

## Preference Dissemination by Sharing Viewpoints : Simulating Serendipity

Guillaume Surroca, Philippe Lemoisson, Clement Jonquet, Stefano A. Cerri

► **To cite this version:**

Guillaume Surroca, Philippe Lemoisson, Clement Jonquet, Stefano A. Cerri. Preference Dissemination by Sharing Viewpoints : Simulating Serendipity. KEOD: Knowledge Engineering and Ontology Development, Nov 2015, Lisbon, Portugal. IC3K 2015, 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 7th International Conference on Knowledge Engineering and Ontology Development (2), pp.402-409, 2015, Volume 2 : KEOD. <lirmm-01223078>

**HAL Id: lirmm-01223078**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01223078>**

Submitted on 2 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Preference Dissemination by Sharing Viewpoints : Simulating Serendipity

Guillaume Surroca<sup>1</sup>, Philippe Lemoisson<sup>2</sup>, Clément Jonquet<sup>1</sup> and Stefano Cerri<sup>1</sup>

<sup>1</sup>*Laboratory of Informatics Robotics Micro-Electronics of Montpellier, France*

<sup>2</sup>*UMR Territories, Environment, Remote Sensing and Spatial Information, CIRAD, Montpellier, France*

*{surroca, jonquet, cerri}@lirmm.fr, philippe.lemoisson@cirad.fr*

**Keywords:** Knowledge Representation, Knowledge Discovery and Dissemination, Serendipity, User-Centered Knowledge Engineering, Collective Intelligence, Web 2.0, Agents.

**Abstract:** The Web currently stores two types of content. These contents include linked data from the semantic Web and user contributions from the social Web. Our aim is to represent simplified aspects of these contents within a unified topological model and to harvest the benefits of integrating both content types in order to prompt collective learning and knowledge discovery. In particular, we wish to capture the phenomenon of Serendipity (i.e., incidental learning) using a subjective knowledge representation formalism, in which several “viewpoints” are individually interpretable from a knowledge graph. We prove our own Viewpoints approach by evidencing the collective learning capacity enabled by our approach. To that effect, we build a simulation that disseminates knowledge with linked data and user contributions, similar to the way the Web is formed. Using a behavioral model configured to represent various Web navigation strategies, we seek to optimize the distribution of preference systems. Our results outline the most appropriate strategies for incidental learning, bringing us closer to understanding and modeling the processes involved in Serendipity. An implementation of the Viewpoints formalism kernel is available. The underlying Viewpoints model allows us to abstract and generalize our current proof of concept for the indexing of any type of data set.

## 1 INTRODUCTION

Since Web 2.0 has democratized the sharing, recommendation and creation of content via social networks, blogs and fora, and since semantic Web technologies have begun to structure the knowledge deposited, generated and stored on the Web, two kinds of content have emerged. These types of content differ in the ways they are produced and structured. On one hand, contribution-based social Web platforms allow the production of a wealth of data with little or no structure; these data evolve rapidly (e.g., folksonomies (Mika, 2007)). On the other hand, highly structured knowledge is constituted consensually by circles of experts (e.g., ontologies (Karapiperis & Apostolou, 2006) or linked data (Bizer, Health, & Berners-Lee, 2009)). With the Viewpoints approach, our objective is to create a knowledge representation formalism that retains the best qualities of each type of content. Our objective is to support and give value to both (i) the structure which characterizes semantic Web datasets and (ii) the evolution and maintenance rates of

shared knowledge on the social Web as proposed in Gruber (Gruber, 2008) or (Freddo & Tacla, 2009). We aim to contribute to knowledge representation approaches by designing a system involving Web agents (human or artificial) who share “viewpoints” linking system resources (identified by a URI). We ask ourselves the following questions:

- Which Web browsing strategies allow the most optimal diffusion of user preference systems?
- What should the conditions be to favor incidental learning, a.k.a., Serendipity, in the study of preference systems?

We define the preference system of an agent by the expression of his tastes and attractions in terms of proximity or distance relationships between Web resources. In a previous contribution (Lemoisson, Surroca, & Cerri, 2013), we demonstrated the learning ability of a knowledge base built with an initial version of our formalism. However, this proof of concept was based on a poor behavioral model of agents who navigated and contributed to the knowledge base; we were only interested in the

agents' satisfaction and did not take into account their preference systems. In another contribution, we showed how Viewpoints allow the search and discovery of knowledge through a search engine prototype for scientific publications (Surroca, Lemoisson, Jonquet, & Cerri, 2014). In the newest model, we include a "Serendipity acceptance" factor in the behavior of agents, defined as the tendency of an agent to turn to resources outside of his preference system. This allows us to assess the diffusion of preference systems, depending on whether an agent is open-minded or focused on what he knows and prefers. Using this model, we build a simulation based on individual behavior rules (microscopic level) in order to observe the effect on collective learning and on the diffusion of preference systems (macroscopic level). This simulation illustrates the advantages of using Viewpoints to "merge" the essence of data semantics and the social Web.

The rest of this article is organized as follows: section 2 presents the background and inspiration for our approach by introducing the notion of Serendipity in computer systems. In our review of the state of the art, we also briefly compare Viewpoints to several other knowledge representation approaches. Then, we briefly present the Viewpoints formalism in Section 3. Section 4 explains our behavioral model of Web users and our representation of their preference systems: we show how we simulate the evolution of the Web as a knowledge graph and discuss a set of hypotheses on the impact of individual browsing strategies. Section 5 presents a simulation in which three agents (the Princes of Serendip) contribute to building a 'toy' knowledge graph with resources of different shapes, sizes and colors; then we discuss our current results relative to our assumptions and our research objectives. Section 6 concludes and presents potential perspectives for this work.

## 2 STATE OF THE ART

### 2.1 Knowledge Representation

Several studies have focused on the merging of the Semantic Web and the Social Web (Gruber, 2008). We synthetically compare our approach to these studies as follows: in addition to incorporating the (human or artificial) Agent as presented in (Mika, 2007), our representation of knowledge considers it a central constituent. We explain how in the formalism section. Moreover, our knowledge representation considers Viewpoints micro-

expressions of individual semantics. However, our mechanism for evaluating and confronting Viewpoints does not use any additional contribution as is the case in (Limpens & Gandon, 2011). Thus, the emphasis is placed on what emerges from the knowledge graph, as reported in (Aberer et al., 2004; Noh, Park, Park, & Lee, 2010); indeed, these authors studied the possibility of the emergence of a collective representation of knowledge with a "bottom-up" vision of system interactions.

Finally, we define a metric distance over the set of resources formed by the knowledge providers (Agents), supports (Documents) and descriptors (Topics) while semantic distances found in the literature apply to homogeneous subclasses such as distances between tags or ontology concepts (Lee, Shah, Sundlass, & Musen, 2008). The resulting Viewpoints Knowledge Graph (KG) is constituted by resources connected by viewpoints, and can be seen as a wide, evolving, associative memory enabling collective intelligence, metaphorically replicating a brain, where all learning processes are supported by the evolving strength of synapses (Edelman, 1987). Instead, we adopt a topological approach and compute semantic distances on top of the Viewpoints in a manner similar to (Pedersen, Pakhomov, Patwardhan, & Chute, 2007).

### 2.2 Serendipity, the Incidental Learning

The term 'Serendipity' is derived from an ancient Persian tale entitled 'The Three Princes of Serendip' (Merton & Barber, 2006). Recently, Perriault said that "the Serendipity effect (...) consists in nimbly and randomly happening upon something we did not search for ". We are then led to make abductive inferences in order to build a theoretical framework which encompasses, via appropriate aggregation, information which used to be disparate (Perriault, 2000). We note that the notion of luck or chance is important in the Serendipity phenomenon. However, "it does not only depend on a divine dice roll" as explained in (Fine & Deegan, 1996) and takes place only at the border of what is already known. Thus, incidental learning is greatly facilitated when new knowledge is in the vicinity of existing knowledge and may be interpreted by someone who knows this neighborhood. We share the vision that knowledge does not guarantee serendipitous discovery, but that it makes it more likely. We therefore introduce the notion of Serendipity proximal zone, which is similar to the concept of proximal development zone

(Vygotsky, 1978) in learning and education sciences. We will show below how the Serendipity acceptance factor helps us to capture Serendipity in our model.

When considering the huge amount of information available on the Web and the ways in which one may get lost while browsing, Serendipity seems to be a realistic phenomenon. One may talk about serendipitous Web-based learning, as explained hereafter. The search for knowledge through serendipitous learning can succeed by chance or as an offside activity of a main task (Bowles, 2004). For instance, a user who makes an initial query may be progressively led into an unexpected path that ultimately proves more productive than the initial search. In such cases, Bowles writes that serendipitous learning occurs (Bowles, 2004). This is exactly the phenomenon we model and observe in our section 4 with multiple navigation strategies. In addition, according to Allen Tough, almost 80% of learning is informal and unplanned (Tough, 1999). Serendipitous navigation is an "intellectual lottery (...) with small chances but with big potential payoff" (Marchionini, 1997). In the latter work, the parallel with our Viewpoints approach is made explicit: "We also gain new viewpoints and associations for our problem by browsing alternative sources using different tools, techniques and data structures."

Recommender systems (Adomavicius & Tuzhilin, 2005) are increasingly interested in Serendipity, because the variety of recommendations is as important as their accuracy. Serendipity goes beyond what recommendation systems offer, thanks to the surprise, variety and novelty of the proposed results. Additionally, many recommender systems have begun to implement Serendipity principles. The folksonomy-based recommendation in (Yamaba, Tanoue, & Takatsuka, 2013) allows users to tag books and go beyond the traditional classification, and therefore add new books to the Serendipity proximal zone of other users. However, to our knowledge, except from work proposed in (Corneli, Pease, & Colton, 2014) on the theoretical framework for the phenomenon of Serendipity, the literature on the formalization and the measurement of this phenomenon is lacking. Based on our review, there is currently no exploitable model of Serendipity.

### 3 VIEWPOINTS FORMALISM

Viewpoints is a formalism for subjective knowledge; it holds that any proximity or distance relationship

between two resources is expressed by an agent as a viewpoint. A typed viewpoint connects these two resources. These viewpoints are individually interpreted by a perspective chosen by the user / contributor. This perspective allows assigning a weight to each viewpoint, depending on who issued it, on when it was created, and on its semantic type or other more complex criteria. Therefore, Viewpoints is a knowledge representation formalism centered on equally considered human (e.g., Web users) or artificial (e.g., data mining tools, knowledge extractors, ontologies) agents. Resources (providers, descriptors and knowledge supports) are bound by the viewpoints on the knowledge graph. The KG is a bipartite graph consisting of a set of resources  $R$  and a set of viewpoints  $V$  connecting these resources. The resources in  $R$  are either agents (knowledge providers, i.e., viewpoint creators), knowledge descriptors (topics, tags) or knowledge supports (documents, videos, Web pages, messages, posts, etc.). A viewpoint is a tuple  $(a \rightarrow \{r1, r2\}, \theta, t)$  containing the following information:

- $a$ , the agent who issued the viewpoint;
- $\{r1, r2\}$ , the couple of resources semantically connected by  $a$ ;
- $\theta$ , the viewpoint's type, used to interpret (i.e., assign a weight to) it;
- $t$ , the viewpoint's creation date.

For instance,  $(Guillaume \rightarrow \{Diffusion\ systems\ [...] \ views, \ acm:Knowledge\ representation\ and\ reasoning\}, \ dc:subject, \ 27/02/15)$  means that the agent Guillaume associates this article to the Knowledge representation and the reasoning concept of ACM's taxonomy with the relation DublinCore subject.  $(Mario \rightarrow \{Mario, \ Luigi\}, \ foaf:knows, \ 1985)$  means that Mario elicited that he has known (as in FOAF) Luigi since 1985. To identify the meaning of the data represented in the form of Viewpoints, we adopt, when possible, existing Semantic Web types.

### 4 VIEWPOINTS EXPLOITATION

The aggregation of all connections between two resources created by the different agents form a semantic proximity link named synapse. The strength of the synapse is based on the aggregation of the weights of each viewpoint in the synapse. The two functions of evaluation (Map) and aggregation (Reduce) of viewpoints form a perspective which allows the exploitation of subjective knowledge. For the same KG, several interpretations, defined as Knowledge Maps (KM), can be made dependent on how agents evaluate and aggregate viewpoints. The

Knowledge Map is a graph made of resources (R) and synapses (S) to which common graph algorithms can be easily applied. The perspective is unique to each user, who decides to interpret the KG any way he wants. The two functions of evaluation and aggregation of viewpoints can be extended at will to suitably match one's needs. Figure 1 illustrates the interpretation process of KG. In the following simulation we use: (i) a direct neighborhood function that returns all the resources directly connected by viewpoints to a specified resource, and the weight of the synapses binding this resource to its direct neighbors; (ii) an indirect neighborhood function based on the Dijkstra algorithm and which, for a resource  $r_i$ , returns all resources  $r_j$  on all the paths starting from  $r_i$  with a length less than a specified threshold,  $m$  (simulation-specific parameter).

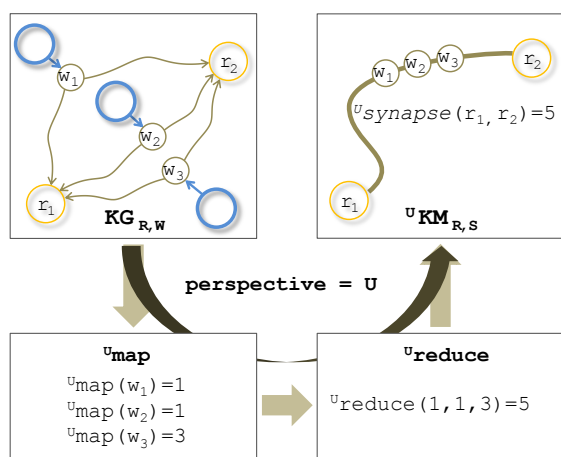


Figure 1: Interpretation of Knowledge Graph (KG) into Knowledge map (KM).

An important aspect, directly inspired from the Web 2.0, lies in the built-in feature for integrating agent feedback. Within their perspective, agents use any Viewpoint for browsing KM and reversely update the KG through viewpoints expressing their feedback. Along these exploitation/feedback cycles, shared knowledge is continuously elicited against the beliefs of the agents in a selection process. The knowledge map is defined as a graph in which semantic similarities within the knowledge resources are computed according to a given perspective.

## 5 SERENDIP SIMULATION

Our goal is to simulate the evolution of a knowledge base – such as the Web – from individual behavior rules that describe agents browsing the Web and

disseminating their preference systems. First, we explain how we represent the preference systems in a Viewpoints KG; then, we propose a behavioral model simulating different configurable navigation strategies. This model is based on calculations of direct and indirect neighborhoods. Finally, we observe the effect of this set of individual rules on the macroscopic level of knowledge represented in the resulting KG.

### 5.1 Preference Systems Representation

In this simulation, each resource is characterized by a shape, a size and a color. Shape and size information will already be included in the KG at the beginning of the simulation; this information simulates the Semantic Web data. Color information is introduced step-by-step during the simulation by three agents, the princes of Serendip, who know and like a different color each (red, green, blue); this information simulates social Web contributions. The preference system of a prince is the set of all the viewpoints he has issued to make same-color resources get closer to him or closer to one another. We consider two kinds of viewpoints: (i) the first kind links two same-color resources ( $vps:knows$ ); (ii) the second associates a prince of a specific color with a resource of the same color ( $vps:likes$ ). The dissemination of a preference system is therefore equivalent to the distribution of the color information in the graph, i.e., the more colored the graph becomes, the more a preference system has been shared. Thus, when the graph “learns” a color, it illustrates the collective intelligence of the community.

For example, if the red prince searches a red resource  $r$  and retrieves a red-color resource  $r'$ , he issues the two following viewpoints ( $RedPrince \rightarrow \{Redprince, r\}, vps:likes, \tau$ ) and ( $RedPrince \rightarrow \{r, r'\}, vps:knows, \tau$ ). In the next section we will present the different navigation strategies which allow princes to disseminate knowledge about their color.

### 5.2 Behavioral Model of the Serendip Princes

The state automaton in Figure 2 describes the behavior of the princes when they are navigating in the KG and disseminating their preference systems (viewpoints emission). More generally, this automaton simulates the behavior of a user when he

is exploring the contents of a knowledge base such as the Web.

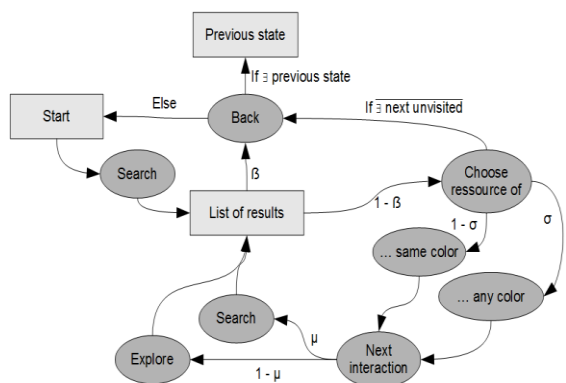


Figure 2: Behavioral automaton of the Princes of Serendip

We capture behaviors such as: querying a search engine, exploring the results, following links included in these results and querying the search engine again. In our simulation the behavior of a Prince corresponds to a specific configuration of the  $\beta$ ,  $\mu$  and  $\sigma$  parameters; we call this a *navigation strategy*. Our simulation is divided into cycles that correspond to successive explorations of the KG. At the beginning of a cycle, a prince begins interacting with the KG; we simulate the use of a search engine: A resource of the KG is randomly selected and the indirect neighborhood function is used to retrieve a list of results (other resources) sorted by semantic proximity. From the proposed results, the prince continues (low  $\beta$ ) or abandons this search and undertakes a new one (high  $\beta$ ). If he continues, he must evaluate these results one by one (comparing them to the color corresponding to his preference system) and select the first non-visited result based on the  $\sigma$  parameter. If the prince accepts Serendipity (high  $\sigma$ ), he does not systematically select resources of his own color; if he does not accept Serendipity (low  $\sigma$ ), he will instead focus on resources of his color only. Once a resource is chosen, the prince moves to the next stage of his journey: Depending on  $\mu$ , he will either perform a direct search on this resource (high  $\mu$ ) or explore locally around this resource (low  $\mu$ ). The first interaction simulates the act of opening a Web page as the result of a previous search; the second interaction simulates either a new search, e.g., with the title or content of the current page, or clicking on a Web link within a page. In the simulation, princes start with an initial budget of interactions; this budget is decreased with each

interaction (research and exploration). It represents the amount of effort princes are willing to make when navigating. When princes wish to go backwards, three scenarios will lead to the end of the cycle: There are no previous steps; or, all resources have been visited; or, the initial interaction budget has been spent.

These strategies simulate Web browsing. In terms of graph traversal, a high  $\beta$  corresponds to a breadth-first approach, whereas lower  $\beta$  corresponds to a depth-first approach. In an information search process, the breadth-first approach would superficially assess all the results; instead, the depth-first approach would rather focus on what would seem to be the best result and dig deeper.  $\mu$  determines the navigation style. A high  $\mu$  value means princes mainly use SEARCH engines that sort results according to a global approach; a low  $\mu$  means princes will carry out a step-by-step exploration by collecting unsorted local results (EXPLORATION). For example, navigating from one suggested YouTube video to another is a good illustration of a step-by-step exploration, in which as a succession of Google searches illustrate a BREADTH traversal. We represent the Serendipity acceptance factor ( $\sigma$ ) as a third dimension. High  $\sigma$  means princes are mainly OPEN and are willing to visit both the resources that match their preferences and the resources that do not but could lead to chance discoveries. Low  $\sigma$  means princes are mainly CLOSED to the latter prospect and are entirely guided by their preferences when browsing.

## 6 SIMULATION DYNAMICS

### 6.1 Initial Conditions

A fixed-size KG is generated. In addition to their specific color (red, green, blue), the resources of the KG are characterized by their size (small, medium, large) and their shape (square, circle, triangle). For each possible size, shape and color combination,  $N$  resources are created. Therefore, there are initially  $27N$  resources. Two artificial agents, called peons are added to the KG. One of them shares his appreciation of shapes in the knowledge graph, connecting all the same shapes of resource pairs by viewpoint types  $vps:initial$ . The other peon does the same for size. Thus, after the peons have shared their appreciations, the KG does not “know” colors because resources are only tied by size and shape

characteristics. Finally, the 3 princes are added to the KG. Each of them is characterized by a unique color, and has the ability to appreciate colors and share this assessment by issuing new viewpoints such as `vps:like` and `vps:knows` in the KG. Thus, there is an implicit understanding that the princes are only able to share by issuing viewpoints as feedback.

## 6.2 Dissemination of Preference Systems

The simulation parameters are summarized in Table 1. The princes follow the behavioral model previously described and disseminate their preferences (knowledge of their own color) by issuing `vps:like` and `vps:knows` viewpoints. The weight assigned to each type of viewpoint is shown in Table 1. The aggregation capability of viewpoints for the calculation of the value is the sum of the synapses. At the end of each cycle, the following measures are calculated to evaluate the dissemination of color (preference) knowledge in the KG:

- M1 Color X: This is the ratio of the average distance between any resources over the average distance between X-colored resources.
- M2 Color X: This is the probability of getting a resource of the same color in the neighborhood of X-colored resource.

Table 1: Simulation parameters.

Categories	Parameters	Values (if fixed)	
Scale parameters	Scale factor (N)	3	
	Number of cycles	100	
	Number of iterations per cycle	50	
Perspective parameters	Weight of viewpoints with type <code>vps:initial</code>	1	
	... type <code>vps:knows</code>	2	
	... type <code>vps:like</code>	1	
Navigation strategy parameters	$\beta$		
	$\mu$		
	$\sigma$		
Activity distribution	Red prince	33%	80%
	Green prince	33%	10%

Given the large number of parameters (Table 1), we present the results (curves) of several simulations with the parameter configurations which we consider the most significant for navigation strategies. However, we explain the effects of

specific parameters in the discussion section. Other fixed parameter values are given in Table 1.

## 6.3 Hypotheses

Princes progressively share their color assessments with other users through the feedback mechanism. We aim to observe how the KG "learns" (at the global level) the notion of color that was not originally in the knowledge represented by the `vps:initial` viewpoints. Thanks to viewpoints, each individual preference system becomes part of the collective knowledge represented in the KG, where it coexists with the preference systems of other princes. Our goal is to experiment with different navigation strategies and demonstrate that preference systems do not neutralize each other when concurrently broadcast. We also want to measure the effect of Serendipity. Thus, we expect M1 to increase; in other words, the average distance between same-color resources will decrease more quickly than the average distance between any resource. M2 should also increase as it reflects the probability of finding the same-color resource in the m-neighborhood of a resource.

## 7 RESULTS AND DISCUSSIONS

### 7.1 Impact of the Serendipity Acceptance Factor

We start by assessing the impact of  $\sigma$  on the dissemination of the color red thanks to measures M1 and M2 Red. One can notice (Figure 3) that when search engines are mainly used, M1 and M2 increase at a faster rate when Serendipity acceptance is low; conversely, when Serendipity acceptance is high, they reach higher final values. Therefore, Serendipity acceptance allows a wider dissemination of color knowledge. Indeed, while the search indirectly returns results and allows the creation of viewpoints that have not already been issued, Serendipity acceptance increases the potential for creating new original viewpoints.

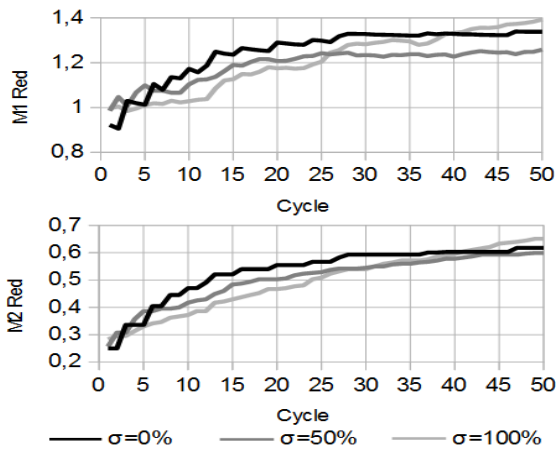


Figure 3: M1 Red and M2 Red evolution when princes mainly use search engines ( $\mu=70\%$ ,  $\beta=10\%$ )

These new associations are expressions of preference systems that would likely not have been generated if the princes had been guided only by their preferences to navigate. In contrast, we observe (Figure 4) that when mainly local exploration results are used to navigate from, Serendipity acceptance does not affect either M1 and M2 value increases or final values. This strategy's idea is to explore local and in-depth results; moreover, going through less interesting results along the road tends to slow the spread of preference systems. The  $\mu$  effect (navigation device) is very important for Serendipity. However, we realize that the relative homogeneity of our graph does not realistically represent the Web's structure. We believe that, under more realistic conditions, Serendipity can produce more substantial gains than it does in our "toy" knowledge graph. In this simulation, the three princes are active (33%) and  $\beta = 10\%$ .

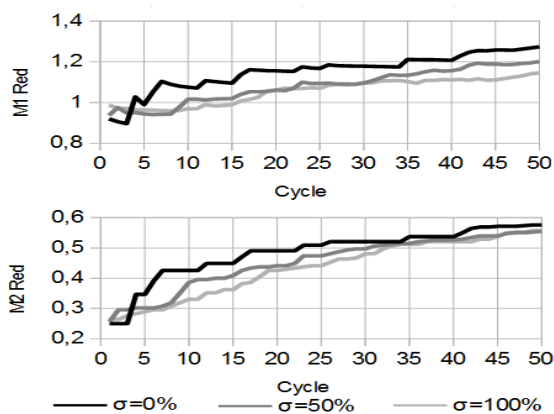


Figure 4: M1 Red and M2 Red evolution when princes mainly use link-by-link exploration ( $\mu=30\%$ ,  $\beta=10\%$ ).

## 7.2 Adaptation to Real Web Data

We also conducted a similar experiment with real data on movies and user ratings. We studied a Web dataset (MovieLens), in which explicit semantics were mixed with social contributions. This dataset consisted in two sets of 100,000 and 1,000,000 ratings which had been collected by the GroupLens Research Project at the University of Minnesota. In our MovieLens experiment, users elicited preferences when they associated movies with ratings. Initially, each movie was linked to other movies by metadata such as actors, directors or genres. For instance, the genre characteristic corresponded to the shape characteristic in our Serendip simulation. All films, as well as other resources such as genres, were initially added to the KG. During each cycle, a portion of the ratings was added to the KG as viewpoints, once again simulating the contributions of the social Web. We observed knowledge crystallizing progressively around the reviewers. This experiment showed us that when working with such a recommendation system, we may observe that structured data (genres, actors, director) do bootstrap the creation of subjective (social) knowledge. Integrating user data such as gender, age group, job and movie metadata (genre, release year) showed us new relations. User was closing movies and movies were semantically reproaching users. One of the goal we gave to us with ViewpointS was also to observe dynamics in an evolving represented knowledge.

## 8 CONCLUSIONS AND PERSPECTIVES

After presenting and positioning our approach of subjective knowledge representation, we studied the phenomenon of Serendipity and its current influence on the Web. With the Princes of Serendip simulation, we presented an experiment for modeling Serendipity on the Web. We recognize that this behavioral model of Web users may not fully represent the reality and diversity of Web exploration methods. Nonetheless, we hope that we have demonstrated the ability of the Viewpoints knowledge graph to learn. Our simulation results allowed us to assess the contribution of the Serendipity acceptance factor to various navigation strategies and its impact on the dissemination of preference systems; we consolidated the Viewpoints proof of concept by confronting it with a more



realistic use of modeling and simulation. We are planning for several applications which may help us evaluate the Viewpoints approach: Amongst them, (i) one will consist in cross scientific discovery of agronomic knowledge (CIRAD) and (ii) another will deal with biomedical data within the SIFR project (<http://www.lirmm.fr/sifr>). We are finishing also several IR benchmarks (recall, precision and f-measure) on a film recommendation scenario comparing our semantic neighborhood methods to classic indexation and research methods such as Vector Space Model. We will soon publish benchmarks results in one the scenarios we previously mentioned.

## ACKNOWLEDGEMENTS

This work was supported in part by the French National Research Agency under JCJC program, grant ANR-12-JS02-01001, as well as by University of Montpellier, CNRS and the CIRAD.

## REFERENCES

- Aberer, K., Cudr, P., Catarci, T., Hacid, M., Illarramendi, A., Mecella, M., ... Scannapieco, M. (2004). Emergent Semantics Principles and Issues. In D. Lee, YoonJoon and Li, Jianzhong and Whang, Kyu-Young and Lee (Ed.), *Database Systems for Advanced Applications* (Vol. 2, pp. 25–38). Springer Berlin Heidelberg.
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.
- Bizer, C., Health, T., & Berners-Lee, T. (2009). *Linked Data - The Story So Far*. In *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (Vol. 5, pp. 1–22).
- Bowles, M. (2004). Relearning to E-learn: Strategies for Electronic Learning and Knowledge. *Educational Technology & Society*, 7(4), 212–220.
- Corneli, J., Pease, A., & Colton, S. (2014). Modelling serendipity in a computational context. *arXiv Preprint arXiv:1411.0440*.
- Edelman, G. (1987). *Neural Darwinism: The theory of neuronal group selection*.
- Fine, G. A., & Deegan, J. G. (1996). Three principles of Serendip: insight, chance, and discovery in qualitative research. *International Journal of Qualitative Studies in Education*, 9(4), 434–447.
- Freddo, A. R., & Tacla, C. A. (2009). Integrating social web with semantic web: ontology learning and ontology evolution from folksonomies. *KEOD 2009 Proceedings*, 247–253.
- Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1), 4–13.
- Karapiperis, S., & Apostolou, D. (2006). Consensus building in collaborative ontology engineering processes. *Journal of Universal Knowledge Management*, 199–216.
- Lee, W.-N., Shah, N., Sundlass, K., & Musen, M. (2008). Comparison of ontology-based semantic-similarity measures. *AMIA, Annual Symposium 2008*, 384–8.
- Lemoisson, P., Surroca, G., & Cerri, S. (2013). Viewpoints: an alternative approach toward Business Intelligence. In *eChallenges e-2013* (p. 8).
- Limpens, F., & Gandon, F. (2011). Un cycle de vie complet pour l' enrichissement sémantique des folksonomies. In *Extraction Gestion de Connaissance EGC 2011* (pp. 389–400).
- Marchionini, G. (1997). *Information Seeking in Electronic Environments* (Cambridge., p. 224). Cambridge university press.
- Merton, R. K., & Barber, E. (2006). *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science* (Princeton., Vol. 2006, p. 313).
- Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1), 5–15. doi:10.1016/j.websem.2006.11.002
- Noh, T., Park, S., Park, S., & Lee, S. (2010). Learning the emergent knowledge from annotated blog postings. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4), 329–339.
- Pedersen, T., Pakhomov, S. V. S., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40(3), 288–99.
- Perriault, J. (2000). Effet diligence, effet serendip et autres défis pour les sciences de l'information. In *Pratiques collectives distribuées sur Internet*.
- Surroca, G., Lemoisson, P., Jonquet, C., & Cerri, S. A. (2014, May 13). Construction et évolution de connaissances par confrontation de points de vue: prototype pour la recherche d'information scientifique. *IC - 25èmes Journées Francophones d'Ingénierie Des Connaissances*.
- Tough, A. (1999). *Reflections on the Study of Adult Learning*. WALL Working Paper.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes* (Vol. 1978, p. 159).
- Yamaba, H., Tanoue, M., & Takatsuka, K. (2013). On a serendipity-oriented recommender system based on folksonomy. *Procedia Computer Science*, 22, 276–284.