is currently based on Gaussian Process Bandits and uses only colour features [GH14, HT14]. The ongoing work on the system involves combining implicit relevant feedback from mouse clicks with explicit feedback coming from the eye gaze, replacing currently used colour features with deep neural network features, developing a new interface that allows the user a “peek” into the future.

The first version of the *Team Building* student-supervisor matching system has been implemented [GI15]. The system allows students to look for possible dissertation supervisors by ranking keyphrases suggested by the system based on the initial query. Based on the keyphrases selected by the user, the system next re-ranks all the names of the faculty members who submitted their research papers to the system. Currently we are working on an extension of the system that will allow the faculty members to manage their profile in the system.

6 Content Report: WP3 Sensing and Analytics Platforms

The goal of this work package was to facilitate the distributed computing environment for Re:Know Prototypes. This includes the stored big data preprocessing frameworks, developing necessary machine learning algorithms for the distributed computing platform to operate on the stored data and streams. A cluster of seven servers were set up for fast processing.

Big data preprocessing: Huge volume of unstructured data processing for big data platforms , such as Spark and Hadoop, is a challenge. In the Re:Know project, we are dealing with Terabytes of unstructured data from M-Brain, Wikipedia dump, and arXiv. We are working toward a Spark based data pre-processing framework. In the first phase of the project, we implemented data preprocessing and key-word extraction using Spark for using with Lucin, Solr, SciNet, and IntentStream systems.

Stream processing: Along with the stored content large-scale processing of streams is another challenge in Big Data. The key issue here is dynamics of individual streams. Therefore synchronisation of them is important. During the first phase of the project, we developed a Distributed Physiological Signal processing framework, where an individual machine can receive and process thousand of streams. The framework is being updated with various APIs for physiological information modelling.

Linear systems: Linear systems are the building blocks of different machine learning algorithms. We developed basic matrix related algorithms to facilitate the development of machine learning algorithms. This is being used in further ongoing research. Several Master Theses are in preparation.

A survey of machine learning tools and platforms for big data analysis is in preparation.

7 Content Report: WP4 Applications and Integration

The WP4 objectives were:

- Integration of work in different WPs
- Understanding the larger design challenges related to knowledge work applications
- Utilising co-design and user groups to produce designs for application prototypes

During the 20 months of the project, WP1 focused on delivering these challenges through five work tracks:

- Continuous coordination of work between different WPs
- Co-design work with different stakeholder groups
- Trends & Needs in knowledge working tools
- Open ecosystem modelling
- Organising various stakeholder and internal project group events for sharing knowledge and integrating work

Continuous coordination work was realised by Patrik Floréen and assisted by Kai Kuikkaniemi. This work consisted of regular weekly project meetings with principal researchers, quarterly big researcher meeting with the full project staff, board meetings and smaller focused meetings.

Co-design work was led by Anu Kankainen who orchestrated co-design workshops about future knowledge work with a developer team and a law firm, as presented at the Steering Group meeting 9.5.2014. The co-design work was further on carried out also in our internal project meetings and it provided critical material for the development of the DiMe concept.

The *Trends & Needs* work has been finalised as a report [LK15] and will be published in relation to an event 4.8.2015. It has been iterated several times during the project and disseminated internally for providing insight for the project team. The trends & needs in knowledge working

tools work is led by Tuukka Lehtiniemi. The Digital Me (DiMe) concept in Re:Know 2 has been inspired by the Trends & Needs work.

Antti Poikola and Tuukka Lehtiniemi have been responsible for producing an *open ecosystem plan* for the knowledge working tools. This work is also aligned with the DiMe model and MyData approach [PK15]. The goal of the open ecosystem model is to identify the core enablers that are needed to create a knowledge work environment that supports open innovation, flexible integration of new tools, portability of data and human-centric control for data. The report [LP15] was distributed to the Steering Group in their final meeting 30.4.2015.

During the project the team has been *collaborating* for example with the Quantified Employee work and organised two larger *events*:

- The "Key Elements in Knowledge Work" workshop at FIOH 26.2.2014 produced ideas on future development of knowledge work and new solutions for handling complex decision problems over dispersed large datasets, with external experts Ossi Kuittinen (Sovelto, Open Knowledge Foundation), Ville Peltola (IBM), and Rauli Kauppinen (HUS) participating. The same topics have been addressed in multiple industrial/occupational group discussions by Kiti Müller as part of her interactive lectures on brain work.
- A public seminar "Information systems at knowledge work: from today into tomorrow's world" was arranged by Re:Know together with the FlowIT – virtaa IT-hankintoihin project on 26.11.2014 at the Finnish National Museum with 71 participants. The afternoon was dedicated to Re:Know. The presentations are available on the website www.reknow.fi/seminar-2014.

In the beginning of the Re:Know we coined a concept *knowlon* to function as a simple design core for our action. The knowlon is a living folder, the atom of knowledge work, digesting for example emails, documents, browsing data, application use data and contextual information including biosignals, and creating an information graph out of this input together with user defined elements, which is then continuously fed to a recommender system for retrieving related information (links, documents, people). The knowlon has many instantiations: project, session, task etc. The knowlon concept was iterated and best parts of it evolved into DiMe.

The DiMe concept originated from the workshops organised in WP4. DiMe was developed to match the needs of future knowledge work identified in the WP4 trends and needs work especially automation of work, diversity of data and BYOT/BYOD. The DiMe concept simplifies the data management, provides solution for human-centric control and portability, and balances the current closed information ecosystems with an open ecosystem approach.

MyData is related to the DiMe concept since they are both focused on human-centric organisation of personal data. HIIT has ongoing research on MyData in other projects than Re:Know. MyData research supports DiMe development, since it captures the data control, legal and privacy questions. MyData is relevant for Dime also in how it can enable data acquisition from existing large data resources.

WP4 supported the design of Spaceify and HIIT's live participation research, so that ultimately they could be integrated more into the Re:Know 2 research. Spaceify (www.spaceify.org) is an open source edge computing platform that enables interaction between personal devices and resources in the physical space. HIIT's live participation research is focused on how audience can participate in performances by using personal computing devices. In Re:Know 2, live participation work is researched in WP4.3 and Spaceify is used for example as a platform in intelligent meeting room applications in WP3.

8 Content Report: Prototype 1

Prototype 1 was developed based on the SciNet search engine. The prototype utilised a skin conductance signal extracted from EDA (Electrodermal Activity) using the Moodmetric ring. Development of the prototype was assisted by the MIDAS platform developed by FIOH. For the prototype, a skin conductance node (node is a processing module, in MIDAS terminology) was developed to implement a method to decompose a raw EDA signal into a “phasic driver” representing an instantaneous measure of arousal; this is described in [BK10].

Figure 8.1. Example of the prototype functionality during a test session. A screen capture from the user’s screen in shown in the top left quadrant. The video recorded from the session is shown on the right. The bottom panel shows the amount of arousal calculated by the EDA node (phasic activity, normalised between 0 and 1) for slightly more than 400 seconds (bottom horizontal axis). The vertical lines in the figure represent arousal artificially elicited to test the functionality of the prototype. A red line indicates that high arousal was elicited, a green line that moderate arousal was elicited and blue lines indicate low arousal. This was done via autobiographical memories (respectively related to an angry, happy or relaxing event). The black line indicates the current time index. In this particular moment, the user was experiencing low arousal, following an event of higher arousal. The number of keywords on the screen was adjusted according to the current level or arousal.

Previous literature suggested that the amount of arousal modulates the amount of information a user can process simultaneously [LS07, SK88]. From this assumption we developed a system that changed the amount of information shown on a computer screen. The information shown, in this context, was the amount of keywords presented in the SciNet radar view. The prototype allowed to specify in which direction to adjust the amount of information depending on the input physiological signal (whether to display more or less keyword as arousal increases). With the prototype, we achieved live adaptation of a system based on a physiological signal (EDA), interpreted to be associated to a specific cognitive state (arousal). See Figures 8.1 and 8.2 for an overview of the implementation.

Scientific findings related to the prototype were tested in an experiment of motivational intensity (designed in parallel with the prototype and run after the prototype was completed). The experiment confirmed that arousal (which is commonly elicited by increases in motivational intensity) modulates the amount of information one person can process. This will be documented in a forthcoming paper. However, in-house testing of the prototype suggested that adapting an interface in real time in a direct response to arousal might not be an optimal strategy. This is because such a tight feedback loop could induce users into an unwanted state (for example, a spurious increase in arousal causes the number of keywords to increase, which will in turn increase arousal, and so on). Instead, we will continue to use the tools developed for the prototype, such as the EDA MIDAS node and the presentation code created for the motivational intensity experiment in other contexts. Arousal measured in this way can be stored in the DiMe (Digital Me) storage central in Re:Know 2. This information can be used in various ways, for example, it can be recorded during a meeting to estimate the overall arousal between a group of people (Task 4.3) or stored to assess well-being on a longer time scale (Task 5.2).

A video “Video displaying adaptation of information seeking on the basis of psychophysiological measurements” explaining the prototype can be found on the project webpage www.reknow.fi/results/videos.

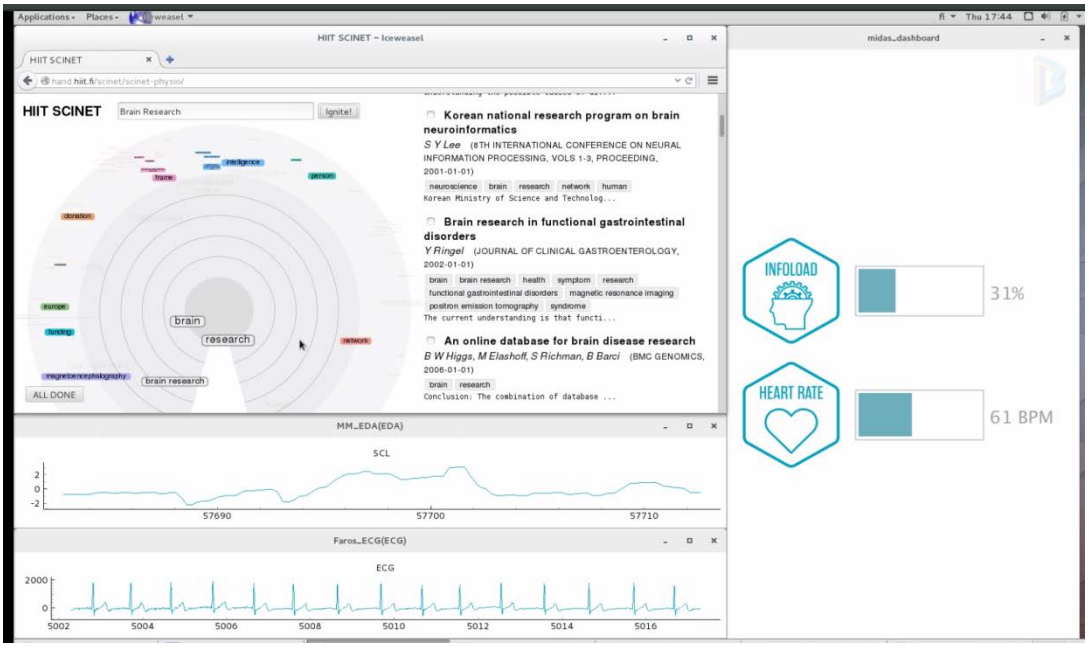

Figure 8.2. A graphically updated version of Prototype 1 was also produced.

9 Content Report: Prototype 2

The Prototype 2 works with a completely different user interface. During the development, different names were used, such as Exploration Wall [KR15], and QueryWall. The QueryWall interface exploits large multitouch screens to allow users to simultaneously compose and combine multiple queries. Several meetings with M-Brain was conducted to get familiar with the search problems they have. We also had extensive news and social media data available from M-Brain. When news data was used on QueryWall, we called it NewsWall. We further developed the prototype into a system called IntentStreams, inspired by recent studies showing that search is usually a multi-task activity [LO11, KO13] and that parallel browsing and branching are very prevalent [HW10, HL12], although surprisingly not well supported by current search interfaces.

IntentStreams was designed to support parallel browsing and branching during exploratory search of news without needing to open new tabs. The news domain was chosen due to the collaboration with M-Brain in order to support the need for better exploration of news, which is a core need of this company. However the prototype design principles were meant to support a wider range of applications not limited to the news field.

IntentStreams, see Figure 9.1, presents parallel streams of searches, where each stream displays results as documents and keywords and displays the underlying queries as keywords representing the search intent of the stream. New streams are initiated by users, where the search intent of a new stream is initialised either by typing a traditional query or by dragging keywords available in any of the streams. In each stream, in addition to the user-chosen keywords the system proposes other relevant keywords and orders them vertically by their predicted relevance. The users can change the relative relevance of keywords in the query intent of each stream, and easily branch new streams, by simply dragging keywords [AK15]. A video “IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search” about the prototype is available at the project website www.reknow.fi/results/videos.

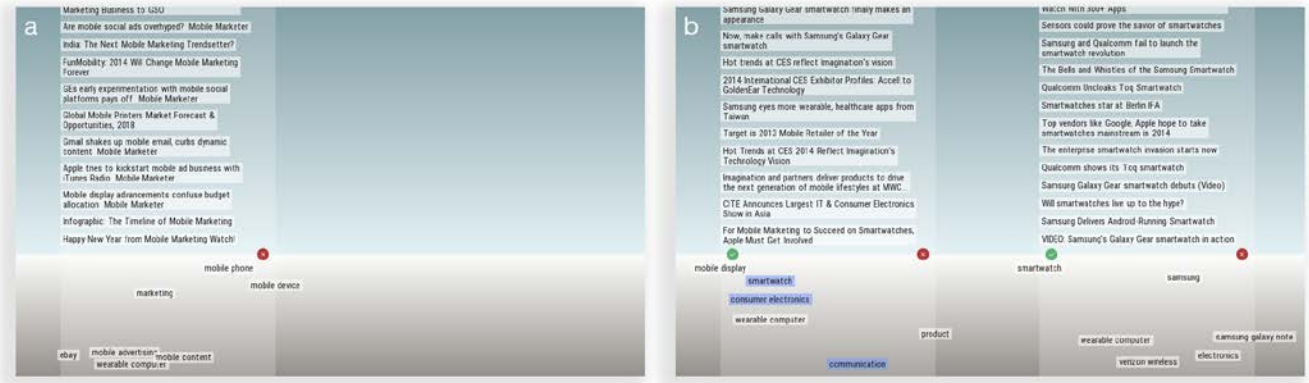

Figure 9.1. (left) A first query (in this case “mobile phone”) returns a search stream composed of news articles most relevant to the query, as well as a set of most relevant keywords extracted from a larger set of related articles. (right) The user can then modify the weight of the keywords by sliding them vertically, the stream will then refresh, updating articles and keywords accordingly. If dropped outside their initial stream, keywords can either trigger a new search stream or be passed to an already existing parallel stream.

We evaluated the system to find out if and how IntentStreams supported parallel browsing and branching behaviour. Participants were 13 volunteers (4 females) with a mean age of 28.4 (SD=4.05). We compared the system against a baseline system with an interface similar to a traditional Google search interface on a 20 minutes long exploratory search task. To evaluate the system, we connected it to a news repository from M-Brain of English language editorial news articles crawled from publicly available news sources from September 2013 to March 2014. The database in total has more than 25 million documents. The documents were originally collected for monitoring media presence of numerous interested parties, and hence the collection had wide topical coverage. The baseline system was connected to the same news repository as IntentStreams. In the baseline system, users could type queries and receive a list of relevant news articles. To start a new parallel query, a new tab would have to be opened.

Parallel search was supported in IntentStreams: Results show that users created more parallel streams than opening new tabs.

Branching was supported in IntentStreams: In IntentStreams more queries and parallel streams were created through branching. Figure 9.2 presents a visual representation of a participant search behaviour, showing the difference between the linear search behaviour in the baseline and the more articulated search behaviour in IntentStreams. We also observe that IntentStreams supports more exploration through more queries.

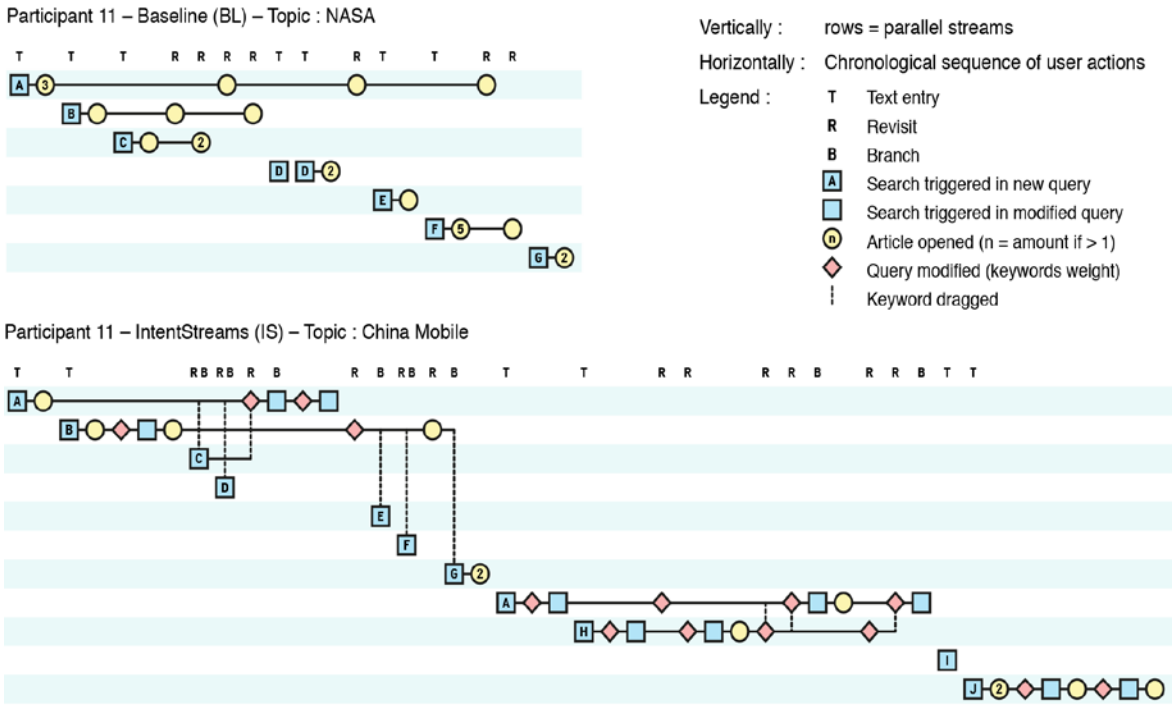

Figure 9.2. Example of branching behaviour from the case study: top - Baseline; bottom – IntentStreams.

10 Summary of Outcomes

Publications

We had an article in the top journal Communications of the ACM [RJ15].

The 21 peer reviewed publications of the project are references [AK15, BE15, ER14, GI15, GH14, HT154, HP14. HP15b, HT14, GS14, KG14a, KG15, KG14b, KR15, KH15, KP14, RJ15, RP14, ST14, TC15, WP15].

In addition, we produced an open ecosystem report [LP15] and a Trends and Needs white paper [LK15], the latter published as part of the Re:Know 2 project.

Software releases

MIDAS: The Finnish Institute of Occupational Health (FIOH) has released the MIDAS (Modular Integrated Distributed Analysis System) system on 13 August 2014. MIDAS is a system for online analysis of streaming physiologic signals and other time series and for easy integration of features (e.g., psychophysiological metrics) into online machine learning frameworks. The MIDAS system consists of distributed processing nodes. A central node implements as RESTful JSON API which can be used to access information from the processing nodes. The MIDAS system is published under the MIT license and can be downloaded from github. More information can be found in the MIDAS wiki.

Spectral Mass: Helsinki Institute for Information Technology (HIIT) released Spectral Mass on 6 October 2014. Spectral Mass is an EEGLAB Extension used to compute the probability that two EEG datasets are similar in terms of *both* power (ERSP in EEGLAB terms) and coherence (ITC). The source code is released under the GPLv2 license and can be accessed from the HIIT github repository, for more information see the Spectral Mass web page.

Maui 2.0: Helsinki Institute for Information Technology (HIIT) released Maui 2.0 on 10 October 2014. Maui 2.0 is a keyword extraction tool, built on the old Maui, a topic indexing tool with several improved added. The source is released under GPLv3 and can be accessed from the HIIT github repository.

MSc Theses

Zhe Xie, “From Exploration to Sensemaking: An Interactive Exploratory Search System,” Aalto University 13.3.2015. Instructor: Jaakko Peltonen. Supervisor: Samuel Kaski.

Public videos

The following videos are available at the project website www.reknow.fi/results/videos:

- About Prototype 1: “Video displaying adaptation of information seeking on the basis of psychophysiological measurements”, https://youtu.be/fYL2LlH5bcU
- About Prototype 2: “IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search”, https://youtu.be/jHM_WKopE2E

Media coverage and press releases

- A press release about the project launch was published 29.8.2013. A news article about the project was also posted on the University of Helsinki homepage.

- Patrik Floréen and Kiti Müller were interviewed about search engines in the radio programme Tieteen taustapeili in Radio Suomi on 27.9.2013.

- Tiede Magazine 3/2014 (11.3.2014), pp. 50-55, “Hyvän & pahan tiedon puu”. This received much attention and also draw the interest of HIS Oy for the commercialisation that led to Etsimo.

- Aalto University Magazine Nr 10 (March 2014), pp. 12-17, ”Datamerestä olennaiseen”.

- Suomen Kuvalehti 31/2014 (1.8.2014), “Informaatiokupla puhki: Suomalaistutkijat kehittävät uudenlaista hakumenetelmää.”

- Helsingin Sanomat 9.11.2014, pp. D4-D5, “Isä pelaa tabletilla ja puhuu puhelimeen”. This was not directly about Re:Know, but Patrik Floréen and Kai Puolamäki were interviewed, but in the final text only Floréen was mentioned and the reference to Re:know was also removed.

- Tekniikka ja Talous 12.12.2014, page 10, ”Googlen haastaja Helsingistä”. This article was about Etsimo.

- A press release about SciNet and Etsimo was published 27.1.2015.

- Helsingin Sanomat 31.1.2015, pp. D10-D11, “Suomalaiset hakukoneet löytävät sinne, minne Google ei”. This was a major 2-page article.

Talks

In 2013: Kiti Müller has given several lectures in which the Re:Know project has also been shortly presented: Duodecim meeting at Hanasaari on 5.9. , Tekes-Liideri programme seminar on 2.10., Tampere area Leadership morning on 18.10. Re:Know and its relevance to future work was also discussed with NordForsk on 25.9. during their visit to FIOH and with Dr David Michaels, director of Occupational Health and Safety (Administration) and assistant secretary of labour of the US government on 21.11. Tekes arranged a morning coffee event in which all three large strategic openings were presented on 23.10. Kiti Müller and Petri Myllymäki presented also the Re:Know project at the DIGILE Foresight Seminar on 27.11.2013.

In 2014: Kiti Müller mentioned the project in several presentations (8.2. Aivojen venymiskyvyn rajat, Mehiläinen Oy konserni, Paasitorni; 21.3. Aivoterveyttä terveysalan ammattilaisille, PlusTerveys konserni Wanha Satama; 25.4. Enemmän luovuutta – vähemmän kilpailua? Tutkimustyön tulevaisuus. Professoriliiton ja tieteentekijöiden liiton seminaari, Pörssitalo) and also in Radio Helsinki on 23.1. “Viekö Teknologia työpaikat”. Patrik Floréen presented 29.9. Re:Know to the National Library of Finland in the “Smart information retrieval workshop” in Mikkeli. Patrik Floréen and Kiti Müller gave 30.9. a keynote to PhD students and postdoctoral researchers at the iCareNet meeting in Tampere.

Several Re:Know persons participated in the Avoin Suomi event 15.-16.9.2014. We tried to get an own stand for Re:Know at the event, but there was much interest and we ended up number 2 in the waiting queue (number 1 on the waiting list was Digile).

Patrik Floréen and Kiti Müller participated 14.10.2014 in the panel discussion "Tiedon vallankumous, robotisaatio ja työelämä" at Tiedekulma, University of Helsinki. The panel was led by Jyrki Kasvi and the third participant was Veikko Eranti.

In 2015: Patrik Floréen presented 26.2. Revolution of Knowledge Work to parliamentarians of the SDP party at the Chancellor's dinner at University of Helsinki. Patrik Floréen talked 4.3. about leading research projects using the Re:Know project as case at an event for the management (rector etc.) of the University of Helsinki. Patrik Floréen talked 21.3. about the future of knowledge work in his invited dinner speech at the student association Spektrum rf's 82nd annual party. The speech had been published in the student association's magazine Spektraklet.

Events for the public

The "Key Elements in Knowledge Work" workshop at FIOH 26.2. produced ideas on future development of knowledge work and new solutions for handling complex decision problems over dispersed large datasets, with external experts Ossi Kuittinen (Sovelto, Open Knowledge Foundation), Ville Peltola (IBM), and Rauli Kauppinen (HUS) participating.

A public seminar "Information systems at knowledge work: from today into tomorrow's world" was arranged by Re:Know together with the FlowIT project on 26.11. at the National Museum with 71 participants. The afternoon was dedicated to Re:Know. The presentations are available on the website www.reknow.fi/seminar-2014.

Organisation of a scientific workshop

Re:Know participated in organising the International Workshop on Symbiotic Interaction 30.−31.10. in Helsinki, with prominent speakers (David Kirsh, University of California San Diego; Roderick Murray-Smith, University of Glasgow), see symbiotic2014.org. The conference proceedings were published by Springer. The next workshop is in Berlin in October 2015.

Commercialisation: Etsimo Oy

On the basis of the article in Tiede magazine, HIS Oy of University of Helsinki took initiative to a US patent application together with Aalto University. A company Estimo Oy was founded, www.etsimo.com. CEO is Thomas Grandell. Etsimo was presented at Slush 2014 in Messukeskus in Pasila 18.-19.11.2014 in the booth of University of Helsinki and was also presented in Yle TV1 Aamu-TV on 19.11.2014.

Re:Know Publications

[AK15] Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, Giulio Jacucci. IntentStreams: Smart parallel search streams for branching exploratory search. In *Proc. 20th International Conference on Intelligent User Interfaces* (IUI '15, March-April 2015). ACM, New York, NY, USA, 300-305. DOI=10.1145/2678025.2701401

[BE15] Oswald Barral, Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Niklas Ravaja, Samuel Kaski, Giulio Jacucci. Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. In *Proc. 20th International Conference on Intelligent User Interfaces* (IUI '15, March-April 2015). ACM, New York, NY, USA, 389-399. DOI=10.1145/2678025.2701389

[ER14] Manuel J.A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Predicting term-relevance from brain signals. In *Proc. 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (SIGIR '14, July 2014), ACM, New York, NY, USA, 425-434. DOI=10.1007/s10618-014-0371-0

[GI15] Yuan Gao, Kalle Ilves, Dorota Głowacka. OfficeHours: A system for student supervisor matching through reinforcement learning. Poster presentation in *IUI Companion Proc. 20th International Conference on Intelligent User Interfaces* (IUI Companion '15, March-April 2015). ACM, New York, NY, USA, 29-32. DOI=10.1145/2732158.2732189

[GH14] Dorota Głowacka, Sayantan Hore. Balancing exploration - exploitation in image retrieval. Poster at the 22nd Conference on User Modelling, Adaptation and Personalization (UMAP 2014, July 2014). http://ceur-ws.org/Vol-1181/umap2014_poster_04.pdf

[HT15] Andreas Henelius, Jari Torniainen. MIDAS: An open source framework for distributed on-line analysis of data streams. Submitted for publication in 2015.

[HP14] Andreas Henelius, Kai Puolamäki, Henrik Boström, Lars Asker, Panagiotis Papapetrou. A peek into the black box: Exploring classifiers by randomization. *Data Mining and Knowledge Discovery* 28(5-6): 1503-1529, 2014. DOI=10.1007/s10618-014-0368-8

[HP15a] Andreas Henelius, Kai Puolamäki, Henrik Boström, Panagiotis Papapetrou. Clustering with confidence: Finding clusters with statistical guarantee. Submitted for publication 2015.

[HP15b] Andreas Henelius, Kai Puolamäki, Isak Karlsson, Jing Zhao, Lars Asker, Henrik Boström, Panagiotis Papapetrou. GoldenEye++: A closer look into the black box. *Proc. Third International Symposium on Statistical Learning and Data Sciences* (SLDS 2015, April 2015), 96-105. http://link.springer.com/chapter/10.1007%2F978-3-319-17091-6_5

[HT14] Sayantan Hore, Lasse Tyrvainen, Joel Pyykko, Dorota Głowacka. A reinforcement learning approach to query-less image retrieval. In *Proc.Symbiotic Interaction*, Springer LNCS 8820, October 2014, 121-126.

[GS14] Giulio Jacucci, Anna Spagnolli, Jonathan Freeman, Luciano Gamberini. Symbiotic interaction: a critical definition and comparison to other human-computer paradigms. In *Proc. Symbiotic Interaction*, Springer LNCS 8820, October 2014, 3-20.

[KG14a] Antti Kangasrääsiö, Dorota Głowacka, Samuel Kaski. Improving controllability and predictability of an interactive user model driven search interface. Poster at NIPS 2014 workshop on Human Propelled Machine Learning. Montreal, Canada, 13.12.2014.

[KG15] Antti Kangasrääsiö, Dorota Głowacka, Samuel Kaski. Improving controllability and predictability of interactive recommendation interfaces for exploratory search. *Proc. 20th International Conference on Intelligent User Interfaces* (IUI '15, March-April 2015). ACM, New York, NY, 247-251.DOI=10.1145/2678025.2701371

[KG14b] Antti Kangasrääsiö, Dorota Głowacka, Tuukka Ruotsalo, Jaakko Peltonen, Manuel J. A. Eugster, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, Samuel Kaski. Interactive visualization of search intent for exploratory information retrieval. Poster at ICML 2014 workshop on Crowdsourcing and Human Computing. Beijing, China, 25.6.2014.

[KR15] Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, Giulio Jacucci, Designing for exploratory search on touch devices. In *Proc. 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI 2015, April 2015), ACM, New York, NY, 4189-4198. DOI=10.1145/2702123.2702489

[KH15] Korpela, J., Henelius, A., Ahonen, L., Klami, A., & Puolamäki, K. Using regression makes extraction of shared variation in multiple datasets easy. Submitted for publication 2015.

[KP14] Jussi Korpela, Kai Puolamäki, Aristides Gionis. Confidence bands for time series data. *Data Mining and Knowledge Discovery* 28(5-6): 1530-1553, 2014. DOI=10.1007/s10618-014-0371-0

[LK15] Tuukka Lehtiniemi, Kai Kuikkaniemi, Antti Poikola, Matti Nelimarkka, Teppo Valtonen, Patrik Floréen, Marko Turpeinen. *Trends of Knowledge Work and Needs for Knowledge Work Tools.* Re:Know White Paper, June 2015.

[LP15] Tuukka Lehtiniemi, Antti Poikola, Kai Kuikkaniemi, Patrik Floréen, Marko Turpeinen, Ben Cowley, Kristian Lukander, Kai Puolamäki. *Re:Know Ecosystem Plan*. Re:Know report, 30.4.2015.

[RJ15] Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, Samuel Kaski. Interactive intent modeling for information exploration. *Communications of the ACM* (CACM) 58(1): 86-92, 2015. http://cacm.acm.org/magazines/2015/1/181621-interactive-intent-modelling/fulltext

[RP14] Tuukka Ruotsalo, Jaakko Peltonen, Manuel J.A. Eugster, Dorota Głowacka, Aki Reijonen, Giulio Jacucci, Petri Myllymäki, Samuel Kaski. Intentradar: search user interface that anticipates user's search intents. In *Proc. CHI '14 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '14, April-May 2014).ACM, New York, NY, USA, 455-458. DOI=10.1145/2559206.2574807

[ST14] Barış Serim, Vuong Thanh Tung, Tuukka Ruotsalo, Luana Micallef, Giulio Jacucci. mailVis: Visualizing emailbox for re-finding emails. Poster at Vis'14, November 2014. http://kar.kent.ac.uk/42075/

[TC15] Jari Torniainen, Benjamin Cowley, Andreas Henelius, Kristian Lukander, Satu Pakarinen. Feasibility of an electrodermal activity ring prototype as a research tool. To appear in *Proc. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EBMS 2015, August 2015).

[WP15] Chirayu Wongchokprasitti, Jaakko Peltonen, Tuukka Ruotsalo, Payel Bandyopadhyay, Giulio Jacucci, Peter Brusilovsky. User model in a box: Cross-system user model transfer for resolving cold start problems. In *Proc. 23rd Conference on User Modelling, Adaptation and Personalization* (UMAP 2015, June-July 2015), 289-301. http://link.springer.com/chapter/10.1007/978-3-319-20267-9_24

Other References

[BK10] Mathias Benedek, Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods* 190(1), 80-91, 2010. DOI=10.1016/j.jneumeth.2010.04.028

[GH10] Philip A. Gable, Eddie Harmon-Jones. The effect of low versus high approach-motivated positive affect on memory for peripherally versus centrally presented information. *Emotion* 10(4), 599-603, 2010. dDOI=10.1037/a0018426

[HL12] Jeff Huang, Thomas Lin, Ryen W. White. No search result left behind: Branching behavior with browser tabs. In *Proc. Fifth ACM International Conference on Web Search and Data Mining* (WSDM'12, February 2012), ACM, New York, NY, 203–212.

[HW10] Jeff Huang, Ryen W. White. Parallel browsing behavior on the web. In *Proc. 21st ACM Conference on Hypertext and Hypermedia*, (HT'10, June 2010), ACM, New York, NY, 13–18.

[LS07] Annie Lang, Nancy Schwartz, Seungjo Lee, James Angelini. Processing radio PSAs: production pacing, arousing content, and age. *Journal of Health Communication* 12(6), 581-599, 2007. DOI=10.1080/10810730701508708

[LO11] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, Gabriele Tolomei. Identifying task-based sessions in search engine query logs. In *Proc. Fourth ACM International Conference on Web Search and Data Mining* (WSDM'11, February 2011),ACM, New York, NY, 277–286.

[LO13] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, Gabriele Tolomei. Discovering tasks from search engine query logs. *ACM Transactions on Information Systems* (TOIS) 31(3), Article 14, 2013. DOI=10.1145/2493175.2493179

[PK15] Antti Poikola, Kai Kuikkaniemi, Harri Honko. *MyData - A Nordic Model for human-centered personal data management and processing*. Ministry of Transport and Communications, Publication 4/2015. http://urn.fi/URN:ISBN:978-952-243-455-5

[SK88] David M. Sanbonmatsu, Frank R. Kardes. The effects of physiological arousal on information processing and persuasion. *Journal of Consumer Research* 15(3), 379-385, 1988.

Revolution of Knowledge Work (Re:Know) is funded by Tekes, as a large strategic opening.

The participating organisations are Helsinki Institute for Information Technology HIIT (Aalto University and University of Helsinki) and the Finnish Institute of Occupational Health (FIOH).

A huge amount of information is spread out in various data silos. The current systems and search engines have inflexible views to the data and they have only a limited ability to study the large data masses, leaving knowledge workers trapped in individual information bubbles. Current systems constrain the work instead of supporting the combined potential of human creativity and the capability of computers to handle big data.

The project combines the multidisciplinary world-class expertise in machine learning, human-computer interaction, distributed computing, cognitive neuroergonomics and human factors at work, available within Helsinki Institute for Information Technology HIIT and the Finnish Institute of Occupational Health. Our objective is to develop Symbiotic Human-Information Interfaces, which pave the way for a revolution of knowledge work.

Symbiotic human-information interfaces combine heterogeneous data sources and utilise the context of use and user actions to jointly with the user determine what information is most likely relevant, and provide the user with a new type of interactive and proactive interface to the data. In the context of knowledge work, we use our know-how on both computational principles and how humans process information to develop a new information management and utilisation paradigm, enabling humans and computers to support each other optimally.