The Web and Complex Adaptive Systems

Maya Rupert
Thompson Rivers University, Canada
mrupert@tru.ca

Salima Hassas
LIRIS–Claude Bernard University, Lyon1 France
hassas@liris.cnrs.fr

Amjad Rattrout
LIRIS–Claude Bernard University, Lyon1 France
amjad.rattrout@liris.cnrs.fr

Abstract

The web continues to grow at a phenomenal rate and the amount of information on the web is overwhelming. Finding the relevant information remains a big challenge. Due to its wide distribution, its openness and high dynamics, the WWW is a complex system, for which we have to imagine mechanisms of content maintaining, filtering and organizing that are able to deal with its evolving, dynamics and distribution. Integrating mechanisms of self-organization of the web content is an attractive perspective, to match with these requirements. This paper explores the web from a complex adaptive system (CAS) perspective. It reviews some characteristic behaviours of CAS and shows how the web exhibits similar behaviours. We illustrate with a model for web content organization adopting the CAS vision and using the multi-agent paradigm.

1. Introduction

The World Wide Web is growing at a tremendous rate. It contains a huge amount of unstructured, distributed, multi-media data. This content provides a great potential source for knowledge acquisition that needs to be filtered, organized, and maintained in order to permit an efficient use. The wide distribution of the web, its openness and high dynamics make any tentative of content organization or maintaining, a task very hard to achieve. The web is a complex open dynamic network, exhibiting a self-organizing adaptive behaviour similar to a complex adaptive system CAS. In this paper, we analyze the web from a CAS perspective and present an approach for organizing dynamically the content on the web adopting the CAS vision and using the multi-agent paradigm. The approach uses the stigmergy mechanism for agents' interactions and communications.

In the next section a review of CAS characteristic behaviours is presented outlining how the web exhibits similar behaviours. In section 3, we propose an approach for organizing the web content from a CAS perspective. Related work is outlined in section 4 and conclusion in section 5.

2. CAS properties and the web

There is no single definition of Complex Adaptive Systems (CAS). CAS are composed of many interacting parts or agents, giving rise to emergent patterns of behaviour. The behaviour is said to be emergent because at the macroscopic level, the system exhibits new complex properties that are not found at the local level of the different components [1]. CAS self-organize and adapt to changes in the environment without central control or rules governing their behaviours. In such systems order can emerge through a process of self-organization.

2.1. Emergent order and self-organization

From the local interaction between the different agents emerges an organized global behaviour of the system. In ant colonies, each ant follows a pheromone trail laid by other ants to get to the food source. The ant reinforces the trail by laying more pheromone of its own for other ants to follow the trail. The pheromone evaporates over time. In the web graph, emergence of scaling is noted in [2]. The probability that a web page has k pages linking to it (indegree) or a web page is linked to k pages (outdegree), follows the power law distribution $P(k)= k^{-\gamma}$. The web graph is not a random network, but it exhibits the property of a scale free network, which develops a degree of self-organization.

In the web, there is no global control or authority governing web page creation. Web authors are free to add and delete pages and websites and create hyperlinks to any page or node in the web graph. Despite this decentralized process, the web self-organizes into web communities based on hyperlinks structure analysis. In [3], Flake defined a web community as "a collection of web pages such that

Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA'06)
1550-445X/06 $20.00 © 2006 IEEE

each member page has more hyperlinks in either direction" within the community than outside of the community".

2.2. Adaptation, co-evolution and dynamics

At the edge of chaos where order starts to disappear, agents need to adapt to a changing environment. They change their internal models and behaviours according to their tempo spatial organization. Since its creation, the structure, content and usage of the web have been coevolving and adapting to each other. Websites personalization and adaptation have emerged as the number of users is increasing constantly and there is a higher need for websites providers to adapt their websites to different usage and to deliver better content. CAS change constantly because of the continuous interactions and interdependence between the different agents and their environment. The web is a dynamic graph and it evolves by constantly adding new pages and removing some old ones (Growth model). There is a higher probability to link new pages to a more connected page (preferential attachment) [2] and a good example is the Google web page.

2.3. Holland's properties

John Holland identified seven basic elements of a CAS:
Aggregation is the property by which agents group to form meta-groups leading to the complex system. Web pages can be grouped into websites, and websites into web communities that emerge and self organize. Aggregate behaviour is also observed in the appearance of hubs and authorities in the web [4]. **Tagging** is the mechanism of assigning attributes or tags used for agent identification. A tag could be the main topic of a web community or the word vector "bags of words" of a web page used in text analysis. **Non-linearity** is the property where the emergent behaviour of the system is the result of a non-proportionate response to its stimulus. **Flows** are the physical resources circulating through the nodes of a complex network. The **diversity** of skills, experiments, strategies, rules of different agents ensure the dynamic adaptive behaviour of a CAS. **Internal models** are the functions agents use to interact with each other and with their environment. **Building blocks** are the component parts that can be combined and reused for each instance of a model. Identifying these blocks is the first step in modelling a CAS.

2.4. Stigmergy and the web

Stigmergy is a concept introduced by Grassé [5]. He noticed that the behaviour of the termites during the construction of their nests is influenced by the structure of the constructions themselves. This mechanism allows the self-structuring of the environment through agents' activity: the environment state and the current repartition of agents in the environment determines their respective future evolutions. Several attempts to associate the web with the notion of stigmergy were analyzed lately especially to create self-organized websites. The site structure would be flexible. Website visitors interact with each other, changing its structure and content with the help of software agents. In [6], the concept of stigmergy was linked to the web, showing that weblogs, web communities, and the search engine Google exhibit a stigmergic behaviour.

2.5. Programming CAS with agents

Hassas has proposed a framework for developing CAS for complex networks such as the internet and the web [7] combining Holland's basic properties to the stigmergy mechanism:

-Using the situated multi-agents paradigm and behavioural intelligence for identifying the building blocks (agents), their internal models and their roles.

-Using the web as the physical environment to facilitate the emergence of aggregates and the flows of information.

-Using a mechanism of communication between agents, based on a spatial representation and mediated by the environment, such as the stigmergy mechanism. This favours the aggregation of control information and its spreading through the distributed environment.

-Maintaining equilibrium between exploration and exploitation in the behaviour of different agents, to allow aggregation (reinforcement) of the building blocks and diversity (randomness).

3. Case study: An approach for dynamic organization of the web content

3.1. The WACO system

The following model illustrates how the CAS principles are applied in the context of web content organization. WACO (Web Ants Content Organization) is an approach inspired by social insects to organize dynamically the web content [8]. In this approach, the web is considered as a complex environment, inhabited by artificial creatures called

WebAnts. These creatures implemented by mobile agents, are organized in a colony and mimic some behaviours of natural ants, namely the collective sorting behaviour and the food foraging behaviour. The content of the web is viewed by WebAnts as a potential source of food that needs to be organized and collected in an efficient way. The WebAnts in WACO are created in a dynamic way and they adapt to their environment and co-evolve. This process requires a mechanism of managing and regulating the population of agents. To do so, the mechanism of energy distribution and consumption is used [9]. WebAnts are sensitive to some notion of order, which is obtained by semantic organization of the web content. The higher the disorder on the web, the more active are WebAnts. Activity of agents is regulated by a mechanism of energy distribution, provided by the environment and directly associated with the notion of order in the global environment. Disorder in the environment, generates energy which is captured by agents making them increase their activity and number. Based on their fitness function defined by order/disorder, two mechanisms direct their life cycle: *duplication* (birth) and *disappearance* (death).

3.2. Documents coding by synthetic pheromone use

Documents contained by the websites are considered as objects to be sorted following their semantic contents, so as to construct semantic clusters, where a cluster is a set of semantically closed documents with respect to a predefined similarity measure. A semantic value is associated to each document. This value is used as a label for a specific pheromone to which WebAnts would be sensitive, when looking for documents with similar semantic values. Each semantic topic is identified by a kind of pheromone. Synthetic pheromone is coded by a structure with these fields:

- *Label* (W_K): characterizes the kind of information coded by the pheromone, which is in our context the semantic value of a document (weighted keyword) [10]

$$W_k = L_C . H_C . T_F . IDF_K$$

T_F is the frequency of the keyword in the current document k, H_C is a Header constant ($H_C > 1$ if the word appears in a title, =1 otherwise), and IDF_K is the inverse of document frequency. The linkage constant L_C ($L_C > 1$ if the word appears in a link, =1 otherwise)

- *Intensity* (τ_{ij}): expresses the pertinence of information. This value is computed at each site i, for each topic j, using the number of documents addressing the same topic, each time (t+1) a new document is added, as:

$$\tau_{ij}(t+1) = \rho_j \tau_{ij}(t) + \Sigma_{k=1, |Dij|} \Delta\tau_{ij}^k(t)$$

ρ_j represents the persistence rate ((1-ρ_j) the evaporation rate), $\Delta\tau_{ij}^k$ (t) the intensity of pheromone emitted by a document k, on the site i for a topic j at time t, and Δ_{ij} is the set of documents addressing topic j on the site i.

- *Evaporation rate*: expresses the persistence rate of information in the environment. The lower its value, the longer is the influence of the spread information

$$\rho_j = |D_{ij}| / |D_i|$$

D_{ij} is the set of documents addressing the topic j on the site i, and D_i is the set of all documents on the site i.

- *Diffusion rate*: expresses the distance to which information is spread in the environment, the higher its value the greater the scope of the information in the environment. We associate to each site i, a distance d_{ij} for each topic j addressed by i, which is computed as the longest path from the site i to the last site addressing the topic j, following a depth first search

$$d_{ij} = Max_k (d_{ij}^{\ k})$$

k is the number of links addressing topic j, from a site i.

3.3. Agents actions and interactions

The web represents a complex uncertain environment that WebAnts explore, and structure by interacting with one another. It is also the medium of their interactions, through the deposit and smelling of the synthetic pheromone. They communicate through the stigmergic mechanism. The building blocks are the populations of WebAnts agents, which mimic the collective sorting behaviour and the foraging behaviour. Aggregation is observed through the multi-pheromone structure.

Four types of WebAnts were created, each assigned a different task (tags associated with each agent): *Explorers WebAnts* look randomly for web documents to sort in perceiving different kinds of pheromones, corresponding to different semantic values. *Collectors WebAnts* maintain and organize semantically collected documents, by computing regularly the synthesis of the site pheromone, and updating the values of its associated parameters (label, intensity, persistence rate, diffusion rate). *Searchers WebAnts* reinforce clusters of collected documents by searching the web for similar documents to add to the cluster. Finally, *Requests Satisfying WebAnts* code the user request into a pheromone value, and search in the environment the appropriate cluster. Simulations were conducted to test the effectiveness of this approach in [11]. Some of the obtained results are presented below.

3.4. Results

Experiment 1: Order increasing and maintained

We studied the notion of order on the web. This notion was associated to the emergence of clusters with similar semantic contents. The function of local order for a given site is expressed by the number of sites in its neighbourhood, with a similar content. Similarity is computed with respect to a specified threshold distance between associated weighted keyword vectors.

On the figure below, the x-axis represents the number of iterations (time scale) and the y-axis represents the number of documents. Disorder decreases regularly in the system whereas new documents apparition increases. Disorder is measured by the total number of documents minus the number of clustered documents.

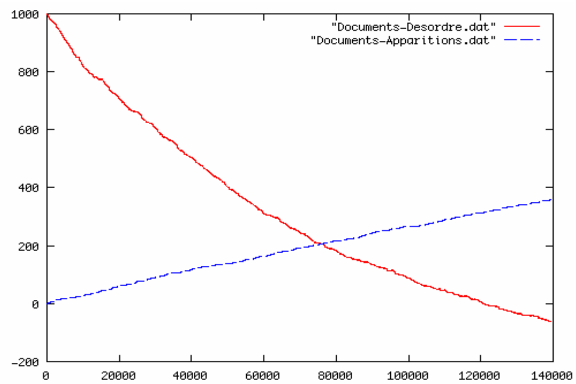

Figure 1. Disorder decreases while new documents are created, and sorting occurs.

Experiment 2: Clusters forming and size evolution

Figure.2 represents the evolution during time (x-axis) of the mean size of clusters (y-axis). We can observe the evolution of clusters forming and the increase of their sizes. This figure shows, when compared to the figure above, how the clustering behavior reinforces the creation of clusters, while the disorder decreases in the system

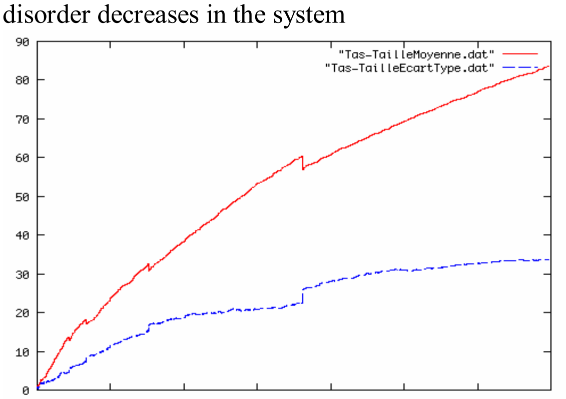

Figure 2. Clusters mean size and its standard deviation evolution.

Experiment 3: Energy distribution and evolution

In figure.3, we show the mean value of energy of specialized agents (Searcher WebAnts). These agents increase their amount of energy during the formation of clusters. When clusters are formed (time 80000), they disappear (energy value = 0). At time 100000, a sudden increase of energy that is associated to the apparition of new clusters, as new documents are discovered.

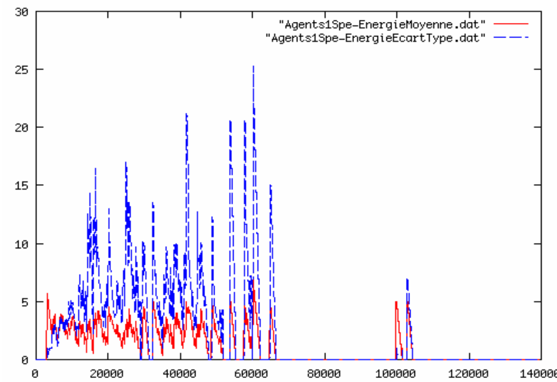

Figure 3. Mean values and standard deviation values of Searchers WebAnts energy evolution

Experiment 4: Evolution of population of agents and regulation of their activities

The following figure (y-axis represents size of population, x-axis represents time) shows the evolution of agents' populations in the system and the proportion of active agents with respect to the whole population.

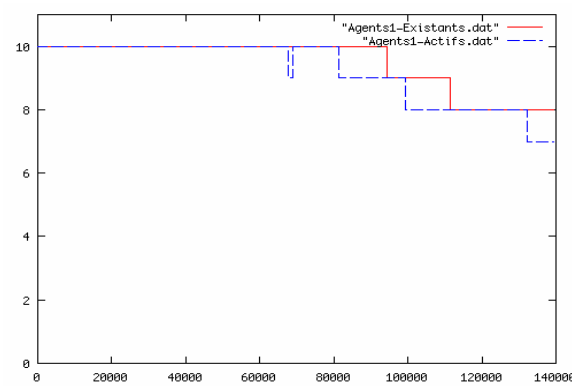

Figure 4. Population evolution: proportion of active agents/total population of agents

We can observe that until time 80000, which is the time of emergence of a global order, all agents are active, and past this time, the number of active agents decreases. This makes inactive agents disappear. All

Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA'06)

agents are active again during the formation of new clusters, as new sites are created (time 100000).

4. Related work

Search engines have been proposed (AltaVista, Excite, Google) providing users with more efficiency, performance and diversity of services offered. However, techniques used by current search engines suffer lack of pertinence of content-based retrieval processes, and inaccuracy with regards to the fast dynamic evolution of web content. This insufficiency is mainly due to the fact that search engines create indexes of web documents, which could be viewed as "snapshots" of the web. These indexes need to be updated as frequently as changes occur on the web. Some systems use crawlers that exhaustively visit and revisit every web page to maintain its content, but as this maintaining is achieved periodically, at any given time an index will be somewhat inaccurate or incomplete. In our approach, the system follows the permanent evolution of the web content and updates its changes as frequently as they occur.

Websites adaptation and personalization [12-15] is another approach to deal with users requests. As web content, and users interests are both highly evolving, personalized search engines must handle long-standing queries by adapting to users profile evolution and contributing to its diversification, by discovering new fields of interests, following the web content evolution. In our approach we deal with these issues by considering different kind of interacting populations of agents, each with a specific role: users requests management, users profiles enrichment, dynamic semantic clustering of web content and content refinement and enrichment.

5. Conclusion

In this paper, we showed that the web exhibits a complex adaptive system-like behaviour. A modeling approach was proposed using Holland's properties combined with the stigmergy mechanism, to overcome its evolving complexity. The WACO model follows the CAS perspective for organizing dynamically the web content. We are currently developing an agent-based approach based on CAS principles that focuses on the association of a semantic to information contained on the web, through the combination of the web content, usage and structure.

6. References

[1] J. Holland, *Hidden Order*: Addison-Wesley, 1995.

[2] A. L. Barabasi and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, pp. 60-69, 2003.

[3] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, "Self-Organization of the Web and Identification of Communities," *IEEE Computer*, vol. 35, pp. 66--71, 2002.

[4] J. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," presented at ACM-SIAM Symposium on Discrete Algorithms, 1998.

[5] P. P. Grassé, "La reconstruction du nid et les coordinations inter-individuelles chez bellicoitermes natalenis et cubitermes, la théorie de la stigmergie," *Insectes Sociaux*, vol. 6, pp. 41-81, 1959.

[6] J. Gregorio, "Stigmergy and the World-Wide Web," 2003.

[7] S. Hassas, "Systèmes complexes à base de multi-agents situés," University Claude Bernard Lyon, 2003.

[8] S. Hassas, "Using swarm intelligence for dynamic web content organization," presented at IEEE Swarm Intelligence Symposium, Indianapolis, USA, 2003.

[9] F. Menczer and R. Belew, "Adaptive retrieval agents: Internalizing local context and scaling up to the web," *Machine Learning*, vol. 9, pp. 203-242, 2000.

[10] S. Brin and L. Page, "The Anatomy of a Large-Scale hypertextual Search Engine," *Computer Networks and ISDN Systems*, 1998.

[11] S. Hassas, "Using swarm intelligence for dynamic web content organization," presented at IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, 2003.

[12] M. J. Pazzani and D. Billsus, "Adaptive Web Site Agents," *Autonomous Agents and Multi-Agent Systems*, vol. 5, pp. 205--218, 2002.

[13] M. Eirinaki, C. Lampos, S. Paulakis, and M. Vazirgiannis, "Web personalization integrating content semantics and navigational patterns," in *Workshop on Web information and data management*: ACM Press, 2004, pp. 72--79.

[14] F. Menczer, "Complementing search engines with online web mining agents," Elsevier Science Publishers B. V., 2003.

[15] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, "Web usage mining as a tool for personalization: A survey," *User Modeling and User-Adapted Interaction*, vol. 13, pp. 311-372, 2003.