Design Towards AI-Powered Workplace of the Future

Yujia Cao[(✉)] ⓘⓓ, Jiri Vasek, and Matej Dusik

Konica Minolta Laboratory Europe, Brno, Czech Republic
{yujia.cao,jiri.vasek,matej.dusik}@konicaminolta.cz

Abstract. The advances of technology have profoundly improved the way people live and work. However, accompanying fast-paced technological development is information overload, which can minimise our capacity for cognitive processing and our ability to make quality decisions. We conducted extensive user research to identify needs and problems of contemporary office workers. Based on the insights of these real needs, the concept of a system called Cognitive Hub has been developed which supports an activity-based new metaphor for work, user state adaptation, smart enterprise search, smart transformation between physical and digital contents and multimodal interaction. Konica Minolta is thus developing Cognitive Hub as a platform that will serve as a nexus for users' information flows within the digital workplace. Cognitive Hub will also provide AI-based services to improve work experience and the well-being of office workers. A demonstrator was created to show the concept in action and illustrate its benefits and value for users.

Keywords: Digital workplace · Artificial intelligence · User-centred design
User state inference · Multimodal interaction

1 Introduction

We live in an era of unprecedented change. The world's population is expected to reach 7.6 billion in 2020, and the number of connected devices is expected to grow to between 20 and 30 billion by the same year as the Internet of Things (IoT) continues to mature. We are observing an exponential increase in available data and ubiquitous information that together are already causing information overload. This overloading can minimise our capacity for cognitive processing and our ability to make quality decisions. It is apparent that we have entered an era in which new human necessities are emerging: we strive to reduce the time spent searching for and memorising reliable information; we struggle with the risks associated with the security of digital information; and we battle to manage a plethora of unforeseen events and adapt to fast-paced changes around us. Artificial intelligence (AI) can provide the answer to many of these needs by offering a system of technologies that can automate information flows and help us to better identify relevant digital contents, make informed decisions and to take advantage of enhanced cognition in a broader sense [17].

Konica Minolta Laboratory Europe is embracing the AI challenges with a focus on the context where most of our skills reside: the workplace. Our proposed solution is a platform called Cognitive Hub. It is expected to become a nexus for users' information flows within the digital workplace and provide augmented-intelligence-based services to improve the work experience and more importantly the overall well-being of office workers. The concept of Cognitive Hub is built upon extensive user research of current problems and needs in the workplace (see Sect. 2), so that it is meaningful and valuable to the users. The concept delivers benefits through features including: supporting a personalised activity-based approach to work; smart enterprise search; user state adaptation; and smart transformation between digital and physical contents and multimodal interaction (see Sect. 3). Cognitive Hub is powered by various cutting-edge AI/HCI (human computer interaction) technologies including: semantic understanding of data, smart data categorisation, machine learning, computer vision, speech recognition, natural language understanding, gaze recognition and multimodal fusion. Section 4 presents a demonstrator that shows the concept of Cognitive Hub in action with a defined set of interaction scenarios.

2   Needs of Contemporary Office Worker

Following a user-centred design approach, a combination of methods has been employed including desk research, design thinking workshops, and interviews with end-users to identify problems and needs in the current workplace. The interview questionnaire consisted of 51 questions that spread over 8 topics, including: task management, time management, information management, team collaboration, communication, productivity, work load, work satisfaction and work-life balance. The interview was conducted with 19 participants in 6 European countries. The participants had varied job positions such as: management, researcher, logistics, accountant, salesman, customer support and IT administrator. The analysis of the interviews revealed that there were 8 categories of needs, described below, that were tightly interconnected and centred around the concept of well-being (see Fig. 1).

The Need to Help with Task Management. Tasks originate from different sources, some are structured (e.g., Microsoft Project) and some are unstructured (e.g., e-mail and chat conversations). People invest significant effort in maintaining an overview of their tasks. Besides relying on memory, people use various methods and tools for tracking their tasks from paper notes and task management systems, to indicating outstanding tasks such as marking emails as 'unread'. People also spend a lot of time performing administrative tasks. When it comes to prioritisation, tasks are typically given priority based on time urgency. This means that workers execute tasks which are urgent, but not necessarily the most important. The link between tasks and their goals/purposes is generally missing or not obviously apparent to workers.

The Need to Have an Integrated Overview of Information. The information that people receive or need to remember comes from multiple sources (e.g., e-mails, passwords, chat conversations, newsletters and web page updates), which makes it difficult
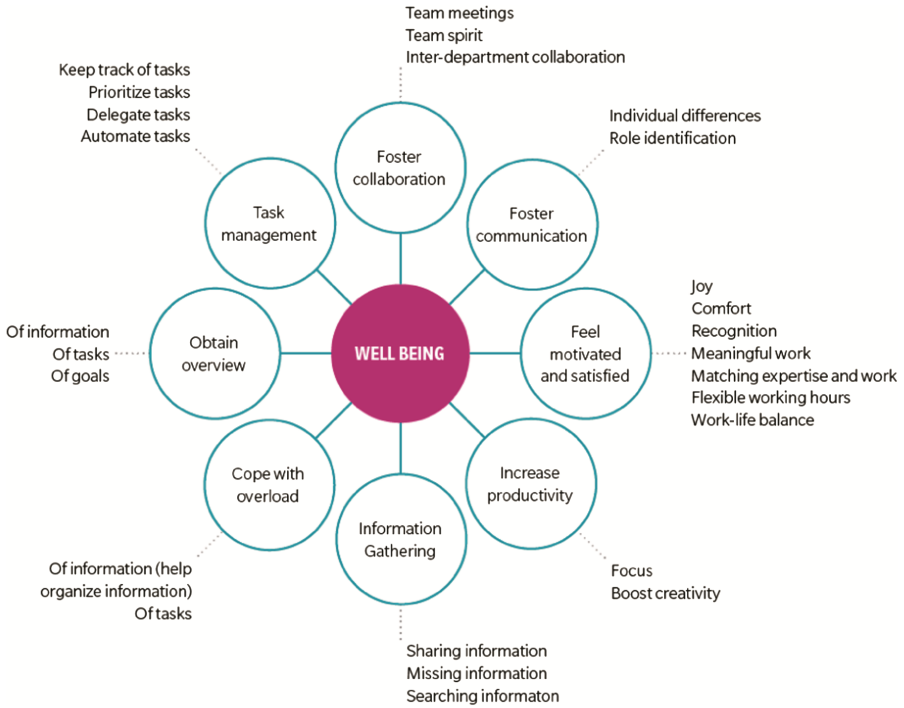
Fig. 1. Needs of office workers

to monitor, organise and integrate. Vast amounts can remain unprocessed. This means that even information which might be relevant for a particular person can remain unnoticed. The organisation of information is often managed via multiple folders (i.e., organisation schemes) which causes duplication of information and ultimately makes the whole information system hard to maintain.

The Need to Cope with Overload. Employees are overloaded with information, as there are too many sources of information to process or remember, including documents, e-mails and passwords. They find it difficult to organise, process and follow, thus missing out on the integration of all those sources. Sometimes there are too many tasks that need to be completed by the same deadline, which then results in people working overtime.

The Need to Help with Information Gathering. There is a lack of transparency in companies concerning knowledge about what other teams, colleagues or departments are working on, and this can make companies inefficient. The process of gathering information is difficult and when information is missing it causes delays. To compensate, employees ask other colleagues and managers to obtain information about contacts, processes and files. Searching for information also takes a lot of time as there are too many sources, inferior quality data (e.g., data that has not been updated) and people need to remember where the relevant and correct information is stored.

The Need to Increase Productivity. Even though people are generally at their best in the morning, they typically do 'small' tasks first so as not to forget about them, however, as a result their most productive time is not used effectively. In addition, they are often not able to focus due to unplanned interruptions, 'quick' tasks such as responding to questions, e-mails, or background noise or conversations from other people in open space office areas. To be productive they need to boost their creativity and this is perceived to increase according to a number of influences including: increased social interactions (e.g., discussions around the coffee machine); in heterogeneous work activities; being part of a great team; taking part in sports activities such as yoga; or having a dynamic environment (e.g., a café, a break room or the presence of music in the workplace).

The Need to Foster Collaboration. Even though collaboration helps people to balance workload and build team spirit, the proper organisation of meetings is difficult. Meetings often have poor agendas, lack structure, take more time than planned, or are organised with more people invited than required. Often the discussion diverts to unplanned topics that sometimes is due to a poor facilitator. In addition, finding a suitable meeting room with appropriate equipment at a time that suits all participants is also challenging. Cross-department collaboration is often tricky due to a lack of open communication, a plethora of 'own agendas', and the different goals of various departments.

The Need to Foster Communication. Problems with communication are caused by individual differences in communication style (e.g., misinterpretation of information) or different levels of language knowledge (e.g., lack of comprehension of information). Unclear role identification means that sometimes people don't know to whom they should address a request because there is a lack of information about other people's responsibilities or skills.

The Need to Feel Motivated and be Satisfied. Joy comes from social interactions and a comfortable office environment. Satisfaction with one's work comes from multiple sources including: recognition of work, professional growth, doing meaningful work, being able to utilise skills in which one excels, not having too much routine work, flexible work hours and a healthy work-life balance. People struggle with not having their desired work-life balance. Work is given higher priority over private activities so that, for example, checking work e-mails at home to prevent surprises is an all too common activity for many workers.

In summary, the well-being of an office worker is to a large extent reflected in the way work is conducted, not only within the work environment and through social interactions, but also through efficiency, effectiveness, productivity, creativity, motivation and satisfaction.

3   Cognitive Hub Concept

Taking into account the needs and requirements reviewed in Sect. 2, Konica Minolta proposes a platform called Cognitive Hub that will become a nexus for user's

information flows within the digital workplace and provide augmented-intelligence-based services to improve work experience and workers' well-being [16]. Cognitive Hub delivers benefits and values to users from the following aspects: it supports a new metaphor of work that goes beyond the desktop metaphor; it is able to detect user's emotional and cognitive states to offer corrective actions or infer user preferences; it applies semantic technology to construct a digital knowledge base about a user, to deliver a personalised experience; it supports seamless transformation between digital and physical information contents; and it allows novel means of multimodal interaction (e.g. voice, touch, gaze, gesture interaction) which is expected to be more natural, flexible, efficient and robust. This section introduces each of these aspects in more detail.

3.1   A New Metaphor of Work

A large proportion of workers' problems and needs that have been identified were related to information management. People have problems to cope with an information overload and to maintain an overview of this information that is coming from many different sources. As a consequence, it is often hard to keep track of everything they need to work on and to prioritise effectively. These problems are partially caused by the desktop metaphor which has dominated the way computers work for more than half a century [15]. Fundamentally the desktop metaphor is file-centric; it treats the computer monitor as if it is the user's desktop upon which objects such as files and applications can be placed into folder systems. To work on one task people often need to manually retrieve files from multiple folders and switch between multiple applications. In the era of information overload, the desktop metaphor is clearly no longer the best means to support the work of contemporary office workers.

Inspired by the 'lifestream' concept [15], Cognitive Hub goes beyond the desktop metaphor to offer an activity-centric metaphor of work. As shown in Fig. 2, the activity-centric metaphor supports 3 levels of work: namely strategic, tactical and operational work. It is proposed here that all people, no matter what job positions they occupy, conduct their work within these 3 levels and dynamically switch between them. However, without any supporting tools, most people are unaware of the levels and they switch between them without making a conscious decision. Cognitive Hub aims to provide a visualised tool to support people in better managing their work tasks at these 3 different levels as described in more detail below.

Strategic Work. This level is about strategically managing the 'purpose and meaning of work'. Users can define and manage their personal goals. Goals should include those work-related objectives that are typically provided by the employer, such as successfully completing a project, or increasing sales by 10%. In addition, users can also define goals to reflect their private targets and interest, such as career growth, learning a new language, keeping fit, etc. Achieving these private goals will increase the overall satisfaction and motivation in life and in turn improve work performance. A user can also set priorities amongst his/her goals. The importance of strategic work is to provide a link between what users do (tasks/activities) and what users want to achieve (goals). This way users are less likely to 'lose the purpose of work' or 'not know what to focus
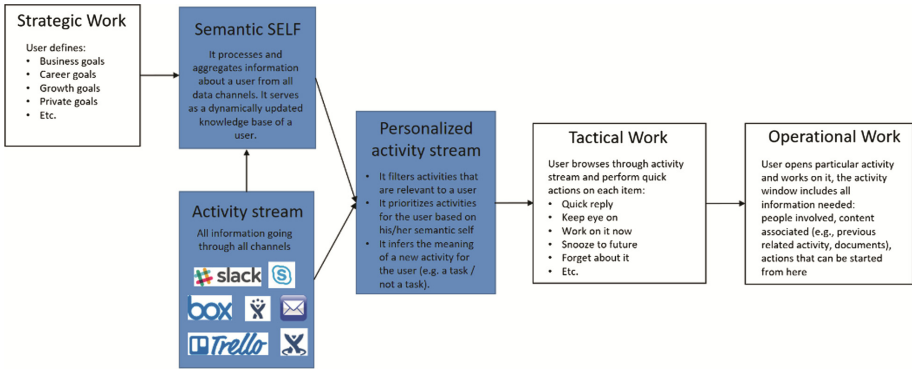
Fig. 2. The activity-centric metaphor of work

on'. Moreover, personal goals are one of the inputs to a Semantic SELF (semantic enrichment and linking framework) that is used for personalising and prioritising activities for a user.

Semantic SELF. Every user has his/her own profile, preference and a unique method of communicating and organising his/her personal enterprise environment. Semantic SELF is a dynamically updated knowledge database of a user, including all the information that can be retrieved from the digital system or inferred from the user's behaviour through the use of the digital system (see Fig. 3). Goals defined by the user are also a part of the input to a Semantic SELF.

Personalised Activity Stream. In workplace, information typically flows through multiple channels including: emails or instant messaging (e.g. MS Outlook, Skype, Slack, etc.); task management (e.g. JIRA, MS Project, etc.); and shared storage systems (MS Sharepoint, Box, Google drive, etc.). Currently, people need to monitor all these channels separately to keep track of everything that they need to do. In the Semantic SELF concept, information from all channels are integrated into one single stream. Then semantic technology is applied to each item within the stream to infer what the user needs to do upon receiving that piece of information. In such an approach the incoming information stream is transformed into an activity stream. For example, if a user receives an email with the subject 'presentation', body text 'Please review it by this Friday' and a MS PowerPoint file as an attachment, the associated activity item would have a title 'Review presentation' and a due-date of 'this Friday'.

The activity stream is personalised for each user based on his/her Semantic SELF. Personalisation includes two aspects: firstly, activity items are prioritised based on the user's goal setting and other preferences; and secondly, a task inference on the same piece of information is performed for each related user separately. For example, a meeting organiser sends out meeting minutes to all invited participants; these minutes contain action items for several people. In this case, each recipient of the meeting minutes would receive an activity item with a personalised title that reflects his/her specific task.
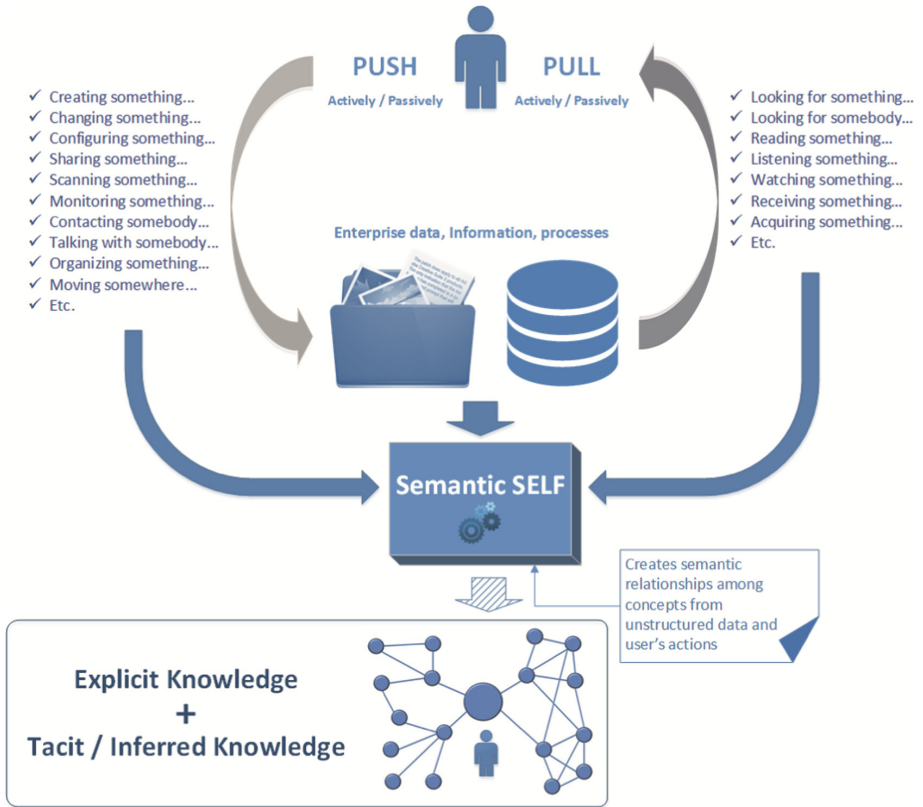
Fig. 3. Semantic SELF

Tactical Work. Once a personalised activity stream is generated, the tactical work level is entered when the user browses through the stream and performs quick actions on each item, such as to make a quick reply, set to snooze for later, or disregard the item. According to the conclusion of the present study people prefer to solve first 'small tasks' at the start of their workday. This way they reduce the number of pending tasks significantly after a small amount of time. Then they can better concentrate on 'big tasks' that take longer time to complete and also require more continuous concentration. Tactical work in the concept presented here supports users to efficiently solve 'small tasks' and plan for 'big tasks'. In addition, a link is proposed here between strategical work and tactical work, so that defined goals can be used as activity filters. The user is then able to see only those activities related to a particular goal.

Operational Work. Operational work is required to spend a longer period of time (e.g. 1–2 h) to continuously work on one 'big task'. An activity item can be extended to a full-screen mode, in which users can see all of the digital contents associated with this activity, including related previous activities (e.g. meetings), communications (e.g. emails, chats), other people involved, documents, images, videos, etc. As mentioned

above, currently people often need to manually search for multiple files related to one task, unless they create a central folder and manually save all related files at one place. It becomes even more complicated to link related meetings, emails and chat messages for one task and make them available in one place. By semantic understanding of information, one is able to infer relationships between activity items and gather related digital contents at one place. Using this approach in an operational work mode, the user has everything he/she needs at hand and can efficiently dive straight into productive work. From the full-screen view, the user can open files, edit files, create new files, write emails, create meetings or even plan a business trip.

3.2 Adaptation to User State

The definition proposed by the authors for a user's state includes the emotional state (e.g. happy, sad, angry, etc.) and the cognitive state (e.g. stress and fatigue). Considering a single-user work station in an open office environment, a non-intrusive and feasible way to detect the user's emotion is to analyse facial expressions based on camera data input. Existing approaches commonly detect the six basic types of emotions – anger, disgust, fear, happiness, sadness and surprise [1, 4, 13]. An eye-tracker is a non-intrusive sensor that can enable recognition of cognitive states, such as arousal, fatigue, and overload [8, 20, 31]. In case the user is wearing a smart wristband, emotional and cognitive states can also be recognised from physiological signals like heart rate, skin conductance, skin temperature, etc. [10, 26]. In private office rooms and meeting rooms, the user state can also be detected via speech input [3, 28].

When the user's state is recognised the system can adapt to it in two manners: a timely manner and a long-term manner. The timely adaptation is mostly used for negative user states (e.g. fatigue, anger) so that the system can propose corrective actions to improve the user state. For example, when fatigue is detected the system can propose to have a break or add more blue-spectral content into the ambient lighting that may help the user to remain vigilant and concentrating [27] (illustrated in Fig. 4). The long-term adaptation is mostly used for positive user states where the system then learns what makes the user happy over a longer period of time and registers this information as a personal preference. An example is illustrated in Fig. 5, where the system detects the task preference of a user and shares it with her team. In addition to the emotional and cognitive states, user preference can also be derived from the mouse and keyboard inputs (how a user interacts with the system, their personal style of work, etc.) and semantic analysis of this and other user-related information.
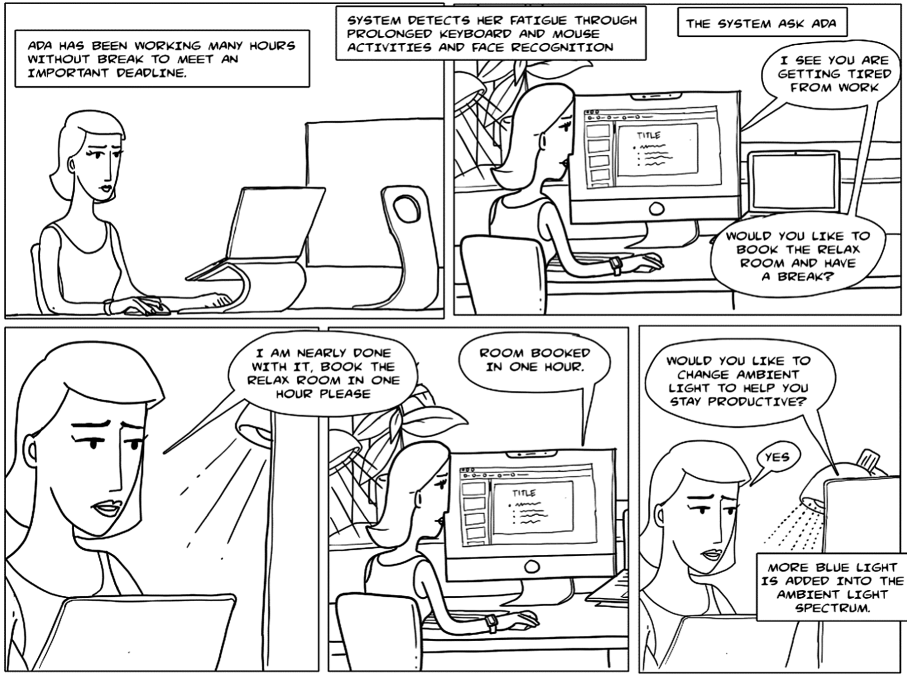
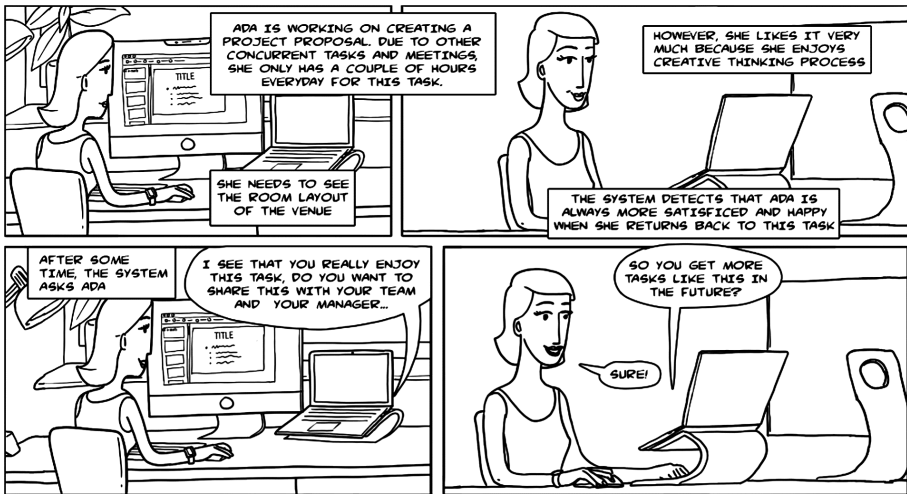**Fig. 4.** Adaptation to detected user fatigue

Fig. 5. Adaptation to detected user task preference

3.3   Smart Enterprise Search

One of the most prominent findings in the present user study (see Sect. 2) is that searching for the right information can be difficult and frustrating in the workplace. This result stands in line with various recent statists on enterprise search [9]: 'workers took up to 8 searches to find the right document and information'; 'employees spend approx. 1.8 h every day—9.3 h per week on average—searching and gathering information'. Or to put it another way, a business hires 5 employees but only 4 effectively contribute and the 5th is away searching for answers and is unable to generate value for the company. While Google and many others have provided high-quality internet search tools for everyone, enterprise search is still left 'in the dark' [25]. Cognitive Hub will provide a smart enterprise search engine to tackle this problem. Enterprise search should cover all resources within an organisation including: emails, all contents on the intranet, employee profiles, meeting rooms, software, licenses, hardware equipment, devices, etc. For each employee, enterprise search should also cover personal files on his/her local drive.

When it comes to searching for documents, which is perhaps the most common search use case in the workplace, the smart enterprise search is content-based but not file-based. Content-based search is enabled by document content extraction technology such as Tika [6], Rockhopper and Magellan (both developed within Konica Minolta). These technologies all index contents within a document including: titles, headings, body text, paragraphs, images and layout structure (the position of all of its contents). Using elastic content-based search, the smart enterprise search engine is capable of searching by several different attributes including: keyword; images; positions of specific contents; functional type (e.g. presentation, contract, invoice, etc.); related dates (creation, modification); related persons (e.g. the originator of the email); and by similarity (e.g. a document which looks like another document).

In case users are unsatisfied with search results after the first attempt, smart enterprise search provides a way to show users how queries are understood by the search engine and how users can adjust the query. As shown in Fig. 6, a search interpretation panel can be extended from the search query field. Users can review the understanding and correct any mistake. They can also specify more information to narrow down the search, for example by entering, 'Peter emailed it to me last week' (people: Petr, time: last week, source: email). This search interpretation panel is especially beneficial when users provide voice queries in natural language, because the natural language understanding technology is still far from being perfect, which is not entirely surprising given that even humans misunderstand each other from time to time.

3.4   Smart Transformation Between Physical and Digital Contents

Despite the observation that more and more paper documents have been migrated to digital versions and that more and more tasks have been being digitalised, recent statistics have shown that office printing volume is still slightly increasing [22, 29]. Based on various research literature [5, 11, 21], paper documents, books and learning materials have advantages that their electronic counterparts are currently not able to offer. In addition, one can observe that many people still carry around paper notebooks to
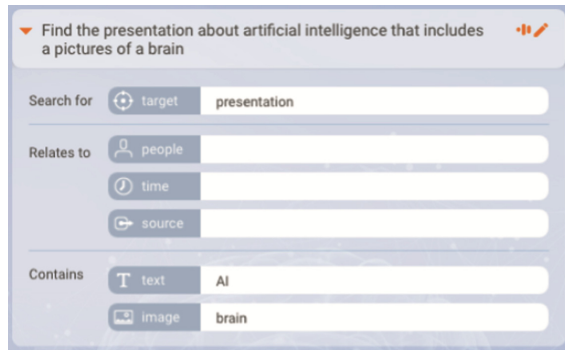
Fig. 6. The search interpretation panel

meetings and that Post-It is still a popular type of office supply. There are reasons to believe that paper documents and associated notes are not going to disappear from the office environment in the immediate future. However, the current workplace provides minimal support on transforming information between physical and digital forms. Printers and scanners typically only make a one-to-one exact transformation without understating the information content.

Cognitive Hub strives to provide a cyber-physical system that allows for smart transformation between digital and physical contents. This requires a device that has a projector, a camera with scanning function and integrated computer vision and machine learning capability. One example product that provides such functionality is offered by Lampix [18]. The following use cases demonstrate some of the features that are required to perform the physical to digital content transformation:

1. Digitising hand-written notes with smart understanding of contents. Assume that a user has a paper note that contains different types of contents, such as text, drawing, flow charts, etc. When the user digitises this note different contents are recognised and imported as separate objects so that the user can deal with them separately later on.
2. Digitising highlighted text in printed documents. A user has read a printed document and highlighted some text with a marker. He is able to digitise only the highlighted text but not the whole document.
3. Search for the digital version of a printed contents. When a user puts a printed document under the camera he can search for the digital version of the document. He can then point to an object on a page (e.g. a chart) and search for the digital version of this object. He can also, for example, point to a photograph of a person and receive more information about this person.
4. Search for keywords on printed document. If a printed document is lengthy and text-heavy, a user can search for a keyword and have it highlighted on the page, for example using an illumination projection system.
5. Paper-based collaborative work. Two people work at different locations can remotely collaborate using the paper medium. As illustrated in Fig. 7, the two people are working on a diagram. They see the same combination drawing on their own

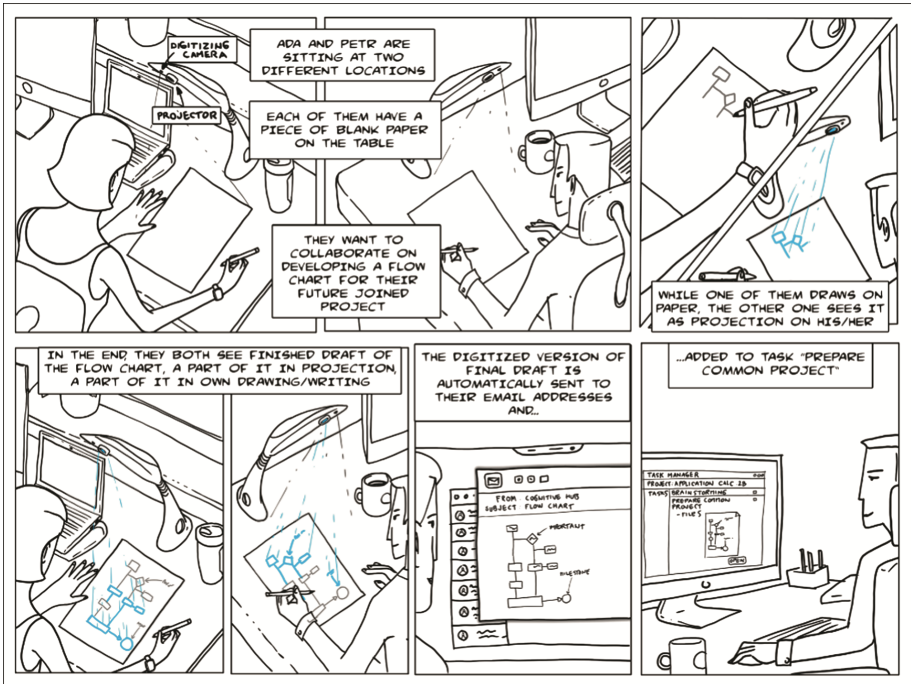paper which is an integration of their own drawing and the projection of the drawing from their collaborator.

Fig. 7. Remote collaborated work on paper

3.5 Multimodal Interaction

The traditional keyboard and mouse are the earliest developed input modalities for human-computer interaction. After the turn of the century, novel modalities such as touch, voice, gesture and gaze have made their way to applications in several domains including consumer electronics, automotive, gaming, advertising, manufacturing, entertainment, etc. [14]. However, in the office environment, human-computer interaction is still solely based on the keyboard, mouse (or trackpad) and the conventional windows-icons-menus-pointers (WIMP) interfaces.

A Cognitive Hub workstation will be well equipped to support voice, touch, gaze interaction and multimodal interaction. Multimodal interaction means that multiple modalities are available at all times, so that users can freely choose modalities based on task characteristics, environment and their personal preference. They can choose different modalities for different tasks or even switch modalities between different steps within a single task. As illustrated in Fig. 8, a user wants to send a scan to Petr. He used voice command in the first step because it saved him many clicks on the display, then

he switched to touch at the second step because it was much faster than describing which Petr to be selected verbally.

Fig. 8. Example of multimodal interaction.

Another valuable attribute that multimodal interaction should deliver is to combine two modalities into one input. Usually one modality is used to determine the object (e.g., looking at a window) and then the second modality identifies the action/command (e.g., speaking the word, "close"). In this case, each modality alone doesn't provide sufficient information about the user's intent. Cognitive Hub workstation will support gaze combined with voice input, gesture combined with voice input and touch combined with voice input. The voice modality provides actions/commands whilst other modalities specify the objects.

Multimodal interaction presents a paradigm shift from the conventional WIMP towards providing users with a more natural interaction and greater expressive power, flexibility, efficiency and robustness [23]. Multimodal interaction is **natural**, because human-to-human interaction is invariably multimodal. Humans tend to interact in a highly multimodal manner, especially in situations when they are frustrated, in tension, or under high workload [2, 12]. Multimodal interaction provides the user with **flexibility** in 'how' to conduct a task. Flexibility in executing the task eliminates potential conflicts of resources as described in the literature [30], because the user can flexibly adapt the response to the context. It is expected that by offering the flexibility in 'how' to conduct a task, the user will find a personalised way of conducting the task that would remain stable [24]. Multimodal interaction is expected to improve **efficiency** of both the short-term memory (i.e., operational/working memory) and procedural memory [7, 19]

because a more natural interaction requires less hierarchy and less abstraction. Considering multimodal interaction is natural for humans, the person can focus on 'what' to do rather than on 'how' to do it. The procedural memory responsible for how to do a certain task is then less loaded. Not requiring the user to remember multiple artificial steps (i.e., sub-tasks) within a process also results in increased free capacity of their working memory. As a consequence, the user is able to perform the task more quickly by having direct access to functions with less hierarchy and less abstraction. Building on the benefits described above, multimodal interaction then also improves the **robustness** of the interaction by reducing the occurrence of errors. Errors are more infrequent because task execution contains less steps, it can be completed more quickly and in a more personalised manner.

4 Proof of Concept Demonstrator

Cognitive Hub is still at a research and development stage within Konica Minolta. To illustrate the concept further a demonstrator has been designed that can immerse a user within a set of pre-defined interaction scenarios.

4.1 UI Design

The UI design supports the activity-centric metaphor of work (see Sect. 3.1). As shown in Fig. 9, the interface contains 4 panels. The 'Goals' panel on the left supports the strategic work. Users can add, describe, edit and delete goals. The 'Activities' panel in the centre is where the personalised activity stream is displayed. The most important ones (identified by Semantic SELF) are displayed at the top followed by the rest. Users can filter activities by selecting a goal so that only activities related to this goal are displayed. Each activity item contains an inferred task as its title, the due-date (if specified), sender, sending channel and related goal(s). A set of action icons appears on an item once the mouse hovers on it so that users can book time for it now, snooze it for later, or disregard it. When users click on an item it expands to show the original incoming information and users can perform short tasks such as making a quick reply, accepting a meeting request, etc. For operational work each item can be opened to a full-screen view where users can find all associated activities, people and digital contents. Contents can be created, opened and edited directly from the full-screen view. Actions such as creating meetings, writing emails and chat can also be directly performed from the full-screen view. Cognitive Hub will provide the functionalities described above to support 'quick tasks'; for more complex tasks it will connect to external applications and services including: text editors, instant messaging, financial tools, trip booking services, etc.

At the top of the interface, a timeline of the day shows all planned meetings within that day. This helps users to select appropriate tasks based on the time available until the next meeting. Users can also drag and drop an activity item on to the timeline to block time for it. On the right side of the screen, there is an 'Assistant' panel. This serves as a location for search functions and smart Q&A with a virtual assistant.
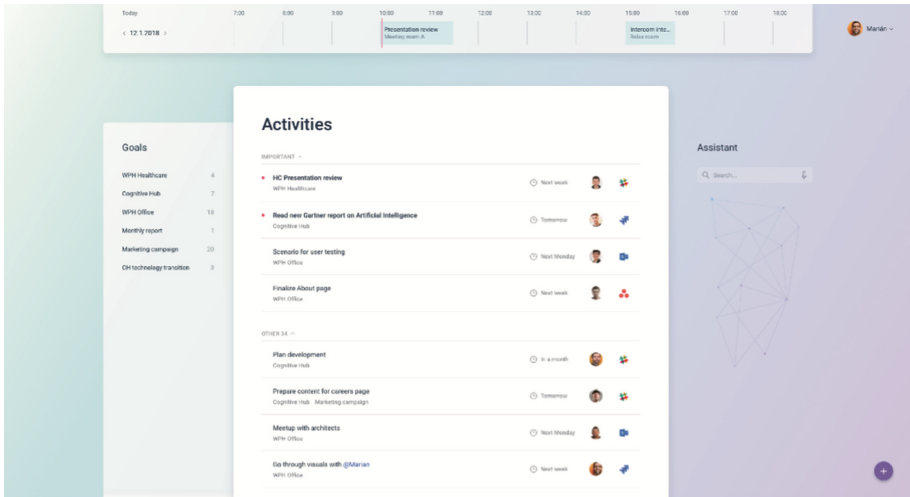
Fig. 9. User interface of Cognitive Hub PoC demonstrator

4.2   Hardware Setup

As shown in Fig. 10, the demonstration is equipped with a standard computer monitor, a keyboard, a mouse, a camera (for emotion detection), two microphones (for voice interaction and face orientation detection, an eye tracker (for gaze interaction) and a Lampix device.
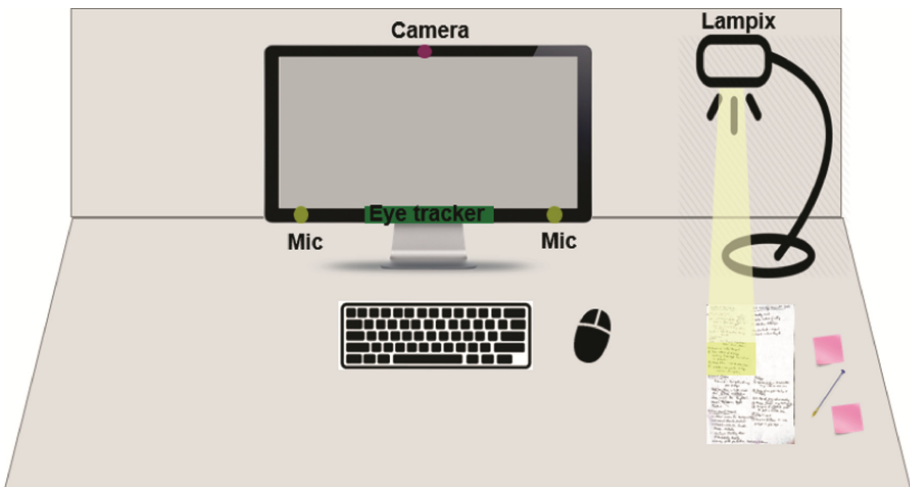

Fig. 10. Hardware setup of the Cognitive Hub PoC demonstration

4.3   Scenarios

Smart Q&A. At the beginning of a work day (which is envisaged as the start of the demonstration scenario), users can ask the assistant to provide a summary of what is new in their activity stream. They can ask questions related to their goals, activities and meetings. They can also ask for suggestions on which activity to work on first. Questions can be typed in the assistant panel or provided by voice. The assistant responds by voice and text.

Interaction with Activities. The activity stream in the demo includes 10 executable tasks. Users can scan through the activity stream and perform tactical work, such as to reply to an email, reply to a chat message, disregard an item, create a meeting, accept a meeting request, snooze an activity until next week, plan a fitness session, plan a business trip, etc.

Multimodal Interaction. In addition to keyboard and mouse, the demonstration also supports voice input, gaze input (based on Tobii eye-tracker [32]) and touch input on a smart surface (based on Lampix [18]). Voice modality can be used for all functionalities, including activity control, Q&A, search and the Lampix control. Keyboard and mouse can be used for all functionalities except the Lampix control. Gaze is not considered as a stand-alone modality in this demo so that gaze always needs to be used in combination with voice. For example, users can look at an activity item and say, "delete" or, "extend" or, "snooze." For the interaction using Lampix, voice and touch are interchangeable so that users can freely choose their preference.

Smart Search. Smart search in this demonstration supports document search and person search, by both typing and voice. Sample user queries include: find the presentation on Cognitive Hub that shows a picture of dirty hands; search for the meeting minutes from a meeting with Michal last week; I am looking for the brochure explaining the term 'digital cortex' near the beginning; or I need the article/paper about gaze interface that was emailed to me by Jane Smart.

Digital-Physical Content Transformation. The Lampix device that is used in this demo allows users to place, for example, a business card under the system and save it as a new contact. Users can ask for a digital version of a printed document. If the document contains multiple pages any page can be used to obtain the digital version. Users can point to a photograph of a person under the Lampix device and create a new email with the right email address pre-filled (the correct person having been identified by face recognition technology). Users can also ask to highlight keywords on printed documents. Interaction with the Lampix system can be performed via voice or touch on the table surface near the device.

Emotion Detection. The emotion detection solution uses camera input and a facial expression recognition SDK by Affectiva [1]. This solution identifies six basic emotions plus neutral faces. Emotion detection runs constantly in the background. When negative emotions such as anger and sadness are detected, the system may propose some

corrective actions. For example, if a user expresses dissatisfaction with a search result, the system may ask, "It seems that you were unhappy with the search results, should I adjust search parameters?" At the same time the search interpretation panel (Fig. 6) is extended for the user to review and make changes. At the very end of the demonstration, the system presents an overview of the user's emotion journey throughout the demo and shows statistics for each type of emotion (e.g. 40% neutral, 30% happy, 10% anger, etc.).

5 Conclusion

This paper presented the concept of a platform called Cognitive Hub, which aims to provide AI-powered services and features to improve the work experience and overall well-being of office workers. A demonstrator was developed to show the concept in action and illustrate its benefits and value. As a future step it is planned to conduct a study with a wide range of users who will experience the demonstrator and provide valuable feedback on the concept. The results of this future study will inform the authors on the necessary refinements of the Cognitive Hub concept. In parallel to this all technological enablers described above will continue to be developed and matured. Konica Minolta is working towards the launch of Cognitive Hub in the horizon of the next 2–3 years.

References

1. Affectiva. https://www.affectiva.com/. Accessed 26 Jan 2018
2. Alibali, M.W., Kita, S., Young, A.J.: Gesture and the process of speech production: we think, therefore we gesture. Lang. Cogn. Process. **15**, 593–613 (2000)
3. Anagnostopoulos, C.N., Iliou, T., Giannoukos, I.: Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artif. Intell. Rev. **43**(2), 155–177 (2015)
4. Anil, J., Suresh, L.P.: Literature survey on face and face expression recognition. In: Proceedings of the International Conference on Circuit, Power and Computing Technologies, ICCPCT, pp. 1–6. IEEE (2016)
5. Mangen, A., Walgermo, B.R., Brønnick, K.: Reading linear texts on paper versus computer screen: effects on reading comprehension. Int. J. Educ. Res. **58**, 61–68 (2013)
6. Apache Tika - a content analysis toolkit. https://tika.apache.org/. Accessed 30 Jan 2018
7. Bernsen, N.O.: Multimodality theory. In: Tzovaras, D. (ed.) Multimodal User Interfaces: From Signals to Interaction. SCT, pp. 5–29. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78345-9_2
8. Chen, S., Epps, J.: Automatic classification of eye activity for cognitive load measurement with emotion interference. Comput. Methods Prog. Biomed. **110**(2), 111–124 (2013)
9. Cottrill Research: Various Survey Statistics: Workers Spend Too Much Time Searching For Information. https://www.cottrillresearch.com/various-survey-statistics-workers-spend-too-much-time-searching-for-information/. Accessed 30 Jan 2018
10. Feel. https://www.myfeel.co/. Accessed 26 Jan 2018
11. Ferris, J.: The Reading Brain in the Digital Age: The Science of Paper versus Screens. Scientific American (2014)

12. Goldin-Meadow, S., Nusbaum, H., Kelly, S.D., Wagner, S.: Explaining math: gesturing lightens the load. Psychol. Sci. **12**, 516–522 (2001)
13. Happy, S.L., Routray, A.: Automatic facial expression recognition using features of salient facial patches. IEEE Trans. Affect. Comput. **6**(1), 1–12 (2015)
14. Jessica, G., Aditya, K.: Artificial Intelligence Use Cases – 215 Use Case Descriptions, Examples, and Market Sizing and Forecasts Across Enterprise, Consumer, and Government Markets. Tractica (2017)
15. Kaptelinin, V., Mary, C.: Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments. The MIT Press, Cambridge (2007)
16. Konica Minolta: Cognitive Hub: the operating system for the workplace of the future. White paper in Artificial Intelligence series (2017)
17. Konica Minolta: The future of work. White paper in Artificial Intelligence series (2017)
18. Lampix. https://www.lampix.co/. Accessed 30 Jan 2018
19. Maragos, P., Gros, P., Katsamanis, A., Papandreou, G.: Cross-modal integration for performance improving in multimedia: a review. In: Maragos, P., Potamianos, A., Gros, P. (eds.) Multimodal Processing and Interaction. MMSA, vol. 33, pp. 1–46. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-76316-3_1
20. Marshall, S.P.: Identifying cognitive state from eye metrics. Aviat. Space Environ. Med. **78**(5), B165–B175 (2007)
21. Myrberg, C., Wiberg, N.: Screen vs. paper: what is the difference for reading and learning? Insights **28**(2), 49–54 (2015)
22. Neville, W.: Office Printing Statistics 2017. https://www.lasersresource.com/blog/office-printing-statistics. Lasers Resource. Accessed 30 Jan 2018
23. Oviatt, S.: Ten myths of multimodal interaction. Commun. ACM **42**(11), 74–81 (1999)
24. Potamianos, A., Perakakis, M.: Human-computer interfaces to multimedia content: a review. In: Maragos, P., Potamianos, A., Gros, P. (eds.) Multimodal Processing and Interaction: Audio, Video, Text. MMSA, vol. 33, pp. 50–89. Springer, Boston (2008). https://doi.org/10.1007/978-0-387-76316-3_2
25. Susan, F., Chris, S.: The high cost of not finding information. ICD white paper (2001)
26. Verma, G.K., Tiwary, U.S.: Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. NeuroImage **102**, 162–172 (2014)
27. Viola, U., James, L.M., Schlangen, L.J., Dijk, D.-J.: Blue-enriched white light in the workplace improves self-reported alertness, performance and sleep quality. Scand. J. Work Environ. Health **34**, 297–306 (2008)
28. Weninger, F., Wöllmer, M., Schuller, B.: Emotion recognition in naturalistic speech and language – a survey. Emot. Recognit.: Pattern Anal. Approach 237–267 (2015)
29. West, M.: Is Office Printing Increasing or Declining?. https://www.printaudit.com/printaudit-blog/premier/is-office-printing-increasing-or-declining-answer-yes. Print Audit. Accessed 30 Jan 2018
30. Wickens, C.D.: Multiple resources and performance prediction. Theoret. Issues Ergon. Sci. **3**(2), 159–177 (2002)
31. Zhang, F., Su, J., Geng, L., Xiao, Z.: Driver fatigue detection based on eye state recognition. In: Machine International Conference on Vision and Information Technology, CMVIT, pp. 105–110. IEEE (2017)
32. Tobii eye tracker. https://www.tobii.com/