Self-Organizing Information Matching in InformANTS

Rainer Hilscher, Sven Brueckner, Theodore C. Belding, H. Van Dyke Parunak
New Vectors LLC, 3520 Green Court, Suite 250, Ann Arbor, MI 48105
{rainer.hilscher, sven.brueckner, ted.belding, van.parunak}@newvectors.net

Abstract

In current information systems, information is passive. People act upon it, either sending it to known destinations or pulling it from known sources. InformANTS makes information active, enabling it to move actively from one user to another. This paper introduces the InformANTS vision and describes one of its major system components, the Information Matching System. Particular emphasis is placed on the distinctive self-organizing processes from which emerge the information exploration and exploitation capabilities of InformANTS.

1 Introduction

InformANTS (Information Agents for Networked Teaming Support) is a *research prototype* system currently developed under funding by the Disruptive Technology Office (DTO), for deployment in the US Intelligence Community (IC). It supports groups of analysts who are independently performing intelligence gathering and dissemination tasks by identifying collaboration opportunities based on overlapping interest and by actively routing relevant information to and between analysts.

The basic notion of InformANTS is a self-organizing information ecology where active information not only organizes into semantic neighborhoods but also drives collaboration by communicating these neighborhoods to appropriate target users.

To realize the above notion InformANTS operates with a very broad definition of a user. Users may be anything from intelligence analysts to databases. InformANTS thus removes one distinction that traditionally hampers information transfer: the distinction between sources and recipients of information.

This paper is organized as follows. Section 2 provides a general overview of the InformANTS architecture. The Information Matching System (IMS) is discussed in detail in section 3. Section 4 introduces ActiveWiki as a testbed for InformANTS and section 5 highlights short and long term challenges for this project.

2 Architecture Overview

Three main components make up InformANTS (see Figure 1): a Knowledge Organizing Environment (KOrE), a User Modeling System (UMS) and an Information Matching System (IMS). This paper focuses on the IMS with its various self-organizing processes.

Documents and users are represented in the KOrE as Information Packages, or InfoPacks. A series of Distillers is responsible for extracting InfoPacks from user actions and document contents.

InformANTS' User Modeling System (UMS) is an example of a Distiller, and constructs InfoPacks representing human users by observing the interactions of those users with contents already present in the KOrE.

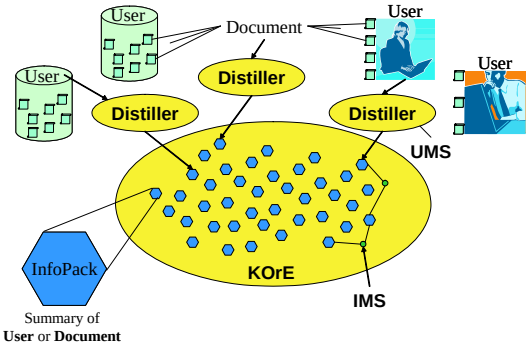

Figure 1: InformANTS Components

InfoPacks comprise three data containers. The "Head" functions as the anchor to the external source and holds the basic identifying information about an InfoPack (time of origin, type of user represented by this InfoPack).

It is also the target and origin for links to/from other InfoPacks. The "Body" houses the "raw" content of an InfoPack, concepts and relations in the form of specialized concept maps. The "Chest" is the location of metadata such as summaries, tags, keywords.

We distill incoming user information or document contents into Specialized Concept Maps. These maps comprise specific types of concepts (currently "Place", "Person", "Action", and "Thing") and a (small) set of relations among these concept types. In a previous project we used generic concept maps [6] to represent the contents of documents or analyst interest and queries, but without the specialization and restriction of concepts and relations, our reasoning tasks would quickly explode in complexity. Here we essentially restrict ourselves to represent geographic networks, social networks, hierarchical task networks, object ontologies, and events that link specific instances across these topologies.

3 Information Matching System (IMS)

The primary function of the Information Matching System (IMS) is to continuously maintain, for each user and each document represented by an InfoPack in the dynamically changing KOrE, a set of pointers to other relevant InfoPacks. Thus, the IMS links users to users (*Virtual Interest Group identification*), users to documents (*information pull*), documents to users (*information push*), and documents to documents (*Topic Group identification*).

The IMS achieves this functionality through a layered set of self-organizing processes that are linked through positive and negative feedback loops, exchanging information in a shared environment. We group these processes into two major classes. The Information Exploration processes maintain a task-independent organization of the (components of the) InfoPacks in the KOrE, while the Information Exploitation processes use this organization to establish the relevance relationships between specific InfoPacks. In other words, the function of the Information Exploitation processes is to make significant relevance relations between InfoPacks explicit for possible presentation to human users.

These two classes of processes are symbiotic. Without the Information Exploration processes, Information Exploitation would have a very low performance since it would operate essentially in a random information space, and without Information Exploitation, Information Exploration would get no performance feedback.

Our design of the Information Exploration and Exploitation processes follows our approach to engineering swarming systems [4] that we have successfully applied in domains as diverse as battlefield prediction, control of autonomous robotic vehicles, or problem-pattern matching for decision support in automotive design. In our approach, we design self-organizing populations of simple individual agents that act and interact only locally in a shared dynamical environment and the desired complex system-level functionalities are then an emergent property of these swarming agent systems.

In the following, we discuss our design for the Information Exploration and Exploitation processes that we are currently implementing for the Information Matching System.

3.1 Information Exploration

InfoPacks are still very complex constructs and thus inappropriate to serve as the atomic building blocks in the self-organizing processes that seek to create an organized information landscape which can be exploited efficiently. Therefore, we base our lowest layer of self-organization on the concepts and relations contained in the "Body" of InfoPacks.

It is our goal in the IMS to find InfoPacks that are closely related to each other, and thus we need to organize them in a common space in which relevance is expressed by distance, using the information contained in the InfoPack's Body. The Information Exploitation processes will then search the InfoPack Proximity Space (IPS) to establish the desired explicit links.

The mapping of an InfoPack from the unorganized (and dynamically changing) set of InfoPacks in the KOrE to the IPS is performed by three processes (concept-to-CPS mapping, CPS adjustment, CPS-to-IPS mapping) that execute in parallel and draw on the three sources of information available in the Body of an InfoPack – concept characteristics, concept relations, and overall map composition. For simplicity of exposition, we consider these processes as though they are three steps in a sequence (Figure 2).

In the first step, we map individual concepts by concept class into their respective Concept Proximity Space (CPS). In the second step, we adjust a concept's location in its CPS by the relations it has to other concepts in the InfoPack. Finally, in the third step, we combine the adjusted locations of all concepts in the Body of the InfoPack in their respective CPS into a unique location of the InfoPack itself in the IPS.

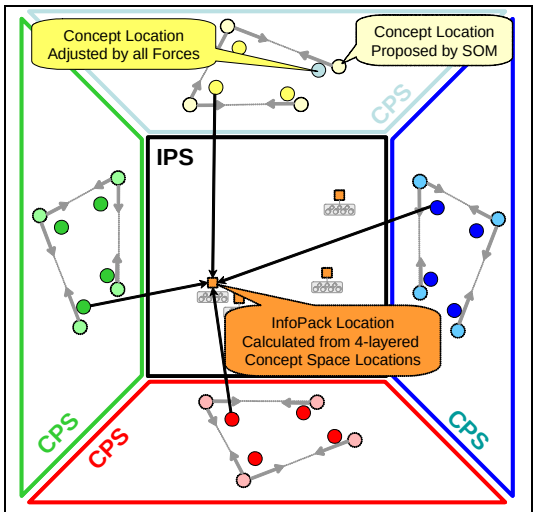

Figure 2: Derived InfoPack Location in IPS

Concepts in the specialized concept maps point to specific locations or regions in a high-dimensional abstract space that is generally non-metric and non-normalized. Consider, for instance, the "Person" concept that points into a space spanned by attributes, such as gender, age, nationality, name(s), or ethnicity. Though it may be possible to construct a similarity measure between such concepts using a subset of these attributes, typically it is very hard to usefully explore this space with sub-symbolic means, and it is computationally prohibitive to explore this space with symbolic means at the scale of our application. Thus, we need a mapping from the space of concepts of a single class as they currently occur in the InfoPacks to a low-dimensional, metric, and normalized space.

We selected Self-Organizing Maps (SOMs) [2], embedded in a 2-dimensional 1-unit-by-1-unit box and trained against the current set of concepts of one class, as our vehicle for the initial concept-to-CPS mapping. With a well-trained SOM, we are able to preserve the relative distance of concepts along the principal components of the high-dimensional space while, at the same time, embedding the concepts into a low-dimensional space that is most suitable for further swarming adjustment.

This SOM mapping, while it preserves the relative distance of concepts according to their specific characteristics (attribute settings), does not take into account the rich network of relationships that may exist between concepts. Therefore, we need to adjust the arrangements of concepts in each of the four CPSs to reflect these relationships.

The second step of the Information Exploration's mapping process uses Force-Based Clustering (FBC) [1] as a generic sub-symbolic method for the dynamic integration of guidance coming from different reasoners. In FBC, a population of particles arranges itself in an "energy-minimizing" configuration under the influence of attractive or repulsive forces of different kinds that are computed between particles. The particles find an "energy-minimizing" configuration in an iterative process where they move at finite speed along the force vector that is the sum of all component forces. In the IMS, particles are individual concepts located in the low-dimensional CPS. We initially consider two kinds of forces that affect the concepts' locations. The "home" force pulls a concept back to the location suggested by its SOM mapping, while the "relation" force attracts concepts to the current location of other concepts if they are linked in the InfoPack.

3.2 Information Exploitation

The Information Exploration processes dynamically form rich landscapes that can be searched efficiently and adjusted by performance feedback from the Information Exploitation processes. These landscapes are the Concept and InfoPack Proximity Spaces (several CPS and one IPS).

In our current system concept for the IMS we are drawing on our Polyagent construct (a single "complex" agent called an Avatar that uses "simple" probabilistic agents called Ghosts to explore complex spaces) [3] to establish useful links among InfoPacks.

The Avatar of an InfoPack's Polyagent continuously sends out Ghosts to explore the various information landscapes the InfoPack is embedded in. Some Ghosts explore the IPS, looking directly for nearby InfoPacks. Other Ghosts select specific concepts in the Body of the InfoPack and look for concepts of other (potentially related) InfoPacks in the concept's CPS. The Avatar collects and dynamically ranks the InfoPacks that have been identified by its Ghosts and thus maintains the desired list of relevant InfoPacks. As users assess this list, the Avatar may receive performance feedback, which it will communicate to the Information Exploration processes that provided the current ordering of InfoPacks.

4 Testbed: ActiveWiki

As a testbed for InformANTS, we intend to apply it to a Wiki, forming an "ActiveWiki". We are exploring an in-house Wikipedia server that will be augmented with the developing InformANTS system. The Wiki will initially be seeded with a full copy of the English version of Wikipedia.

In the ActiveWiki, the UMS observes users operating the Wiki interface and creates adaptive user model InfoPacks. These user models are stored in the KOrE, where the IMS enables the information to organize itself using swarming algorithms and dynamically identifies highly-relevant InfoPacks. The UMS processes these clusters to generate personalized virtual interest groups and data recommendations that the Wiki interface presents to the user.

5 Challenges

This paper has laid out a vision and a research agenda; there are many short- and long-term challenges that will have to be addressed to implement this vision.

5.1 Short Term

In the context of information exploration we must identify a function that will compute a meaningful location for an InfoPack in the IPS from the locations of each of its constituent concepts in their CPSs. Another short-term challenge is to fine-tune the relative strengths of the SOM and FBC in computing the location of a concept. Finally, we must develop meaningful non-numerical metric spaces for such concepts as "role".

5.2 Long Term

In the first stage of InformANTS implementation we concentrate on clustering InfoPacks according to the content of their bodies. Information exploration based on "Chest" content (meta-information) of InfoPacks is planned for the second stage. Self-Organizing Data and Search (SODAS) is a technology for dynamic, decentralized, any-time hierarchical clustering of objects, based on a defined set of object features [5]. We plan to integrate SODAS with InformANTS in order to produce multiple hierarchical cluster trees of InfoPacks, based on the features or tags contained in the InfoPacks' chests. Each cluster tree would provide a different "view" of the InfoPacks, clustered according to different criteria; these trees constitute a third type of space in addition to the CPSs and the InfoPack space.

6 Conclusion

In this paper we have presented our initial design for the IMS as a main component of the InformANTS architecture. The IMS comprises a synergistically linked set of self-organizing processes which dynamically maintain relevance relations between InfoPacks, representing information about users or documents outside the InformANTS system. Specifically, we distinguish between Information Exploration processes that produce an organized information landscape and Information Exploitation processes that search this landscape and link InfoPacks that are relevant to each other.

ACKNOWLEDGMENTS

This research was sponsored by the Air Force Research Laboratory, Air Force Materiel Command, USAF, under Contract number FA8750-06-C-0193. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

7 References

[1] P. Eades. "A heuristic for graph drawing." *Congresses Numerantium*, 42:149-160, 1984.

[2] T. Kohonen. *Self-Organizing Maps*. Heidelberg, Springer-Verlag, 1995.

[3] H. V. D. Parunak and S. Brueckner. "Modeling Uncertain Domains with Polyagents." In *Proceedings of International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'06)*, Hakodate, Japan, ACM, 2006.

[4] H. V. D. Parunak and S. A. Brueckner. "Engineering Swarming Systems." In F. Bergenti, M.-P. Gleizes, and F. Zambonelli, Editors, *Methodologies and Software Engineering for Agent Systems*, pages 341-376. Kluwer, 2004.

[5] H. V. D. Parunak, R. Rohwer, T. C. Belding, and S. Brueckner. "Dynamic Decentralized Any-Time Hierarchical Clustering." In *Proceedings of Proceedings of the Fourth International Workshop on Engineering Self-Organizing Systems (ESOA'06)*, Hakodate, Japan, Springer, 2006.

[6] P. Weinstein, H. V. D. Parunak, P. Chiusano, and S. Brueckner. "Agents Swarming in Semantic Spaces to Corroborate Hypotheses." In *Proceedings of AAMAS 2004*, New York, NY, pages 1488-1489, ACM, 2004.