Towards a model for serendipitous discoveries

Chan Gu Kang, Hyeonmin Park, and Seongbin Park*

Department of Computer Science, College of Informatics, Korea University

**Corresponding Author: hyperspace@korea.ac.kr*

1. Introduction

What is serendipity?

Serendipity is an experience of encountering unexpected things that leads to new ideas of solving problems, research, etc. Serendipity also can be found in daily lives such as finding books in library, readings, and even in conversations through phone.

Systemic approach of serendipity

- By analyzing cognition of a user encountering serendipity in a system, definition of serendipity can be presented as a series of processes [1]

Perception ⇨ Attention ⇨ Interest ⇨ Explanation ⇨ Bridging ⇨ Valuation

- We give devised definition of serendipitous information encounter.
- To devise the method, we study on user navigation of information space, where users are encountering information only by links between information node.
- We take the path-based approach of the graph model. For serendipity evaluation, we quantitatively assess each certain path in graph, and therefore be able to make structural calculation of an information space.

2. A model for the serendipity of an information space

Information Space

Definition 1 (Information Space). An n-level information space S is a directed graph $S = (V, E)$ satisfying the following three constraints:

1. *V can be partitioned into sets $V_0, V_1, \dots, V_n$ and satisfies the following.*
   a. *The family of sets $\{V_0, V_1, \dots, V_n\}$ is a partition of V*
   b. *$V_0 = \{v_0\}$ where v_0 is root node (having no information or a meaning)*
   c. *For $i = 0,1,\dots,n$, we define V_i to be set of nodes at level i.*
2. *E can be partitioned into sets $E_0, E_1, \dots, E_n$ and satisfying the following.*
   a. *The family of sets $\{E_0, E_1, \dots, E_n\}$ is a partition of E*
   b. *For $i = 0,1,\dots,n-1$, $E_i \subseteq V_i \times V_{i+1}$*
   c. *$E_n \subseteq V_n \times V_n$*
3. *For $i = 0,1,\dots,n-1$, $G_i = (V_i \cup V_{i+1}, E_i)$ is a forest whose connected components are rooted trees,*
   where every node $v_i \in V_i$ should be root node of one rooted tree of forest G_i.

- Our intention of constraints is to make each node in the space to be specifically classified by its upper-level node.
- This implies that nodes in V_n to be do-main-unique.
- Our Model can be seen as abstract Linked Data [6] where the relations between in-formation are explicitly stated and so that the hierarchy of information by domain can be formed.

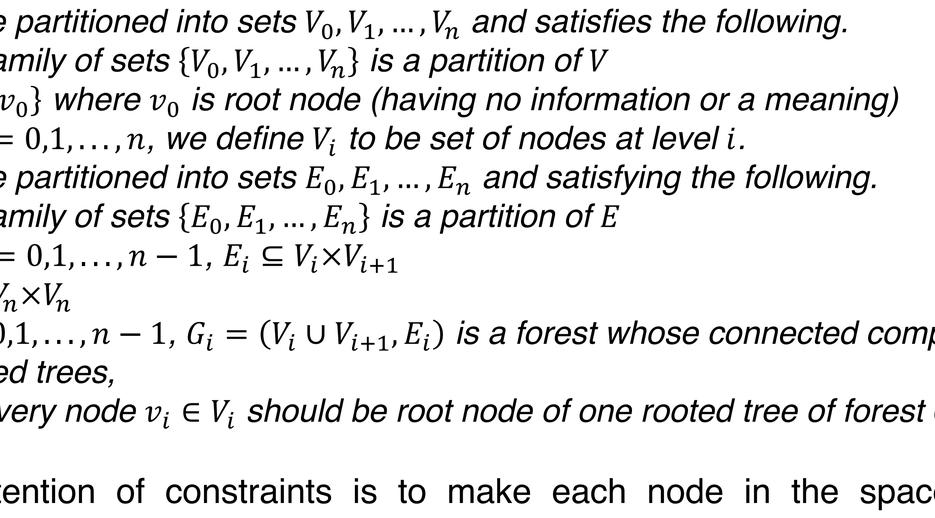

Fig 1. An example of a 4-level information space of Definition 1

Serendipitous Path

- Consider paths within the information space with special properties as follows. The definition is based on systemic approach of serendipity.

Definition 2 Given an information space $S = (V, E)$, a path $p'$of $\mathrm{v_i} \rightarrow^{p'} \mathrm{v_j}$ where $v_i, v_j \in S.V_n$ is serendipitous if it satisfies the following three conditions. For explanation, let $v_0 = v_i$, $v_n = v_j$, and $p' = v_0, v_1, v_2, \dots, v_n$.

1. *(**Well Discovered**) The start node v_1 is easy to be discovered. In terms of information space, a node may be easily discovered if a user can be more likely to find or encounter the node while an information browsing or navigation.*
2. *(**Interest**) The sequence of nodes $v_1, v_2, \dots, v_n$ is interesting enough for a user to keep navigating through the nodes of path p'. That is, each node in the sequence should keep users to ride on the sequence without abortion or distraction in navigation en route.*
3. *(**New Connection**) The destination node v_n has enough possibility of leading user to 'new connection' (such as new ideas, knowledge etc.). Suppose a user is navigating within information space. We can think of a case a user is likely to navigate information nodes which belongs to his/her background domain. To a user, a node leads to new ideas if it is something different with user's existing knowledge. Thus, while navigating, we can say that an information node that offers 'new connection' is one that offers links to nodes that belong to a different domain that the user belongs to. We may call such link 'cross-domain links'.*

Evaluation

- We focus on 'sidetracked path' p' from user's normal navigation.
- The metric is calculating how much p', satisfies the condition of serendipitous path.
- The evaluation object is limited to 'potentially serendipitous' path, that only satisfies ***Well Discovred*** and ***Interest*** condition of Definition 2.

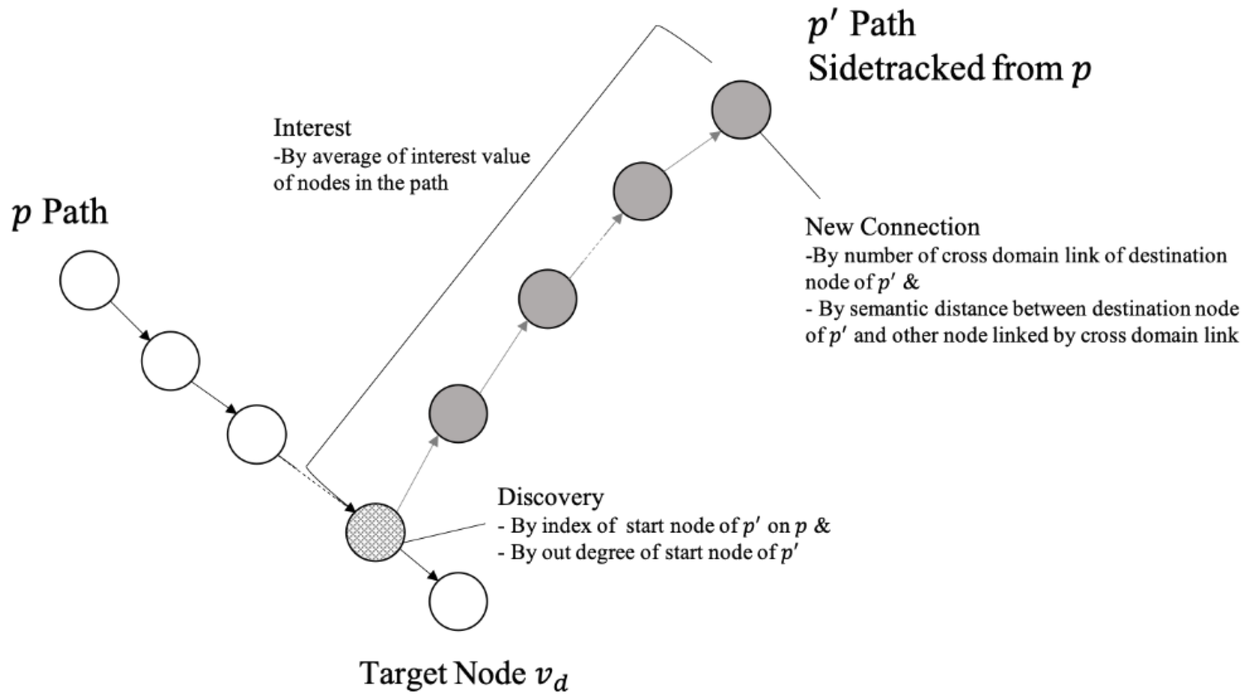

Fig 2. An Illusatration of Evaluation Ideas

3. Conclusion

- We defined serendipity to be consisted of discovery, interest, and possibility of new connection.
- The idea of our metrics is to numerically calculate how serendipitous a path that is sidetracked from a path towards a target node is.
- Using our method, it is expected that a serendipitous information space can be designed so that while navigating the space users can encounter an unexpected but impressive information frequently.

Implementation

https://github.com/ckhp/SerendipityEvaluationModel

References

1. J. Corneli, A. Jordanous, C. Guckelsberger, A. Pease, S. Colton, Modelling seren-dipity in a computational context, arXiv preprint arXiv:1411.0440.
2. Linked Data *Evolving the Web into a Global Data Space*, Morgan & Claypool Pub-lishers, 2011.