Available online at www.sciencedirect.com
ScienceDirect
Journal of Computer and System Sciences 74 (2008) 133–145
JOURNAL OF COMPUTER AND SYSTEM SCIENCES
www.elsevier.com/locate/jcss

The web from a complex adaptive systems perspective ☆

Maya Rupert [a,b,*], Amjad Rattrout [b], Salima Hassas [b]

[a] *Thompson Rivers University, Kamloops, BC, V2C 5N3, Canada*
[b] *LIRIS-University of Lyon 1, 69622 Villeurbanne Cedex, France*

Received 15 June 2006; received in revised form 31 October 2006

Available online 20 April 2007

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 web 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. Self-organized complex systems can be programmed using situated multi-agent systems with a coupling between the agents' social organization and spatial organization. This paper explores the web from a complex adaptive system (CAS) perspective. It reviews some characteristic behaviors of CASs and shows how the web exhibits similar behaviors. We propose a model and a prototype of a system that addresses the dynamic web content organization, adopting the CAS vision and using the multi-agent paradigm.

Keywords: Web; Complex adaptive systems; Multi-agent systems; Web content organization

1. Introduction

The 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 efficient use. The web content's organization and maintenance are tasks hard to achieve because of the wide distribution of the web, its openness and high dynamics. The web is a complex open dynamic network, exhibiting a self-organizing adaptive behavior similar to a complex adaptive system (CAS). In this paper, we analyze the web from a CAS perspective and present a methodology for programming complex systems using the multi-agent paradigm. Adopting the CAS vision, we propose WACO, an approach inspired by social insects for organizing dynamically the content on the web.

The paper is organized as follows: in the first section, a review of CAS characteristic behaviors is presented, outlining how the web exhibits similar behaviors. A new methodology for programming complex systems is presented in the second section and in section three, we describe a model for organizing the web content in which we adopt the

☆ Expanded version of a paper presented at the Workshop Web And Mobile Information Systems (WAMIS 2006).

* Corresponding author.
E-mail addresses: mrupert@tru.ca (M. Rupert), amjad.rattrout@liris.cnrs.fr (A. Rattrout), hassas@liris.cnrs.fr (S. Hassas).

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2007.04.001

CAS perspective and based on the methodology outlined in section three. In section four, we present experiments to demonstrate the effectiveness of this approach, and discuss some observed results. Related work is outlined in section five, and we conclude in section six.

1.1. Complex adaptive systems

There is no single definition of Complex Adaptive System (CAS). Most researchers in this field [1,2] agree that CASs are composed of many interacting parts, giving rise to emergent patterns of behavior. The behavior 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. CASs self-organize and adapt to changes in the environment without central control or rules governing their behaviors. CASs are non-linear systems where the "whole is more than the sum of its components" [1]. Each CAS is formed of many interacting agents. From the local interaction of individual agents arises a global behavior or a pattern that cannot be predicted at the local level. In such systems, order can emerge through a process of self-organization [2]. The study of CASs has been applied to different fields such as economies [3], organizations [4], ecologies [5], biology, the immune system and the brain. CASs share several common characteristic features. In the next sections a review of these characteristics will be presented, outlining how these features apply to the web.

1.2. Emergence and self-organization

The main characteristic of a complex system is that it has many interacting components or agents. Examples include atoms, molecules, neurons in the brain, and ants in a colony. These semi-autonomous agents are interconnected, and they interact with each other to form the behavior of the whole system. In the web, there is a large number of agents including users, web authors, search engines, web pages, hyperlinks, web services and software agents interacting with each other in a non-linear way.

From the local interaction between the different agents emerges an organized global behavior of the system. The whole is more than the sum of its parts, and reductionism theory cannot be applied to understand this emergence. The web has been represented by a directed graph whose nodes or vertices are the web pages and whose arcs or edges are the hyperlinks between the web pages [6]. In the web graph, the emergence of scaling is noted [7], i.e. 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 $P(k) = k^{-\lambda}$. 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. This topology shows how order is emerging on the web.

In a CAS, there is no centralized control that dictates the system's overall behavior. Agents govern their own rules of behavior at the local level, adapt to their environment, and at the macroscopic level, order emerges. 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. A web community can be defined as "a collection of web pages such that each member page has more hyperlinks in either direction within the community than outside of the community" [8].

1.3. 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 behaviors according to their temporal-spatial organization. They also co-evolve to ensure survival in the new environment. In fact, the adaptive behavior of the system cannot be the result of completely random dynamics. Holland [9] describes the evolution of these systems as the result of a strategy combining (random) exploration to maintain a certain diversity, and exploitation to reinforce promising tracks, allowing adaptation. 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 in order to deliver better content.

A CAS changes constantly because of the continuous interactions and interdependence between the different agents and their environment. The dynamics are non-linear, and one cannot predict the behavior of a CAS. The web is a dynamic graph, and it evolves by constantly adding new pages and removing some old ones (Growth model). There