A distributed model for tacit design knowledge exchange

Ann Heylighen[1], Francis Heylighen[2], Johan Bollen[3] & Mathias Casaer[1]

[1] *Katholieke Universiteit Leuven, Dept. of Architecture,*
<ann.heylighen|mathias.casaer@asro.kuleuven.ac.be>
[2] *Vrije Universiteit Brussel, Center "Leo Apostel", <fheyligh@vub.ac.be>*
[3] *Old Dominion University, Computer Science Dept., <jbollen@cs.odu.edu>*

Abstract. The distributed cognition approach, and by extension the domain of social intelligence design, attempts to integrate three until recently separate realms: mind, society, and matter. The field offers a heterogeneous collection of ideas, observations, and case studies, yet lacks a coherent theoretical framework for building models of concrete systems and processes. Despite the intrinsic complexity of integrating individual, social and technologically-supported intelligence, the paper proposes a relatively simple 'connectionist' framework for conceptualizing a distributed cognitive system. This framework represents shared information sources (documents) as nodes connected by links of variable strength, which increases interactively with the number of co-occurrences of documents in the patterns of their usage. This connectionist learning procedure captures and uses the implicit knowledge of its community of users to help them find relevant information, thus supporting an unconscious form of exchange. The principles are illustrated by an envisaged application to a concrete problem domain: the dynamic sharing of design knowledge among a multitude of architects through a database of associatively connected building projects.

Keywords: distributed cognition, connectionism, architectural design

1. A general perspective on social intelligence design

Social intelligence design is one of an emerging cluster of approaches, originating in different disciplines and application domains, that attempt to bring together three until recently separate realms: *mind* (cognition, intelligence), *society* (social interaction, organizations, institutions), and *matter* (objects, tools, technologies). Related approaches go under the names of distributed cognition [Hutchins 1995], situated and embodied cognition [Susi & Ziemke 2003], activity theory [Hasan et al. 1998], social construction of knowledge [Berger & Luckmann 1967], extended mind [Clark 1995], and collective intelligence [Levy 1997, Heylighen 1999]. What they all have in common is that they extend the cognitive processes that characterize learning, memory and intelligence outward: from the individual brain to the surrounding social and physical environment.

One dimension of extension is the collectivity or social system formed by individual minds interacting. It is a commonplace that most of our knowledge, concepts, and problem-solving procedures develop in interaction with others, and that an individual cut-off from society (e.g. a child raised by wolves) would be virtually powerless to comprehend our complex environment—if capable of surviving at all. Thus, we can understand these higher forms of cognition only at the social level, as aspects of a collective intelligence emerging from the interactions between a multitude of individuals.

More recently, the awareness has been growing that cognitive processes moreover heavily rely on structures in the physical environment, i.e. on various forms of symbols, tools, and spatial arrangements [Kirsh & Maglio 1994]. These structures, exemplified by items such as scribbled notes, traffic signs, file cabinets, and computers, are used to store, process, communicate or organize information, thus reducing the burden on our limited mental capacities. Since most of these structures have been consciously designed for that purpose, we may call them *information technologies*.

However, this should not mislead us to immediately think in terms of digital computers or electronic networks. When shepherds put pebbles in a bag to count their sheep or drop them along the path to remember where they passed, they are using a cognitive artefact just as much as if they were calculating with a palmtop computer or determining their location with GPS. In both cases, they are manipulating their physical environment so that it can hold information, thus augmenting their interior capacity. Similarly, architects who design a building so that major features such as exits or bathrooms are easy to find are creating an information structure intended to minimize the cognitive load on its users. For an even more vivid example of the need for external structure to support cognition, imagine a household in which all individual items, such as cooking utensils, clothes, cleaning products, cans with food, books, pens, etc. would be distributed in a completely random order over the different cupboards, drawers and shelves of the house: since our brain would never be able to memorize all their locations, these items would effectively become irretrievable and thus useless.

The social and material dimensions of extended cognition are inextricably linked. Exchanging information with others necessarily entails a shared physical medium. In the simplest case, this is just the air that carries the sounds of our speech. But as soon as the information becomes too complex to be easily remembered, we begin using more structured media, such as paper, computer discs or spatial arrangements, which can accurately register and shape the information being shared so as to facilitate reuse. In the 'random distribution' example above, even if after years of diligent training we had come to memorize the locations of all household items, we would never be able to communicate this knowledge to another person. A 'common sense arrangement' (e.g. clothes together in the wardrobe, cleaning products under the kitchen sink), on the other hand, would merely need a few general indications to be used by others.

This brings us to the perspective of *distributed cognition*: to understand complex information processes, we must consider the distributed organization constituted by different individuals with different forms of knowledge and experience, the social network that links them together, and the artefacts, media or information technologies that support their individual thought and interindividual communication. A classic example of this approach can be found in the work of Hutchins, who used ethnographic methods to study a Navy warship as a distributed cognitive system, examining in detail how its navigation depends on the cognitive activity of many people coordinated by means of various instruments, ship navigation manuals, specialized communication channels, formal and informal procedures [Hutchins 1995].

The distributed cognition approach, and by extension the domain of social intelligence design, as yet offers little more than a heterogeneous collection of ideas, observations, and case studies. It lacks a coherent theoretical framework that would provide a solid foundation for building models of concrete systems and processes [Heylighen et al. 2004b, Susi & Ziemke 2001]. In spite of the intrinsic complexity of the problem of integrating individual, social and artefact-supported intelligence, we believe that a few well-chosen concepts and methods can help make the issue much more tractable, offering a relatively simple paradigm to conceptualize a distributed cognitive system. In the remainder of this paper, we will first briefly sketch our ‘connectionist’ conceptual framework, and then describe in more detail its envisaged application to a concrete problem domain: the dynamic sharing of design knowledge among a multitude of architects by means of a database of associatively connected building projects. The paper closes by situating our approach in the context of related research and outlining directions for future work.

2. A connectionist perspective on distributed cognition

The essence of connectionism is to model cognitive systems as networks of nodes connected by (typically weighted and directed) associations or links. This static representation is complemented by dynamic rules that govern the short-term interaction between the nodes, and the long-term effect this has on the links. This connectionist perspective is inspired by the organization of the brain, where neurons play the role of nodes, synapses the role of links, and interaction occurs by the transmission of electrical activation from neuron to neuron, with a strength proportional to the ‘conductivity’ of the connecting synapse. This conductivity is adjusted according to the simple principle that successfully used links grow stronger, while unsuccessfully used ones weaken, thus continuously improving the overall ‘success rate’ or efficiency of the network. This can be formalized most simply by the Hebbian rule, which states that the increase in strength of a link is proportional to the strength of the co-activation (product of activations) of the nodes it connects [Heylighen & Bollen 2002]. Each experience of use thus leaves a trace in the linking pattern. The network therefore functions as a self-organizing, dynamic memory, which becomes an increasingly reliable guide for dealing successfully with situations similar to those experienced in the past.

While this is of course a gross simplification of the actual processes in the brain, research on artificial neural networks has shown that this approach allows us to successfully model most fundamental cognitive processes [McLeod et al. 1998]. The advantage of the relative abstractness and simplicity of this connectionist representation is that we can easily extend it to systems that do not consist of actual brain tissue. In particular, we propose to extend it to distributed cognitive systems that consist of human individuals and physical information sources (‘documents’), linked by communication channels.

For the sake of simplicity, we will focus in this paper on the mechanisms of distributed memory, learning and information retrieval, assuming that more complex processing either takes place in individuals or emerges from the collective dynamics of the network (see [Heylighen et al. 2004b] for a somewhat more in-depth look at these emergent processes). We further assume that all individuals have in principle read/write access to all documents (information sources). Thus, the documents play the role of a shared memory, through which individuals can exchange information, access, and contribute to, the whole of their collective knowledge [Heylighen 1999]. The network of linked documents thus constitutes a medium for indirect cognitive collaboration between its users. (The connectionist perspective can also be used to model the direct communication between socially or technologically linked individuals, but since our research on this topic is still very preliminary [Van Overwalle et al. 2004, Heylighen et al. 2004b], we will limit ourselves here to document-mediated forms of distributed cognition.)

Given the present information explosion, documents can easily number in the millions (or even billions if you consider the present web). As the example of household items illustrates, to make this memory in any way usable, we need an intuitive ordering that makes sense to all users. E.g. although an individual could easily memorize that the cleaning products are stored in the wardrobe, and the trousers under the kitchen sink, most people would look for the trousers in the wardrobe, and vice-versa. But for an ever growing collection of novel and abstract data, there are no such conventional location schemes, like wardrobes or kitchen sinks. We need to develop a *collective mental map* [Heylighen 1999], that indicates the relative locations of all documents, and that constantly adapts to changes in the information content or user preferences.

The basic advantage of the connectionist model is that it allows us to have such a mental map self-organize. We simply represent each document by a node, and then determine the strength of the link between two nodes depending on the degree to which they are mutually relevant or associated. The location of a node is then determined by its ‘surroundings’, i.e. the nodes it is most strongly linked with. To retrieve a document, you start from the already known document that is most associated with it, and then move ‘upstream’, each time selecting the link that seems most strongly associated with the target. This is how people navigate through hypertext networks [Bollen 2001]. The effectiveness of the strategy depends on how well the actual hyperlinks reflect the users' intuitive associations. And this depends on how well the collective mental map represented by the linking patterns captures the collective knowledge of its users. The Hebbian learning rule sketched above allows the network to assimilate the knowledge implicit in how the network is being used [Heylighen & Bollen 2002]. We simply assume that two documents A and B are *co-activated* if they are consulted by the same user within a relatively short time interval. This indicates that B is likely to be relevant for A, and that it is worth strengthening the link A → B. The next time a user consults A, B will be easier to reach, thus reducing the burden of searching for relevant documents. In that way, the network continuously adapts to better reflect the intuitive expectations of the users collectively.

Co-activation, in the sense of being consulted (nearly) simultaneously, can be generalized to *functional co-occurrence*: appearing in the same category of usage. Some examples of functional categories in which documents may co-occur: being cited in the same paper, being bought by the same customer, being present in the same library section, having the same author, having the same keyword in the title, sharing specific features such as document type or subject... Some of these data are already being used to find related documents, e.g. co-citation to locate scientific papers on the same subject, or co-purchase to recommend relevant books to customers of web bookshops. (This is a form of ‘collaborative filtering’ [Shardanand & Maes 1995].) In our generalized connectionist model, all these co-occurrence data are potentially useful to indicate associations between documents, i.e. to efficiently structure the collective mental map and thus to facilitate collective information retrieval and exchange. Combining all data available should improve the quality of the associations, especially in the beginning stages when the network has not yet had much time to ‘learn’ from the way it is being used.

The only problem is to determine the relative importance of the different contributions: is having the same author a better or worse indication of mutual relevance than being bought by the same customer, or having the same keyword? A possible solution would be to consider the degree of association implied by each type of co-occurrence as a variable, and then to calculate the correlation coefficients between all these variables over the whole collection of documents. Applying the statistical technique of factor analysis to the matrix of cross-correlations, we can then compute a ‘principal component’. This is a linear combination of all variables that maximally co-varies with each of them. The load of a variable on this ‘principal’ dimension can then be interpreted as an indication of its importance for the overall calculation of associative strength. Co-occurrence variables that correlate poorly with the others will have low loads, implying that they are unreliable indicators of relevance.

In conclusion, our connectionist approach is able to intuitively structure any—even random—collection of documents, first by aggregating the various features that two documents share (co-occurrence) to determine their initial degree of association or ‘distance’ in the collective mental map. This network of associations is then continuously improved, updated and fine-tuned by letting it learn from the actual way it is used by its community of users. This makes it much easier for the users to find the documents they are interested in, in addition to providing a global ordering for the collection that allows various more complex forms of use, such as categorizing documents by means of cluster analysis, finding the most relevant documents with particular features, producing tailor-made, context-dependent recommendations for individual users, and discovering global trends [Heylighen & Bollen 2002]. The method relies on an implicit, technology-mediated collaboration between the different users, which does not demand any effort from them except the one they have to invest anyway to use the network purely for their individual sake.

3. A concrete application

Having sketched our connectionist framework to conceptualize a distributed cognitive system, this section switches attention to a concrete application of this framework in the domain of architecture. In this domain, design ideas are developed as much through interaction as by individuals in isolation. This observation inspired the development of a Dynamic Architectural Memory On-line (DYNAMO), an interactive platform to share ideas, knowledge and insights in the form of concrete building projects among architects in different contexts and at different levels of expertise.

Interaction with professional and student architects revealed this platform to suffer from at least two thresholds. First of all, making projects available to other platform users takes time, effort, and specific skills. Secondly, architects tend to sense a psychological threshold to share their ideas and insights with others. This section points out how the framework outlined above, and in particular the notion of co-occurrence can be instrumental in addressing both thresholds. Equipping DYNAMO with a self-organizing mental map that reflects the relative locations of all its projects, and that constantly adapts to changes in its content or user preferences, would enable the platform to benefit from its users’ insights without any extra effort or even awareness on their part.

In the subsections which follow the basic concepts of DYNAMO are presented, followed by their implementation and results from experiments with various user groups. Based on this foundation, we point out the power of the connectionist framework to help the platform effectively fulfill its original ambition.

3.1 DYNAMO in a nutshell

DYNAMO (htpp://dynamo.asro.kuleuven.ac.be) is a web-based design assistant for students and professional designers in the field of architecture, which aims to incorporate quite literally the view of cognition underlying Case-Based Design (CBD) and at the same time to extrapolate it beyond the individual [Heylighen & Neuckermans 2000].

Firmly rooted in the Theory of Dynamic Memory [Schank 1982], the CBD approach propounds that people's knowledge does not only consist of abstract, generally applicable principles, but also of specific experiences, so-called cases [Riesbeck & Schank 1989, Kolodner 1993]. Moreover, it claims that human memory is dynamically changing with every new experience. Several years of observing and analyzing people's remindings have nurtured the hypothesis that experiencing, understanding, remembering and learning cannot be separated from one another. Our understanding grows by trying to integrate new things with what we already know. As a result, understanding causes us to come across old experiences as we process new ones. A significant side-effect of this process of understanding is that memory never behaves exactly the same way twice, since it changes as a result of its own experiences. As experiences are recalled and used, memory gets an opportunity to try out the knowledge associated with them. This allows memory to re-organize and re-define itself dynamically, in other words to learn from its experiences [Kolodner 1993].

Learning from experience can occur in different ways [Riesbeck & Schank 1989]. New episodes are stored in terms of old expectations generated by previous experiences. Eventually expectations that used to work may have to be invalidated. Indices to unique experiences that were once useful will cease to do so because similar experiences have been encountered. In short, memory learns from experience by acquiring new cases, grouping similar cases, or re-indexing cases stored improperly at first.

Inspired by the cognitive model underlying CBD, DYNAMO is conceived as an (inter)active workhouse rather than a passive warehouse. It actively develops and is interactively developed by architects' design knowledge, in that it stimulates and supports several modes of interaction:

- interaction among building projects, for projects are labeled and linked to related projects by various features architects address during design;
- interaction between (human) designer and (computer) memory, for users cannot only consult projects in DYNAMO, they can improve its content in various ways;
- interaction among individual designers in different contexts and at different levels of expertise, for DYNAMO is meant for collective use by students in architecture schools and professionals in design firms;
- and thus also interaction between practice and education in architectural.

3.2 Implementation

From a technical point of view, DYNAMO can be thought of as a learning content management system. The platform is designed to support the creation, storage, use and reuse of learning content in the granular form of building projects. All learning content is organized by means of a dynamic metadata classification system and stored in a data repository embodied by a relational database, which is subdivided into four sub-databases as follows (see Fig.1):

The *cases* database labels every building project by various features: project name, architect and location, but also aspects of form and space, function, construction and context. These metadata serve as filter criteria during retrieval and as links to projects with analogous characteristics. In order to avoid confusion, a clear distinction should be made between *categories* and *values*. The term category refers to the name of an index, e.g. spatial configuration. Each category provides a place for a project to characterize itself with one or more values, e.g. cluster, linear, radial, etc. A value thus refers to the concrete realization of a category for a certain project. It characterizes a specific project but can, and in many cases will be the same for several projects. In other words, whereas categories are chosen generally and shared by all projects, the values for these categories are assigned to each project specifically. For some categories, materials for instance, a single project can have multiple values.

The *files* database contains all information about the files documenting the projects (file name, author, source, file type, etc). Projects can be documented with a combination of various media, ranging from sketches and drawings, over digital models and animation, to pictures, video and text. The advantages of this combination are manifold. Compared to written data, visual and spatial representations better fit the architect's designerly way of thinking and working [Cross 1982]. Furthermore, mixing multiple modes of representation provides users with a richer learning experience [Vora & Helander 1997], which results in a better understanding of the project.

The *log* base keeps a log of all user interactions with the platform. For every user (inter)action, the log base stores parameters such as the identity of the user (user name), the location of the page that registered the action, the query string used during the action (which typically contains variables such as the project and/or file id), the date and time the action was performed, the client ip address, and the type of action performed (such as save file, view project, search, etc.).

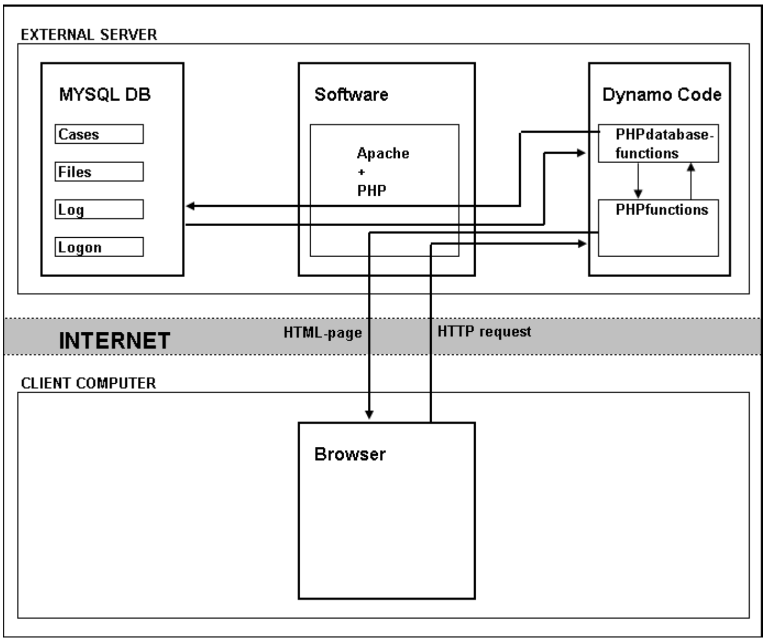

Figure 1: Schematic overview of DYNAMO's implementation. As we write, DYNAMO represents 536 projects, 7032 files, 1911 keywords (index values), 713 registered users, and 85126 logs.

The *logon* database takes care of user privileges and administration. The current version of DYNAMO distinguishes between three types of users. DYNAMO *users* have limited privileges: they can consult the platform, feed it with new projects, project features and documentation, and create new categories. *Monitors* have extended privileges in that they can approve, alter or delete user-added materials. *Administrators* have access to all DYNAMO features, including user and monitor administration.

3.3 Use

DYNAMO is being used and developed by various groups and individuals in architecture schools, and provides an emerging knowledge pool for design practice. As such, DYNAMO holds great potential to become a distributed cognitive system for student and professional designers in architecture. The question to what extent this potential is effectively exploited, motivated a series of experiments to assess the added value of the platform in supporting architectural practice and education.

This longitudinal study has been reported more fully elsewhere [Heylighen et al. 2004a] and only the findings relevant in the context of this paper are described here. To give a little background: the study monitored and analyzed various user groups (students and practitioners, both novice and expert designers) while using the platform in solving concrete design tasks. To this end, a variety of data was drawn on, including think aloud protocols, surveys and log files.

Protocol analysis revealed some interesting differences between novice and expert designers. While novices tend to scan DYNAMO for inspiration on interesting concepts, experts rather try to project their own concepts onto the projects available in the platform. In a comparative analysis, the novice under consideration did not consult DYNAMO because he was looking for anything in particular, but to get inspired by concepts in other projects. By contrast, the expert explicitly looked for concepts related to his design ideas. His search was more structured in that he tried to match his own concepts with those in the project collection. This seems to suggest that novice and expert designers make different associations between projects: novices tend to rely, primarily, on superficial similarities (e.g. the function) between the design task at hand and projects in DYNAMO, hoping that the latter will provide useful concepts for their task; experts, for their part, seem to draw more insightful parallels (e.g. the underlying concept or spatial organization) between projects that may look quite different at first sight.

The study also revealed several obstacles that prevent DYNAMO from being used as originally intended. Worth mentioning here are the obstacles encountered in establishing a community of users who exchange and learn from each other’s insights and experiences. In principle, we expected users to consult DYNAMO’s project collection, but also to (inter)actively participate in feeding it, be it by adding, supplementing, connecting or commenting on projects. Yet, apart from students submitting an obligatory project analysis, hardly any user so far exploited the opportunities for (inter)action. Perhaps a good term to characterize this pattern of use is ‘free-ridership’ [Agre 2003]: users who do not participate in contributing to developing DYNAMO’s content nonetheless benefit from it.

One reason for this ‘free-ridership’ may relate to the submission procedure, which users seemed to find rather complex at first. In general, making material available to others takes time, effort, and sometimes specific skills [Van House 2003]. Especially the articulation of appropriate metadata, which are key to identifying inter-project relationships, is far from trivial a task. Moreover, making material available seems inconsistent with the habits and priorities in architectural practice. Professional architects turned out to be highly skeptic about sharing information on their own projects with other platform users, anxious to give away 'the secret of their success'. If practitioners are not willing to share their knowledge, insights, and expertise, this may become a heavy burden on DYNAMO's future.

3.4 Towards self-organization

In the light of the connectionist model, however, this 'free-ridership' does not need to be as problematic as first meets the eye. According to this model, DYNAMO can be thought of as a collective mental map in which the meaning of a single project is defined by the whole of its associations with other projects rather than by its independent content. In other words, projects can be considered as nodes in an associative network and the strength of the links between them as the degree to which they are mutually relevant or associated.

At this point, these links are determined by the projects' characteristics, i.e. by the metadata explicitly specified by users. As a result, functional co-occurrence among projects is necessarily limited to having the same value for one or more categories, such as being designed by the same architect, using the same material, being built in the same location or period, having the same function, and the like. Moreover, once these values have been specified, the degree of association between two projects stays the same, and may alter only when a user explicitly creates a new category.

Following the framework outlined above, however, this collective mental map may be considerably enriched by exploiting the various modes of interaction DYNAMO aims to stimulate and support. So far, we have mentioned, interaction remains a quite limited phenomenon, in the sense that most users of the platform act as 'free-riders'. In a slightly different sense, however, there is in fact a lot of interaction—across different projects and between user and platform—that currently remains implicit, yet holds great potential to determine the degree of association between projects in a more sophisticated and dynamic way.

A first possibility would be to determine this degree by taking into account not only those features of a project that are explicitly specified as category values, but also features that are contained in the files documenting that project. For instance, while the user submitting the project 's Hertogenmolens in Aarschot (Belgium) by Noa Architects may forget (or fail) to label it explicitly as making use of corten steel, a journal article documenting this project is likely to mention this material, as it plays an instrumental role in realizing the architects' design concept. (Along the same line, one could even think of distilling numerical information out of digital models, or using image recognition techniques to derive visual features from pictures or drawings.)

In addition to features extracted from files that document projects, DYNAMO could exploit information on how users actually interact with projects, as each user (inter)action is automatically recorded by the log base. By analyzing the records of each project consultation, the degree of association between projects could be computed on the basis of relations such as being consulted by the same user or being consulted within the same time slot [Heylighen & Bollen 2002]. Also, if any documents related to the projects are requested, one could derive the user's relative measure of interest in each project by counting the number of requested files or the total time spent on the project, and this interest gives a direct indication of the degree of 'activation' to be accorded in a connectionist model [Heylighen & Bollen 2002, Heylighen 1999].

Added together, these various forms of association between projects form determine an overall, fine-grained association matrix, which can be put to use to guide DYNAMO users in various ways [Heylighen & Bollen 2002]. To start with, it would be possible to append to each project a list of the projects that are most strongly associated with it, in decreasing order of association. In this way, a user who discovers a relevant project will immediately get a pointer to the projects that are most likely to be relevant as well, even though they may be labeled incompletely or differently. Another application is the clustering of projects. Submitting the association matrix to a clustering algorithm should allow DYNAMO to automatically create categories of projects, even when those categories have not been formally recognized yet. Finally, by keeping track of all the projects a particular user has consulted, together with an estimate of the degree of interest for each of these projects, we can define an 'activation vector' representing that user's present interest profile. Multiplying this vector recurrently with the association matrix implements the connectionist process of 'spreading activation', where new projects are activated proportionally to the degree that they are associated with the complete interest profile of the user, rather than just the last consulted project. In this way, DYNAMO can produce at any time a tailor-made recommendation for the user, which is updated with each further consultation activity.

These applications, if successfully implemented, would considerably help DYNAMO in tackling the obstacles it is currently facing. First of all, the extraction of relations between projects from documents and log files would free users at least in part from the cumbersome task of articulating appropriate metadata [Heylighen & Bollen 2002], while considerably enriching the content of the platform. As DYNAMO grows, this extraction becomes even more important if users are to see the wood for the trees. Indeed, on the level of data management, the continuously growing data pool demands proportional efforts from the DYNAMO monitors and administrators to keep the information provided of an acceptable quality and trustworthiness. Furthermore, DYNAMO users are confronted with ever larger selections of projects, calling for more sophisticated search and navigation facilities.

Secondly, and more importantly, equipping DYNAMO with a self-organizing mental map would allow benefiting from the expertise of *all* architects who use the platform, including those who do not release information on their own projects. Indeed, such map would capture and exploit the insights that expert designers rely on while consulting projects without any extra effort or even awareness on their part, thereby making these insights readily available to other DYNAMO users. Each user, or rather each usage would interpret and transform the previous organization of DYNAMO's collective mental map, not in its entirety but piecemeal. Each usage would produce additional associations in the network of projects through the (often tacit) insights that users rely on to guide project selection and consultation. As pointed out above, the usage of experts seems to differ from that of novices, not because they are applying different rules or strategies, but simply because they see other, less obvious relationships across projects. Enabling DYNAMO to exchange this kind of insights effortlessly and unnoticed, i.e. without bothering the users to explicitly formulate their insights, would be a considerable step towards establishing a distributed cognitive system for architecture.

4. Related work

In making this step a reality, there are plenty of insights from other research we can draw on. In particular, the production of self-organizing mental maps is strongly related to techniques applied in the domains of recommender systems and data mining of web usage data.

Research on recommender systems can be split into two main categories according to how they regard the relationship between individual users and their context in a user community: personalization on the one hand and collaborative filtering on the other hand. The former category attempts to provide users with interface customizations that relate to individual preferences. Implicit in this approach is the tendency to stress the needs and preferences of individual users, and to isolate them from top-down design verdicts imposed by system engineers and other users even when WWW interfaces are involved [Rucker & Polanco 1997]. The bias of these approaches is against communal standards and for individual needs. Personalization systems have been designed to make use of log and usage data [Takano & Winograd 1998, Harvey et al. 1998, Mobasher et al. 2001].

Collaborative filtering on the other hand aims to produce recommendations by comparing users to other individuals in a user community and deducing possibly interesting items from that comparison [Wasfi 1999, Konstan 1997]. If two users are similar, they may like similar items. Hence, items that one user preferred are recommended to other, similar users [Herlocker et al. 1999]. The application of collaborative filtering algorithms therefore implies that a metric of user similarities is defined and that it positions individual users within a contextual community of other users. In other words, individual preferences are addressed on the backdrop of a communal, semantic model [Mock & Vemuri 1997], leading to a simulation of the social processes involved in retrieval [Harvey et al. 1998], such as word-of-mouth [Shardanand & Maes 1995], human recommendation [Twidale et al. 1997] and distributed problem solving [Bouthors & Dedieu 1999]. The literature on collaborative filtering is thus quite relevant to the mentioned principles of self-organizing mental maps and their application.

In addition, the principle of collaborative filtering can be transposed to generate item similarities based on usage patterns. Rather than the similarity of users being assessed on the basis of the items they downloaded, the similarity of items can be assessed on the basis of the users that downloaded them [Sarwar et al. 2001, Heylighen 1999]. This procedure is similar to the mentioned principle of log analysis: two items are deemed similar if they were frequently downloaded by the same sets of users. Although in many cases sufficient usage data are missing to generate dense networks of item relationships, this can be alleviated by associative retrieval techniques [Huang et al., 2004].

This procedure, however, does not address the issue of temporal order. The sequence in which items are accessed or co-occur may be an equally important indicator of relatedness as simple co-occurrence in a user session. Widespread efforts have thus been made to generate models of the user click stream [Xiao & Dunham 2001] and to apply them to recommender systems [Bollen 2000, Bollen & Rocha 2000], including the use of Hidden Markov Models [Levene & Loizou 1999] and datamining for the discovery of association and sequential rules [Mobasher et al. 2000, Gery & Haddad 2003]. Such models can be used to generate item recommendations based on a user's past navigation history or to dynamically restructure web sites' link patterns [Yan et al. 1996, Masseglia et al. 1999, Albanese 2004].

The central issue in most analysis of log data concerns data validity: usage logs can be quite noisy [Pitkow 1997]. Robots, proxies, caching, and user temperament can all greatly increase the difficulty to identify session boundaries. Bollen and Nelson [2002] identify a method to generate self-organizing mental maps by having autonomous information objects monitor their local usage and automatically update their links to other objects, which may be appropriate to turning DYNAMO into a distributed cognitive system.

5. Summary and future directions

In response to the lack of a coherent theoretical framework for the distributed cognition approach, and by extension for the domain of social intelligence design, we have proposed a relatively simple 'connectionist' paradigm that provides a solid foundation for conceiving and building models of concrete distributed cognitive systems, and the processes through which they self-organize and adapt. Subsequently, we have zoomed in on the envisaged application of this paradigm to a concrete system: a multimedia platform with associatively connected building projects that aims at the dynamic sharing and exchange of design knowledge by a multitude of architects.

This exchange turns out to suffer from at least two thresholds: a physical threshold caused by the effort and time needed to feed the platform, and a psychological threshold to share ideas and insights with other designers. Both thresholds could be largely overcome by conceiving the platform as an associative network of projects, and exploiting the information and insights that are implicitly available in the project documentation and the data on user (inter)actions to determine and continually update inter-project relationships. This should allow the platform to learn from its previous experiences and progressively increase its ability to satisfactorily support its users.

Further evidence is needed to prove the value of the scenario proposed in this paper. A first step is to define a working strategy to calculate the association matrix on the basis of usage logs and co-occurrence data. Subsequently, we will need to analyze the resulting associations and test whether they are useful in guiding and serving the user. If so, DYNAMO's interface must be redesigned to visually integrate the degree of associations between projects.

One challenge in implementing the proposed scenario derives from the fact that the algorithm generates the best results if used frequently and recursively. When a feedback loop is created, the mining process will further fine-tune itself to enhance its performance. This means that the usefulness of the approach cannot be evaluated before it has more or less been integrated into the interface and tested over a significant period of time.

Moreover, a key challenge will be to determine whether all log data are worth to take into account and trigger a new recursion. Indeed, tracing the 'interesting' information in the log base, i.e. insightful associations between projects as made by expert designers, in a system with 713 users is far from trivial a task. We do not necessarily want the algorithm to take into account every single usage, but only those interactions that can help enrich the content of the platform. On the other hand, we can expect that poor-quality data, representing users browsing randomly or not knowing very well what they are doing, will be averaged out by taking all data together, which would leave only the statistically significant trends.

Awaiting the implementation of this scenario, our purpose in presenting the connectionist framework and its envisioned application to a concrete system, is to point out the potential of using a few relatively simple yet well-chosen analogies with dynamic and distributed cognitive processes to better understand social intelligence design, as both the connectionist paradigm and DYNAMO do quite literally. In our view, it is precisely these common roots in the functioning of the human brain that makes the paradigm highly compatible with DYNAMO (and vice versa) and the scenario outlined in this paper worth pursuing.

Acknowledgements

Ann Heylighen is a Postdoctoral Research Fellow of the Fund for Scientific Research-Flanders (FWO).

References

Agre P: 2003. Information and Institutional Change: The Case of Digital Libraries, *Digital Library Use. Social Practice in Design and Evaluation,* The MIT Press, Cambridge (Mass), 219-240.

Albanese M: 2004. Web personalization based on static information and dynamic user behavior, *Proceedings of the 6th ACM international workshop on Web Information and Data Management* (WIDM04), ACM Press.

Berger P L & Luckmann T: 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Anchor.

Bollen J: 2000. Group user models for personalized hyperlink recommendation, *LNCS 1892 - International Conference on Adaptive Hypermedia and Adaptive Web-based Systems* (AH2000), Trento, Aug 2000. Springer Verlag, 39-50.

Bollen J: 2001. *A Cognitive Model of Adaptive Web Design and Navigation: A Shared Knowledge Perspective*. PhD Thesis, Vrije Universiteit Brussel.

Bollen J & Nelson ML: 2002. Adaptive networks of smart objects, Bollen J & Nelson ML (eds), *Proceedings of Workshop on Distributed Computing Architectures for Digital Libraries* (ICPP2002), Vancouver (BC), Canada, Aug 18-21, 2002, IEEE, 487-496.

Bollen J & Rocha LM: 2000. An adaptive systems approach to the implementation and evaluation of digital library recommendation systems, *LNCS – 4th European Conference on Research and Advanced Technology for Digital Libraries* (ECDL2000), Lisbon, September 2000, Springer Verlag.

Bouthors V & Dedieu O: 1999. Pharos, a collaborative infrastructure for web knowledge sharing, Abiteboul S & Vercoustre AM (eds), *Research and Advanced Technology for Digital Libraries,* 3rd European Conference, ECDL '99, Paris, France, Sept 22-24, 1999: Proceedings, vol 1696 of Lecture Notes in Computer Science, Springer-Verlag, 215-233.

Clark A: 1997. *Being There: putting brain, body, and world together again*, Cambridge (Mass), MIT Press.

Cross N: 1982. Designerly ways of knowing, *Design Studies*, 3(4), 221-227.

Gery M & Haddad H: 2003. Evaluation of web usage mining approaches for user's next request prediction, *Proceedings of the 5th ACM international workshop on Web information and data management*, ACM Press, 74-81.

Harvey CF, Smith P & Lund P: 1998. Providing a networked future for interpersonal information retrieval: Infovine and user modeling, *Interacting with Computers*, 10, 195-212.

Hasan H, Gould E & Hyland P: 1998. *Information systems and activity theory: tools in context*, 7-18. University of Wollongong Press, Wollongong.

Herlocker JL, Konstan JA, Borchers A & Riedl J, 1999, An algorithmic framework for performing collaborative filtering, *Proceedings on the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley (CA), Aug 15-19, 1999, 230-237.

Heylighen A & Neuckermans H: 2000. DYNAMO: Dynamic Architectural Memory On-line, *Educational Technology and Society*, 3(2), 86-95.

Heylighen A, Neuckermans H & Casaer M: 2004a. ICT revisited—from information & communication to integrating curricula?, *ITcon*, 9, 101-120.

Heylighen F: 1999. Collective Intelligence and its Implementation on the Web: algorithms to develop a collective mental map, *Computational and Mathematical Theory of Organizations* 5(3), 253-280.

Heylighen F & Bollen J: 2002. Hebbian Algorithms for a Digital Library Recommendation System, *Proceedings 2002 International Conference on Parallel Processing Workshops*, IEEE Computer Society Press.

Heylighen F, Heath M & Van Overwalle F: 2004b. The Emergence of Distributed Cognition: a conceptual framework, *Proceedings of Collective Intentionality IV*, Siena (Italy).

Huang Z, Chen H & Zeng D, 2004. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering, *ACM Trans. Inf. Syst.*, 22(1), 116-142.

Hutchins E: 1995, *Cognition in the Wild*, MIT Press.

Kirsch D & Maglio P: 1994. On distinguishing epistemic from pragmatic action, *Cognitive Science* 18, 513-549.

Kolodner JL: 1993. *Case-Based Reasoning*, Morgan Kaufman, San Mateo.

Konstan JA, Miller BN, Maltz D, Herlocker JL, Gordon LR & Riedl J: 1997. Grouplens: applying collaborative filtering to usenet news, *Communications of the ACM*, 40(3), 77-87.

Levene M & Loizou G: 1999. A probabilistic approach to navigation in hypertext, *Information Sciences*, 114, 165-186.

Lévy P: 1997. *Collective Intelligence*, Plenum.

Masseglia F, Poncelet P & Teisseire M: 1999. Using data mining techniques on web access logs to dynamically improve hypertext structure, *SIGWEB Newsletter*, 8(3), 13-19.

McLeod P, Plunkett K & Rolls ET: 1998. *Introduction to connectionist modeling of cognitive processes*, Oxford (UK): Oxford University Press.

Mobasher B, Cooley R & Srivastava J: 2000. Automatic personalization based on web usage mining, *Communication of the ACM*, 43(8), 142-151.

Mobasher B, Dai H, Luo T & Nakagawa N: 2001. Effective personalization based on association rule discovery from web usage data, *Proceedings of the 3rd international workshop on Web information and data management*, ACM Press, 9-15.

Mock KJ & Vemuri VR: 1997. Information filtering via hillclimbing, wordnet, and index patterns, *Information Processing and Management*, 33(5), 633-644.

Pitkow J: 1997. In search of reliable usage data on the www, *Proceedings of the 6th International WWW Conference*, Santa Clara (CA), April 7-11, 1997.

Riesbeck CK & Schank RC: 1989. *Inside Case-Based Reasoning*, Lawrence Erlbaum, Hillsdale.

Rucker J & Polanco MJ: 1997. Personalized navigation for the web, *Communications of the ACM*, 40, March 1997, 73-75.

Sarwar B, Karypis G, Konstan J & Reidl J: 2001. Itembased collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web*, ACM Press, 285-295.

Schank RC: 1982. *Dynamic Memory—A Theory of Reminding and Learning in Computers and People*, Cambridge University Press, Cambridge (Mass).

Shardanand U & Maes P: 1995. Social information filtering: algorithms for automating “word of mouth”, *ACM Conference Proceedings on Human Factors in Computing Systems*, Denver, CO, May 7-11 1995, 210-217.

Susi T & Ziemke T: 2001. Social Cognition, Artefacts, and Stigmergy: A Comparative Analysis of Theoretical Frameworks for the Understanding of Artefact-mediated Collaborative Activity, *Cognitive Systems Researc*h, 2(4), 273-290.

Takano H & Winograd T, 1998. Dynamic booksmarks. Managing personal navigation space by analysis of link structure and user behavior, *Hypertext 98*, Pittsburgh (PA), ACM, 297-298.

Twidale MB, Nichols DM & Paice CD: 1997. Browsing is a collaborative process, *Information Processing and Management*, 33(6), 761-783.

Van House NA: 2003. Digital Libraries and Collaborative Knowledge Instruction, *Digital Library Use. Social Practice in Design and Evaluation*, The MIT Press, Cambridge (Mass), 271-296.

Vora PR & Helander MG: 1997. Hypertext and its implications for the Internet, Helander MG, Landauer TK & Prabhu P (eds), *Handbook of Human-Computer Interaction*, Elsevier Science, Amsterdam, The Netherlands, 878-914.

Wasfi AMA: 1999. Collecting user access patterns for building user profiles and collaborative filtering, *Proceedings of the 4th international conference on Intelligent user interfaces*, ACM Press, 57-64.

Xiao YQ & Dunhamn MH: 2001. Efficient mining of traversal patterns, *DATA AND KNOWLEDGE ENGINEERING*, 39(2), 191-214.

Yan TW, Jacobsen M, Garcia-Molina H & Umeshwar Dayal U: 1996. From user access patterns to dynamic hypertext linking, *Computer Networks and ISDN Systems*, 28, 1007-1014.