COMPUTATIONAL MODEL FOR SERENDIPITY

A. Anguera, M.A. Diaz, A. Gutierrez

Abstract: *In recent years there have been various attempts and studies that are eager to serendipity in Computer Science. Authors such as Campos and Dias have tried to model the serendipity in order to get serendipitous behaviors in Information Retrieval (IR). There have been attempts to introduce serendipity in Recommender Systems (RS), although the latter proposals have led efforts using metrics for measure Serendipity in those RS, rather than to emulate Serendipitous behaviors in recommender systems. However, so far there haven´t been succeeded in designing a model which can be applied to different Web browsing. The main problem we have detected analyzing the proposals in this field is that the solutions provided do not take into account the two aspects of the concept of serendipity. Do not consider that in addition to the accidental discovery of information that is not sought for, is also required some characteristics in the user like sagacity, perception, flexible thinking or intensive preparation. If we could develop a model that support any search engine or search tool, we would facilitate an incredible advantage to the user by offering specially information that the user is not focused upon. The aim of this paper is to propose a computational model that supports Serendipity and induce and facilitate serendipity through the use of a special-purpose designed system.*

Keywords*:* Supporting serendipity, : Designing Serendipity Serendipia, Data Minig, Artificial intelligence, SOM, Models of Computation

ACM Classification Keywords: H.2.8 Data Mining, F.1.1 Models of Computation, I.2 Artificial intelligence

Introduction

Making a journey through the history of discoveries, scientists or not, appears again and again, a phenomenon that is sometimes overlooked and sometimes is neglected, component by chance, coincidence, luck or even a "fluke" that entails the seemingly mysterious aspects it contains. This phenomenon, called Serendipity, refers to the accidental discovery of something valuable, find by, perhaps, luck, chance, fluke and sagacity of something that was not expected. It is also defined as the ability to make important discoveries by accident, or the ability to extract knowledge of random events. There are many examples of the value of serendipitous discoveries.

Although the importance of the phenomenon called serendipity, in the field of computer science there are not many jobs about it. What´s more, the few we founded are focused in the field of information retrieval and how to provide users of such systems serendipitous discoveries. Computer science, assessing the role of serendipity in discoveries, has tried to design systems that procure the serendipity. In any case, this research had its focus mainly on only one aspect of serendipity: "that of chance encounters". A step in the study of this subject is to consider not only the user's role in the perception of serendipity, but also how user perception should be assumed by the system itself to increase the chance of serendipity.

Applying agents to web browsers, some systems have designed trying to facilitate or induce serendipity. The first, Letizia [Lieberman 95], although no mention of the word serendipity, is described as an attempt to anticipate

what items may be of interest to the user, through inferences from his owns navigation behaviors. It doesn´t have predefined objectives or goals, and Letizia guides a search for linked documents , making recommendations to anticipate future needs.

Another relevant work in this field is the one done by Campos and Dias [Campos, 2001]. The authors, on the one hand, they develop what they call "Serendipity Equations", which would be better to call pseudoformal expressions, which help to discriminate whether a discovery is serendipitous or not. Campos and Dias identified four different types of serendipity, namely pseudoserendipity, serendipity, serendipity without "inspirational idea" and serendipity from incorrect knowledge. The authors provided a succinct description of the phenomenon in the following equations:

P1⊂(KP1) $\Rightarrow$ S1⊂(KP1, KM,KN) M⊂(KM)	Pseudoserendipity
P1⊂(KP1) P2⊂(KP2) $\Rightarrow$ M⊂(KM) S2⊂(KP2,KM,KN)	Serendipity
P2⊂(KP2) P1⊂(KP1) $\Rightarrow$ S2⊂(KP2, KN)	Serendipity wihtout an inspiring metaphor
P2⊂(KP2) P1⊂(KP1, EP1) $\Rightarrow$ S2⊂(KP2, KN)	Serendipity wiht incorrect knowledge

Table 1: Serendipity ecuations

P1	Original Problem.
KP1	Knowledge domain problem *P1*.
S1	Solution *P1*.
M	Metaphor.
KM	Knowledge domain *M*.
P2	Problem 2 inspirated by *M*.
KP2	Knowledge domain *P2*.
S2	Solution *P2*.
KN	New knowledge.
EP1	Incorrect knowledge for *P1*.

Table 2: Serendipity ecuations terms

From the formulation of the Serendipity equations, Dias and Campos developed Max, a software agent that uses information retrieval techniques and heuristic searches in information space to find information on the Internet that can motivate and promote serendipitous discoveries [Campos and Dias, 2001]. The mission of Max is to stimulate the user with unexpected information. They postulate six possible categories in the value of Max's suggestions:

- Category 1: the suggested pages were already known the suggestion has no value at all;
- Category 2: the suggested pages were not known and did not belong to any domain of interest, the suggestion has little value. Although not interesting at the present moment, the suggested pages may have some usefulness in the future (in the context of lateral thinking).
- Category 3: the suggested pages were not known, but belong to some domain of interest the suggestion has little value, since the user could have reached those pages otherwise (ex.: using a search engine);
- Category 4: the suggested pages were not expected, but they are slightly related to some domain of interest the suggestion is valuable, and it could hardly have been found by the user.
- Category 5: the suggested pages were not expected they did not belong to any current domain of interest of the user, but they sparked in him a new interest , the suggestion is extremely valuable since it is very improbable that the user would ever find the page on himself.
- Category 6: the suggested pages establish a new connection between two current domains of interest, the suggestion is extremely valuable (an insight occurred);

Another significant work about serendipity in the world of digital data and media is the one done by Beale [Beale, 2007], using artificial intelligence approaches to reduce the distance between the user interface and the system making it more natural for the user to be able to achieve their goals. Beale developed two differente systems which are able to, using appropriate technologies, keep the user at the centre of an interaction and still support them in making new discoveries in different ways, making for a more serendipitous environment.

For its part Ertzscheid and Gallezot (Ertzscheid and Gallezot, 2003) suggested serendipity is an increasingly valuable tool in helping in the search field to find information. Their starting point is the definition of serendipity as a leading model of access to information in the context of a search process or one of its iterations, considering that this model gives results that satisfy user´s needs in a casual way.

Serendipity can also occur during directed web search. Spink et al (Spink, 1998) suggest that results "partially relevant" may lead to the generation of new ideas and guidance for those seeking information. In this same field of Web search, Paul André and Teevan (André, 2009) conducted a study to determine if the results of web search results contain potentially serendipity. The participants were asked to categorize the results they obtained in their searches in two scales: Relevance (relevant, partially relevant and irrelevant), and degree of interest (Interesting, interesting part is not interesting.) The starting working hypothesis is that the results judged as not very relevant, but at least interesting, is an area where you can be serendipity. In addition to collecting relevant trials were also able to obtain additional information about each query and each result by examining search logs.

One of the areas where they have also developed different projects over the past years regarding the serendipity has been the Recommendation Systems (Recommender System, RS). Next we are going to expose some of these works of particular relevance.

In information recommendation, a critical aspect is to provide useful information to the user. These systems have, as a first objective, to provide personalized recommendations to improve user satisfaction.

In the last decade, several recommender systems have been developed and used in a variety of domains. As more and more are proposed recommendation techniques, researchers and practitioners face the problem of how to estimate or measure the value of the recommendations. From the perspective that good recommender systems recommend items that suit the preferences of the user, they are evaluated by using metrics related to accuracy. Two popular metrics for measuring accuracy are precision and recall, which were developed in the research area of information retrieval. In addition to precision and recall, various predictive accuracy metrics are used in recommender systems.

However, some recent papers argue that accuracy is not the only value that should be considered for the evaluation of recommender systems and that there are other important aspects to consider, and that assessments should focus on the future. The evaluation metrics for measuring user satisfaction move beyond accuracy to include novelty and serendipity

Computational Model

Although in recent years there have been efforts for implementing computer systems that directly support serendipity in search engines and information retrieval and also in recommendation systems, there has not been developed any computational model system for serendipity. The main cause of the lack of this model is one that, as pointed out Van Andel (Van Andel, 1994) the very fact of "program" serendipity invalidate the concept itself.

Another limitation detected by analyzing the scientific literature on this subject, is that in general, there has not been considered the dual nature of the concept of serendipity in the designing of these systems. The model we propose consider the concept from its duality, namely, first an unexpected event should occur, by chance or fluke, and also the necessary observer attitude that is necessary to make possible a discover of a new opportunity for generating knowledge.

Advances in information technology, especially in terms of the potential of current hardware and the new opportunities offered by the current in the field of human-machine interface, can face the design of this model from a new perspective that includes both the ability of the system itself to generate serendipity as the willingness of the user before it. It is from this perspective that arises from the computational model for serendipity.

As a first approximation to Serendipitous computational model, we describes the steps that should be reflected in to accommodate both the possibility of generation of serendipity as to the attitude that every time the person concerned to show that time that the system will offers. Therefore, the model should provide the opportunities that the computer system is implemented on he has the ability to present serendipitous results to requests for information made by the user, and the different answers you may give these results, the model must understand first the possibility that during the search for information arising actions not directly related to the exact terms and search domain user and, secondly, what action does the user to this new fact (not expected by him).

Before formulating the model, it should indicate that it should be collected in different situations about serendipity can occur. First, it may not occur either because the user serendipity is not interested in investigating the unexpected or because, even so, does not reach any valid conclusions (IF POSSIBLE, IF WORKABLE, undesirable outcome). In this frame of mind, you can find situations where the unexpected event

pseudoserendipia help solve the problem initially in the search (IF POSSIBLE, IF WORKABLE, undesirable outcome). Second and more interesting for what we are concerned, pure serendipity can occur if the act emerged by chance helps users find a solution to a problem that already existed but was not the target of their search or investigation of the discover so much a problem as a solution for the same had not been raised so far (IF POSSIBLE, IF WORKABLE, IF DESIRED).

1. Identifying a research problem.
2. Investigation of this problem: finding a solution.
3. In the research phase of a problem, the appearance of an unexpected event.
4. With the emergence of an unexpected event can take the following actions:
    A. Not investigate the surprising fact (IF POSSIBLE, IF WORKABLE, undesirable outcome). The user may not realize the importance that presents the surprising fact, not investigate. In this case there will be no serendipity. However, there are two situations:
        - Discarded, which implies that the user does not forget and we can be of any use in subsequent investigations (IF POSSIBLE, IF WORKABLE, undesirable outcome).
        - Record it, which that user or another we can be of any use in subsequent work. (IF POSSIBLE, IFWORKABLE, undesirable outcome).
    B. Investigate the surprising fact. In this case the user is aware of the importance of the event and therefore the need to investigate (IF POSSIBLE, IF WORKABLE, undesirable outcome). In this case there are alternatives:
        - You can investigate and do not come to any conclusion. In this case, there is no serendipity. These findings can be recorded and may be used for future research. In any case, will become part of the "background" of the user to possible future serendipitous discoveries (IF POSSIBLE, IF WORKABLE, undesirable outcome.
        - You can investigate and come to the conclusion that it is the solution.
            i. the problem he was studying, in which case it serendipity, but pseudoserendipity (IF POSSIBLE, IF WORKABLE, undesirable outcome)
            ii. a problem existed but that he was not studying at that time. In this case, it is Serendipity (IF POSSIBLE, WHICH CAN BE MADE IF, IF DESIRED)
            iii. that is an important discovery. It has just discovered a problem and its solution. Again, this is a case of serendipity (IF POSSIBLE, IF WORKABLE, IF DESIRED)
5. Knowledge generation and use.

Figure 1 shows the overall system architecture in which you can see the different modules that compose it. As shown in the figure, the user is logged by a search request on a topic of interest to him in a particular domain. The system collects the search significant tokens and sends them to a module that is responsible for transforming these tokens in serendipitous tokens enriched. This transformation is done using a neural network responsible for obtaining new words based on the concept of semantic distance. To develop the architecture of this network model we start from the SOM (Self-Organizing Maps) [Kohonen, 2002], [Somervuo, 2004] in which instead of syntactic distances between neurons arise semantic distances.

The tokens obtained in the neural network, process manager searches in launching these tokens in different search engines. Documents obtained from these searches are filtered and categorized by whether they belong to the initial search domain or a different one. These documents are presented to the user making an assessment based on the interest and use that report, both in relation to your original search with other search domains that had not initially plant.

This user interaction the system will learn to guide the user in subsequent searches.

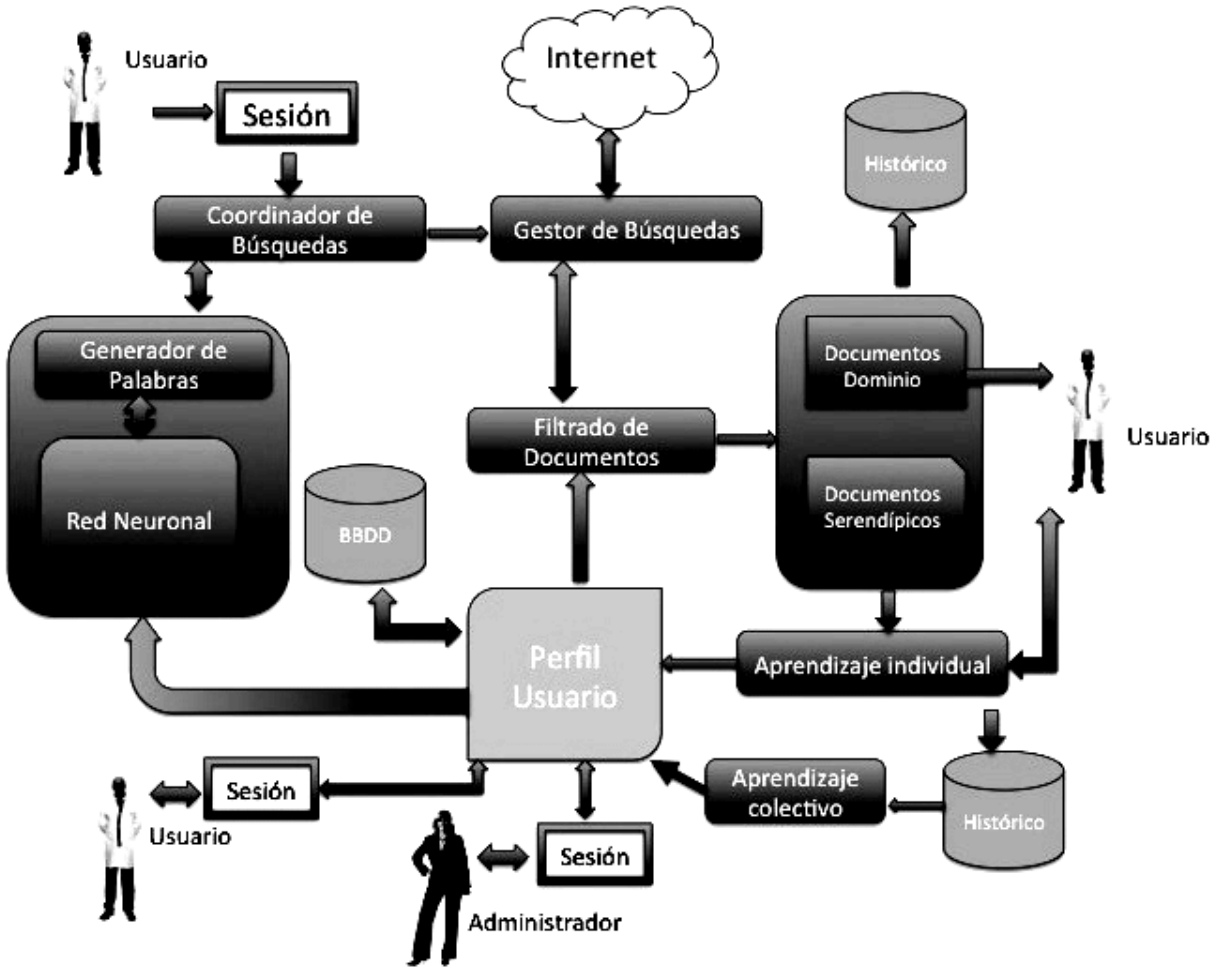

Figura 1: Modelo arquitectonico del sistema

Conclusions

The proposed model allows us to consider new lines of research in relation to the proposed neural network to generate the tokens Serendipic as regards the integration of serendipity in the various current systems in Internet search.

The system allows different forms of serendipity collected in the field equations and Dias, as the user's attitude to what the system offers and the final result is obtained.

The model provides a new way to treat the phenomenon of serendipity computationally integrating the two parts of the same: the emergence of a surprising fact (which provides serendipity generation module) and facilitates the user's predisposition to this fact

Bibliography

[André I, 2009] Paul André, Teevan J, Dumais Susan , Discovery Is Never by Chance: Designing for (Un)Serendipity. C&C09 Octubre 26-30, 2009, Berkeley, California 305-314

[Beale, 2007] Beale, Russell: Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. International Journal of Human-Computer Studies. 5-mayo-2007, pages 421-433

[Campos I, 2001] Campos, J. y Dias de Figuereido, A. 2001. The Serendipity Equations. Proceedings of the Workshop Program at the Fourth International Conference on Case-Based Reasoning., ICCBR 2001, Technical Note AIC-01-003. Washinton, DC: Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence.

[Campos II, 2001] Campos y Dias. Searching the Unsearchable: Inducing Serendipitous Insights. 2001

[Ertzscheid, 2003] Ertzscheid, O. y Gallezot, G. Buscar falsedades y encontrar exactitudes, serendipia y búsqueda de información. CIFSIC Bucarest 2003. Comunicación y complejidad. 2003

[Kohonen, 2002] Kohonen, T., Somervuo, P.J.: How to make large self-organizing maps for nonvectorial data. Neural Networks. Vol. 15. Pp: 945-952.

[Somervuo, 2004] Somervuo, P.J.: Online algorithm for the self-organizing map of symbol strings. Neural Networks. Vol. 17, issues: 8-9. Pp: 1231-1239.

[Spink 1998] Spink, A., Greisdorf, H. & Bateman, J. From highly relevant to no relevant: Examining different regions of relevance. IP&M 34,5 (1998), 599-621

[Van Andel, 1994] Van Andel, Pek. "Anatomy of the unsought finding: serendipity: origin, history, domains, traditions, appearances, patterns and programmability". British Journal for the Philosophy of Science 45 (2): 631–648. 1994

Authors' Information

Aurea Anguera de Sojo Hernández- *Associate professor U.P.M. Crtra Valencia km 7, Madrid 28031, Spain;*

e-mail: aanguera@eui.upm.es

Miguel Angel Díaz Martínez- *Associate professor U.P.M. Crtra Valencia km 7, Madrid 28031, Spain;*

e-mail:mdiaz@eui.upm.es

Abraham Gutiérrez Rodríguez- *Asocciate professor U.P.M. Crtra Valencia km 7, Madrid 28031, Spain;*

e-mail:abraham@eui.upm.es