
Organisation du contenu du Web selon la perspective des Systèmes Complexes Adaptatifs

Maya Rupert***—Amjad Rattrout**—Salima Hassas**

**Thompson Rivers University
Box 3010 – 900 McGill Road, Kamloops, B.C., Canada V2C 5N3
mrupert@tru.ca*

***LIRIS - Bat Nautibus, Université Claude Bernard-Lyon 1
43 Bd du 11 Novembre 1918, 69622 Villeurbanne Cedex
amjad.rattrout@liris.cnrs.fr
hassas@liris.cnrs.fr*

RÉSUMÉ. Le web continue à croître à une vitesse phénoménale. Trouver l'information pertinente sur le web reste toujours un grand défi. La large distribution, l'ouverture et la grande dynamique qui caractérisent le web, en font un système qui exhibe toutes les caractéristiques des systèmes complexes. Afin d'appréhender la complexité d'un tel système, il faudrait alors imaginer des mécanismes de maintenance, de filtrage et d'organisation de son contenu qui puissent prendre en compte sa dynamique évolutive et sa distribution. Intégrer des mécanismes d'auto-organisation est une perspective attractive qui répond à ces exigences. Cet article analyse le web selon la perspective des systèmes complexes adaptatifs (CAS). Il examine quelques comportements caractéristiques des CAS et montre que le web exhibe des comportements similaires. Cette analyse sera illustrée par un système d'organisation du contenu du web, adoptant la vision CAS et utilisant le paradigme multi-agents.

ABSTRACT. The web continues to grow at a phenomenal rate and the amount of information on the web is overwhelming. Finding the relevant information on the web 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. 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.

MOTS-CLÉS : organisation du contenu du web, systèmes complexes adaptatifs, auto-organisation, stigmergie.

KEYWORDS: organizing web content, complex adaptive systems, self-organisation, stigmergy.

1. Introduction

Le web continue à croître à une vitesse phénoménale. Il contient une grande quantité de données non-structurées, distribuées et multi-medias. Ce contenu fournit une source riche en information qui a besoin d'être maintenu, filtré et organisé pour permettre un usage efficace. La tentative d'organiser et de maintenir le contenu du web est une tâche difficile à réaliser vu sa large distribution, son ouverture et sa forte dynamique. Le web est un réseau complexe, ouvert et dynamique exhibant un comportement adaptatif et auto-organisé similaire à un système complexe adaptatif (CAS). Dans cet article, nous analysons le web selon la perspective des CAS et nous présentons une approche pour l'organisation dynamique du contenu du web adoptant la vision des CAS, et utilisant le paradigme multi-agents. L'approche utilise le mécanisme de stigmergie pour l'interaction et la communication entre les agents.

Dans la section suivante, une revue des caractéristiques des CAS est présentée, montrant que le web exhibe des comportements similaires. Dans la section 2, nous proposons une approche pour l'organisation dynamique du contenu du web selon la perspective des CAS. Nous terminerons cet article par une revue des travaux relatifs à notre problématique et une conclusion.

Il n'existe pas une définition unifiée des Systèmes Complexes Adaptatifs CAS. La majorité des chercheurs dans ce domaine (Holland, 1995, Kauffman, 1993) admettent que les CAS sont composés de plusieurs éléments en interaction continue, créant des comportements émergents. Le comportement est dit émergent parce qu'au niveau macroscopique, le système génère de nouvelles propriétés complexes qui ne peuvent pas être observées au niveau local des différentes composantes. Ce comportement global est le fruit d'une agrégation non linéaire des comportements locaux. Les CAS s'auto-organisent et s'adaptent aux changements de leur environnement en l'absence d'un contrôle central gouvernant leurs comportements. Les CAS sont des systèmes non linéaires dont "le tout est plus que la somme de ses parties "(Holland, 1995). Dans ces systèmes, l'ordre émerge à travers des processus d'auto-organisation.

1.1. *Ordre émergent et auto-organisation*

La caractéristique principale d'un système complexe adaptatif est qu'il est formé de plusieurs éléments ou agents interagissant entre eux. On peut citer en exemple : les atomes, les molécules, les neurones dans le cerveau, les fourmis dans une colonie et l'air et l'eau dans le système météorologique. L'interaction entre ces agents semi-autonomes génère le comportement du système. De la même façon, le web met en jeu plusieurs acteurs (agents) y compris les utilisateurs, les auteurs du web, les moteurs de recherche, les pages web, les hyperliens, les services web et les agents web qui interagissent réciproquement de façon non linéaire.

De l'interaction locale entre ces différents agents émerge un comportement global du système. Le comportement émergent se présentant au niveau macro ne pourrait pas être prédit des niveaux micros. Une approche réductionniste ne peut pas s'appliquer pour comprendre ce comportement. Dans les colonies de fourmis, chaque fourmi suit une piste de phéromone déposée par les autres fourmis pour aboutir à la source de nourriture. La fourmi renforce la piste en déposant plus de phéromone pour que d'autres fourmis puissent suivre la même piste. La phéromone s'évapore avec le temps. Ce procédé est basé sur des boucles de rétroactions positives, et négatives et le comportement collectif émergent est dit auto-catalytique. Dans le web, l'émergence est observée à plusieurs niveaux. La probabilité qu'une page web ait k pages entrantes (pointant sur elle : *indegree*) ou qu'une page web pointe sur k pages (*outdegree*), suit la loi de puissance $P(k) = k^{-\beta}$ (Barabasi *et al.*, 2003). Le graphe du web n'est pas un réseau aléatoire, il possède la propriété des réseaux invariant d'échelle (*scale free networks*), qui est la caractéristique type des réseaux résultants d'une activité auto-organisationnelle. Cette topologie montre que l'ordre émerge sur le web malgré le contrôle décentralisé dans l'ajout de nouveaux nœuds. Le réseau s'est révélé robuste dans le sens où des suppressions aléatoires n'affectent pas le diamètre ou la connectivité du graphe mais le système devient vulnérable à des attaques ciblées qui suppriment les nœuds fortement connectés.

Dans les systèmes complexes adaptatifs, le contrôle qui dicte le comportement du système est non centralisé. Les agents gouvernent leurs propres règles de comportement au niveau local, s'adaptent à leur environnement, et au niveau macroscopique, l'ordre émerge. Dans le web, il n'y a pas d'autorité globale qui gouverne la création des pages web. Les auteurs du web sont libres d'ajouter et effacer des pages et des sites web et de créer des hyperliens à n'importe quelle page. Malgré ce procédé décentralisé, le web s'auto-organise automatiquement pour créer des communautés. (Flake *et al.*, 2002) ont défini une communauté web comme étant "une collection de pages tel que chaque page a plus d'hyperliens dans la communauté qu'à l'extérieur de la communauté".

1.2. *Adaptation, co-évolution et dynamique*

Au bord du chaos, où l'ordre commence à disparaître, les agents ont besoin de s'adapter à un environnement changeant. Ils changent leurs modèles et leurs comportements internes suivant leur organisation temporelle et spatiale. Ils co-évoluent pour assurer la survie dans le nouvel environnement. Depuis sa création, la structure, le contenu et l'usage du web ont co-évolué tout en s'adaptant l'un à l'autre. La personnalisation et l'adaptation des sites web a émergé comme le nombre d'utilisateurs augmente constamment, et les créateurs des sites web doivent adapter leurs sites aux différents usages et délivrer un meilleur contenu (Menczer, 2003, Perkowitz *et al.*, 2000).

Les CAS changent constamment à cause des interactions continues entre les agents et leur environnement. Le web est un graphe dynamique qui évolue en ajoutant constamment de nouvelles pages et en modifiant les pages existantes (le modèle de croissance). La probabilité de relier de nouvelles pages à une page populaire est plus grande (le modèle d'attachement préférentiel) ou "le riche devient encore plus riche" (Barabasi, *et al.*, 2003). Un bon exemple de cette situation est la page web de Google. Les dynamiques du web, son âge, sa durée de vie ont été étudiées dans (Brewington *et al.*, 2000, Cho *et al.*, 2000) où il a été montré que les changements des pages du web peuvent être modélisés par un processus de Poisson.

1.3. *Les caractéristiques de Holland*

Comme les CAS sont formés d'un ensemble d'agents interagissant les uns avec les autres, s'adaptant et co-évoluant dans leur environnement, modéliser de tels systèmes requiert une approche ascendante qui consiste à identifier les différents agents, leurs règles de comportements et leurs interactions. John Holland a identifié sept éléments fondamentaux d'un CAS (Holland, 1995):

L'agrégation est la propriété par laquelle les agents se regroupent pour former des des meta-agents qui peuvent à leur tour se combiner à un plus haut niveau (meta-meta-agents) ce qui aboutit à un système complexe. L'émergence du meta-agent est due aux actions réciproques entre les agents du niveau inférieur. Le contenu et la structure basés sur les besoins des utilisateurs sont groupés dans une page web et des pages web dans des sites web et des sites web dans les communautés du web (les meta-agents) qui émergent et s'auto-organisent sans contrôle centralisé. Cette auto-organisation est un résultat d'une interaction rétroactive entre l'usage, le contenu et la étiquettes utilisées pour l'identification structure. **Le tagging** ou l'étiquetage est le mécanisme qui facilite la formation des groupes en allouant des étiquettes aux différents agents. Une étiquette pourrait être le thème principal d'une communauté du web ou le " vecteur de mots-clefs " d'une page web spécifique utilisée dans l'analyse de texte et l'analyse de similarité des pages web. **La nonlinéarité** est la propriété où le comportement émergent du système est le résultat d'une réponse non-proportionnelle à son stimulus. Le comportement global provenant des actions réciproques entre les agents est plus compliqué qu'une simple somme ou moyenne des agents individuels. Ainsi le système ne peut pas être prédit en comprenant simplement comment chaque composante fonctionne et se comporte. La croissance du web est un processus non linéaire. **Les flux** sont les ressources physiques ou les informations qui circulent à travers les nœuds d'un réseau complexe. Dans le cas du web, ce sont les informations, les requêtes, les navigations, etc qui traversent les sites web, considérés comme des nœuds du réseau complexe qu'est le web. **La diversité** des expériences, stratégies, et règles des différents agents garantissent le comportement adaptable dynamique d'un système complexe adaptatif. Le web a un grand nombre de composants réagissant réciproquement et cette diversité dans le web contribue à sa robustesse. Nous observons la diversité dans son usage, structure

et contenu. Dans (Liu *et al.*, 2002) les utilisateurs ont été classés en trois catégories les utilisateurs aléatoires, les utilisateurs rationnels et les utilisateurs périodiques. **Les modèles internes** ou schémas sont les fonctions qui gouvernent l'emploi des agents pour communiquer les uns avec les autres et avec leur environnement. Ces schémas dirigent les comportements des agents. **Les blocs de base** sont les parties composantes qui peuvent être combinées et réutilisées pour chaque cas d'un modèle. (Milo *et al.*, 2002) ont montré que les motifs des sous-graphes forment la structure des blocs de base pour le réseau WWW.

1.4. *La stigmergie et le web*

La notion de stigmergie a été introduite par le biologiste français Grassé (Grassé, 1959), pour comprendre le comportement de coordination des constructions chez les insectes sociaux et l'interaction avec leur environnement. Cette interaction aboutit à un changement dans l'environnement qui affectera les actions futures des agents. "La coordination des tâches, la régulation des constructions ne dépendent pas directement des ouvriers, mais des constructions elles-mêmes. L'ouvrier ne dirige pas son travail, il est guidé par lui. C'est à cette stimulation d'un type particulier que nous donnons le nom de STIGMERGIE ." Ce mécanisme est basé sur l'usage de l'environnement comme moyen d'inscription des effets des comportements passés pour influencer les comportements futurs. Les fourmis dans une colonie utilisent le mécanisme de stigmergie pour communiquer indirectement entre elles en déposant la phéromone pour aboutir aux sources de nourriture. Plusieurs tentatives d'associer le web avec la notion de stigmergie ont été analysées récemment surtout la possibilité d'utiliser des stratégies stigmergiques pour créer des sites web auto-organisés. Les utilisateurs interagissent avec le site changeant sa structure et son contenu à l'aide d'agents logiciels. (Gregorio, 2003) a lié le concept de stigmergie au web en montrant que les weblogs, les communautés web et l'algorithme PageRank utilisé dans Google exhibent des comportements stigmergiques.

1.5. *Programmation des CAS par les agents*

Dans (Hassas, 2003), un cadre a été proposé pour le développement des CAS à base d'agents situés combinant la stigmergie aux propriétés de Holland :

- L'usage du paradigme des multi-agents situés et de l'intelligence comportementale. Ceci permet la définition des blocs de base et des modèles internes ainsi que le mécanisme d'étiquetage des agents (rôle des agents).

- La matérialisation physique de l'environnement et l'utilisation d'une représentation spatiale sous forme d'un réseau complexe. Ceci permet, d'une part, la matérialisation spatiale des agrégats émergents grâce à la situation et l'incarnation des agents et d'autre part, la matérialisation des flux d'informations et de ressources;

- L'usage de la stigmergie comme mécanisme de contrôle et de coordination distribuée. Ceci permet l'exploitation des flux et la mise en œuvre de la non linéarité des interactions intra-niveaux et inter-niveaux.

- Le maintien de l'équilibre entre l'exploration et l'exploitation dans le comportement des différents agents. L'exploitation permet de mettre en œuvre l'agrégation des blocs de base au niveau des comportements et l'exploration permet la mise en œuvre de la diversité.

2. Etude de cas : Une approche pour l'organisation dynamique du contenu du web

2.1. Le système WACO

Le modèle suivant illustre comment les principes des CAS combinés à la stigmergie sont appliqués dans le contexte d'organisation du contenu du web. WACO (WebAnts Content Organization) (Hassas, 2003) est une approche inspirée du comportement des insectes sociaux pour organiser dynamiquement le contenu du web. Dans cette approche, le web est considéré comme un environnement complexe, habité par des créatures artificielles appelées WebAnts. Ces créatures implémentées par des agents mobiles, sont organisées dans une colonie, et miment deux comportements observés dans les sociétés d'insectes, à savoir : le tri collectif et le fourragement pour la collecte de nourriture. Quatre types de WebAnts ont été créés, et à chaque type d'agent, a été alloué une tâche spécifique : les WebAnts *explorateurs* cherchent aléatoirement des documents du web. Les WebAnts *collecteurs* maintiennent et organisent sémantiquement les documents collectés. Les WebAnts *chercheurs* renforcent les clusters de documents collectés en explorant le web afin de trouver des documents semblables à ajouter au cluster. Finalement, les WebAnts *satisfaisant les requêtes des utilisateurs* cherchent le groupe approprié basé sur les requêtes des utilisateurs.

Les WebAnts dans WACO sont créées dynamiquement et s'adaptent à leur environnement et co-évoluent. Ce processus requiert un mécanisme pour la gestion et la régulation de leurs populations d'agents. Dans ce contexte, le mécanisme de distribution et de consommation d'énergie est utilisé (Menczer *et al.*, 2000). Les WebAnts sont sensibles à une notion d'ordre qui est obtenue par organisation sémantique du contenu du web. L'activité des agents a tendance à augmenter quand le désordre (sémantique) augmente sur le web. Le désordre sémantique est caractérisé par l'arrivée de nouveaux documents non encore indexés (classés dans un cluster). L'activité des WebAnts est régulée par un mécanisme de distribution d'énergie fourni par l'environnement et directement associé à la notion d'ordre global dans l'environnement. Le désordre dans l'environnement génère de l'énergie capturée par les agents qui les rend plus actifs. Basés sur leur fonction de fitness définie par l'ordre/désordre, deux mécanismes dirigent leur cycle de vie: duplication

(naissance) et disparition (mort). Ces deux mécanismes sont régis par le niveau d'énergie associé à chaque agent, qui est calculé par le taux de désordre dans le voisinage immédiat de l'agent et qui est régulé par l'activité des agents.

2.2. Codage des documents par l'usage d'une phéromone synthétique

Les documents contenus dans les sites web sont considérés comme des objets à classer suivant leurs contenus sémantiques. Des clusters sémantiques seront construits, où un cluster est un ensemble de documents sémantiquement similaires par rapport à une mesure de similarité prédéfinie. Les WebAnts communiquent par un mécanisme de stigmergie, utilisant une phéromone numérique multi-structurée (phéromone synthétique). La phéromone synthétique est codée par une structure spatiale ayant les champs suivants:

- l'étiquette (W_k) ou label : caractérise la nature de l'information codée par la phéromone, qui est la valeur sémantique d'un document (Brin *et al.*, 1998).

$$W_k = L_c \cdot H_c \cdot T_f \cdot IDF_k$$

où L_c (respect. H_c) est un vecteur de constantes >1 si le mot apparaît dans un lien (respect. dans un titre) ou $=1$ sinon. T_f est la fréquence du mot-clef dans le document k et IDF_k est l'inverse de la fréquence du document.

- L'intensité (τ_{ij}) : exprime la pertinence de l'information. Cette valeur est calculée à chaque site i , pour chaque sujet j , utilisant le nombre de documents ayant le même sujet. A chaque temps $(t+1)$ un nouveau document est ajouté:

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

où ρ_j représente le taux de persistance ($(1-\rho_j)$ le taux d'évaporation), $\Delta \tau_{ij}^k(t)$ l'intensité de phéromone émise par un document k , sur le site i pour un sujet j au temps t , et D_{ij} est l'ensemble des documents traitant le sujet j sur le site i .

- Le taux d'évaporation: exprime le taux de persistance de l'information dans l'environnement. Plus bas est sa valeur, plus longue est l'influence de diffusion de l'information dans l'environnement. C'est le rapport entre les documents traitant le même sujet et l'ensemble de tous les documents du site i :

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

où D_{ij} est l'ensemble des documents traitant le sujet j sur le site i , et D_i est l'ensemble de tous les documents sur le site i . L'idée est de faire en sorte que le regroupement des documents partageant le même sujet soit plus attractif que les documents isolés. Si un site contient des contenus sémantiques hétérogènes, cette information est considérée comme n'étant pas suffisamment pertinente. Donc la phéromone associée s'évaporerait plus rapidement, que celle émise par les contenus homogènes.

- Le taux de diffusion : exprime la distance à laquelle est propagée l'information dans l'environnement. Plus sa valeur est élevée, plus élevée est la portée de

l'information dans l'environnement. Nous exprimons cette distance en utilisant l'information de topologie du lien. En effet, en cherchant un sujet particulier sur le web, on explore des pages suivant des liens se référant au sujet d'intérêt. Nous associons à chaque site i , une distance d_{ij} pour chaque sujet j référencé par i , qui est calculée comme le chemin le plus long du site i au dernier site référençant le sujet j , suivant un parcours en profondeur du graphe.

$$d_{ij} = \text{Max}_k (d_{ij}^k)$$

k étant le nombre de liens référençant le sujet j , sur le site i . L'idée est de diffuser plus largement les sites qui sont un bon point d'entrée pour la recherche.

2.3. *Actions et interactions des agents*

Le web représente un environnement incertain et complexe que les WebAnts explorent et structurent en interagissant entre elles. Il représente aussi un médium pour les interactions entre agents via le dépôt/suivi de phéromone. Les blocs de base sont les populations d'agents, qui miment les comportements de tri collectif et de fourragement pour la collecte de nourriture. L'utilisation de la phéromone digitale correspond au mécanisme d'étiquetage chez Holland et l'agrégation est observée à travers la structure multi-phéromonale.

Les WebAnts Explorateurs perçoivent les différents types de phéromone correspondant aux différentes valeurs sémantiques, mais sont plus sensibles à la valeur de phéromone du dernier document collecté. Les WebAnts Collecteurs calculent la synthèse de la phéromone du site et mettent à jour les valeurs de ses paramètres associés (étiquette, intensité, taux de persistance, taux de diffusion). Les Web Ants Chercheurs sont lancés quand un cluster atteint une valeur seuil et les WebAnts Satisfaisant les Requêtes codent les requêtes des utilisateurs en valeur de phéromone et cherchent dans l'environnement le cluster approprié. La combinaison des comportements de tri collectif et de fourragement permet la structuration permanente du web en des clusters représentant la concentration des documents ayant le même sujet sémantique, et la propagation de cette information sur le web facilite la recherche de l'information pertinente.

2.4. *Résultats*

Des simulations ont été conduites pour tester l'efficacité de cette approche dans (Hassas, 2003). Quelques résultats sont présentés ci-dessous :

Expérience 1: Maintenance de la croissance de l'ordre dans le système

Nous avons étudié la notion d'ordre sur le web en associant cette notion avec l'émergence de clusters ayant des contenus sémantiques similaires. Pour un site donné, on a exprimé la fonction d'ordre local par le nombre de sites dans son

voisinage, avec un contenu similaire. La similarité est calculée par rapport à une distance spécifiée entre les vecteurs des poids des mots clés associés. Dans la figure 1 ci-dessous, l'axe des abscisses représente le nombre d'itérations (l'échelle temps) et l'axe des ordonnées représente le nombre de documents. On remarque sur cette figure, que le désordre décroît régulièrement dans le système, tandis que l'apparition de nouveaux documents croît. Le désordre est mesuré par le nombre total de documents moins le nombre de documents regroupés. Cette figure montre l'efficacité du comportement de regroupement (clustering) des documents.

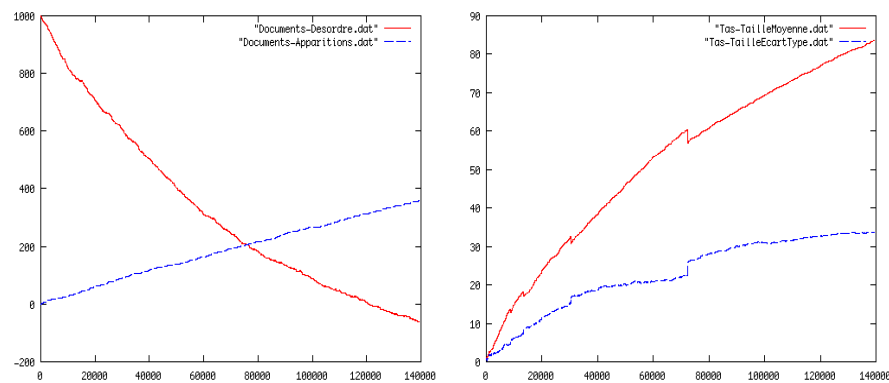


Figure 1. Le niveau de désordre décroît alors que de nouveaux documents sont créés et que l'opération de clustering est en cours.

Figure 2. Evolution de la taille moyenne des clusters et de son écart type

Expérience 2 : Formation des clusters et évolution de leur taille

La figure 2 représente l'évolution au cours du temps (l'abscisse) de la taille moyenne des clusters (l'ordonnée). On observe l'évolution des clusters formés et l'augmentation de leurs tailles. En comparant les figures 1 et 2, on remarque que le comportement de clustering renforce la création des clusters, alors que le désordre décroît dans le système.

Expérience 3: Distribution de l'énergie et son évolution dans le système

La figure 3 ci-dessous montre la valeur moyenne d'énergie des agents spécialisés (WebAnts Chercheurs). Ces agents augmentent leur quantité d'énergie pendant la formation des clusters. Quand les clusters sont formés (temps 80000), ils disparaissent (la valeur d'énergie = 0). On observe au temps 100000, une croissance soudaine d'énergie associée à l'apparition de nouveaux clusters, comme de nouveaux sites sont créés ou de nouveaux documents sont découverts.

Expérience 4 : Evolution de la population d'agents et régulation de leurs activités

La figure 4, où l'ordonnée représente la taille de la population et l'abscisse représente le temps, montre l'évolution des populations d'agents dans le système et la proportion d'agents actifs par rapport à la population entière. Jusqu'au temps 80000, qui est le temps d'émergence d'un ordre global, tous les agents sont actifs, et après ce temps, le nombre d'agents actifs décroît. Les agents inactifs disparaissent, réduisant le nombre d'agents initiaux. Tous les agents sont actifs durant la formation de nouveaux clusters, comme de nouveaux sites sont créés (temps 100000). Après ce temps, on observe un phénomène similaire au temps 80000.

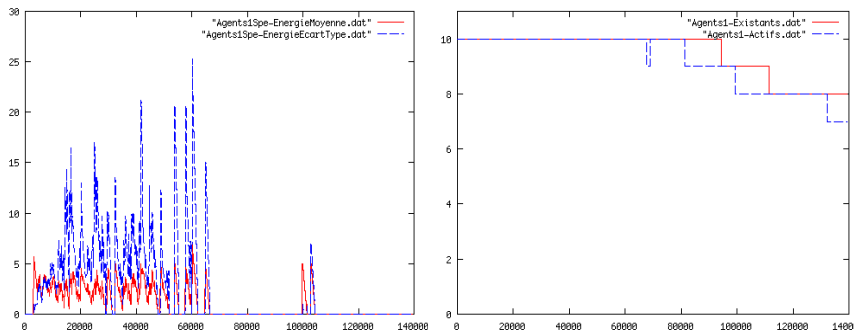


Figure 3. Evolution de la valeur moyenne de l'énergie des agents spécialistes (WebAnts Searchers) et de leur écart type

Figure 4. Evolution de la population: Proportion d'agents actifs/ à la population totale d'agents

3. Travaux Relatifs

Les moteurs de recherche comme Google, Altavista, Excite offrent aux utilisateurs des services de haute performance, efficacité et diversité. Cependant, les techniques utilisées par les moteurs de recherche courants souffrent d'un manque de pertinence pour les processus de recherche basés sur le contenu, et d'une inadéquation à l'évolution dynamique et rapide du contenu du web. Cette insuffisance est due au fait que les moteurs de recherche créent des indexes des documents web, qui peuvent être vus comme des « snapshots » du web. Ces indexes ont besoin d'être mis à jour aussi fréquemment que les changements se produisent sur le web. Certains systèmes utilisent des « crawlers » qui visitent et revisitent exhaustivement chaque page web pour maintenir son contenu, mais comme cette maintenance est réalisée de façon périodique, à un moment donné un index serait inapproprié ou incomplet. Dans notre approche, le système suit l'évolution permanente du contenu du web et met à jour ses changements aussi fréquemment

qu'ils se produisent. Différentes méthodes de Recherche de l'Information sur le web à base d'agents ont été proposées (Eliassi-Rad *et al.*, 2003, Klusch, 2001, Kobayashi *et al.*, 2000, Müller, 1999).

Un autre point à considérer, c'est à quel degré, les systèmes de recherche d'information courants et les moteurs de recherche satisfont les requêtes des utilisateurs. On assiste à l'émergence d'une myriade de systèmes offrant des adaptations et personnalisation des sites web (Eirinaki *et al.*, 2004, Menczer, 2003, Pazzani *et al.*, 2002, Pierrakos *et al.*, 2003). Comme le contenu du web et les intérêts des utilisateurs évoluent en permanence, les moteurs de recherche personnalisés doivent s'appuyer sur une étude à long terme des requêtes émises en s'adaptant à l'évolution des profils des utilisateurs et contribuant à sa diversification en découvrant de nouveaux champs d'intérêts suivant l'évolution du web. Dans notre approche, nous répondons à cette problématique en considérant différents types de populations d'agents interactives, dont chacune possède un rôle spécifique : gestion des requêtes des utilisateurs, enrichissement de leurs profils, clustering dynamique et sémantique du contenu du web et affinement et enrichissement du contenu.

4. Conclusion

Dans ce papier nous avons montré que le web exhibe des comportements similaires à un système complexe adaptatif et présenté une approche pour l'organisation dynamique du contenu du web selon la perspective des CAS combinée à la stigmergie. Nous développons couramment un modèle selon la perspective des CAS, basé sur l'association d'une sémantique à l'information contenue sur le web, par la combinaison du contenu, de l'usage et de la structure.

5. Bibliographie

- Barabasi A. L., Bonabeau E., "Scale-free networks", *Scientific American*, vol. 288, n° 5, 2003, p. 60-69.
- Brewington B., Cybenko G., "How dynamic is the web?" *9th International World Wide Web Conference (WWW9)*, Geneva, 2000.
- Brin S., Page L., "The Anatomy of a Large-Scale hypertextual Search Engine", *Computer Networks and ISDN Systems*, 1998.
- Cho J., Garcia-Molina H., "The Evolution of the Web and Implications for an Incremental Crawler." *the Twenty-sixth International Conference on Very Large Databases*, 2000.
- Eirinaki M., Lampos C., Paulakis S., Vazirgiannis M., "Web personalization integrating content semantics and navigational patterns", ACM Press, 2004.
- Eliassi-Rad T., Shavlik J., "Intelligent Web Agents that Learn to Retrieve and Extract Information", Springer-Verlag, 2003.

- Flake G. W., Lawrence S., Giles C. L., Coetzee F. M., "Self-Organization of the Web and Identification of Communities", *IEEE Computer*, vol. 35, 2002, p. 66--71.
- Grassé P. P., "La reconstruction du nid et les coordinations inter-individuelles chez *bellicositermes natalensis* et *cubitermes*, la théorie de la stigmergie: Essai d'interprétation des termites constructeurs", *Insectes Sociaux*, vol. 6, 1959, p. 41-81.
- Gregorio J., "Stigmergy and the World-Wide Web", <http://bitworking.org/news/Stigmergy>, 2002.
- Hassas S., "Systèmes complexes à base de multi-agents situés", University Claude Bernard Lyon, 2003.
- Hassas S., "Using swarm intelligence for dynamic web content organization", *IEEE Swarm Intelligence Symposium*, Indianapolis, IN, USA, 2003, IEEE Press, p. 19-25.
- Holland J., *Hidden Order*, Hidden Order, Addison-Wesley, 1995.
- Kauffman S., *The Origins of Order: Self-Organization and Selection in Evolution*, The Origins of Order: Self-Organization and Selection in Evolution, Oxford University Press, 1993.
- Klusch M., "Information agent technology for the Internet: a survey", Elsevier Science Publishers B. V., 2001.
- Kobayashi M., Takeda K., "Information retrieval on the web", *ACM Comput. Surv.*, vol. 32, n° 2, 2000, p. 144--173.
- Liu J., Zhang S., Ye Y., "Understanding emergent web regularities with information foraging agents", *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, 2002, ACM Press, p. 457--458.
- Menczer F., "Complementing search engines with online web mining agents", *Decision Support Systems*, vol. 35, n° 2, 2003, p. 203-242..
- Menczer F., Belew R., "Adaptive retrieval agents: Internalizing local context and scaling up to the web", *Machine Learning*, vol. 9, 2000, p. 203-242.
- Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D., Alon U., "Network motifs: simple building blocks of complex networks." *Science*, vol. 298, 2002, p. 824-827.
- Müller M. E., "An Intelligent Multi-Agent Architecture for Information Retrieval from the Internet", Technical report, U. of Osnabrück, Germany, 1999.
- Pazzani M. J., Billsus D., "Adaptive Web Site Agents", *Autonomous Agents and Multi-Agent Systems*, vol. 5, n° 2, 2002, p. 205--218.
- Perkowitz M., Etzioni O., "Towards adaptive Web sites: Conceptual framework and case study", *Artificial Intelligence*, vol. 118, n° 1-2, 2000, p. 245-275.
- Pierrakos D., Paliouras G., Papatheodorou C., Spyropoulos C. D., "Web usage mining as a tool for personalization: A survey", *User Modeling and User-Adapted Interaction*, vol. 13, n° 4, 2003, p. 311-372.