

## Kid website assessment using knowledge discovery

Fauzan Nooraddin

Muna A. Ali

Rasheed Nassr

College of Information and Communication Technology

International Islamic University Malaysia

Kuala Lumpur, Malaysia

### Abstract

*Knowledge management and discovery play great role in business to improve the quality of information and help decision making process to issue accurate decision. However; those tools may employed in internet safety for kids, families need to verify whether specific kid website is an appropriate and helpful for kids. The main challenge of those tools is how to scan content and figure out the nature, content-suitability for kids and quality of information published in that content. this paper tries to investigate the features that suppose to be available in the website to be suitable for kids, those features will be used in proposed framework that is may used to assess suitability and usefulness of the website for kids.*

**Keywords:** content suitability, knowledge management, ontology

### *i. Introduction*

Internet becomes vast and uncontrollable source of information and allows people to get allot of information faster and easier, there is no restriction on type and amount of information that put in it. Even though Internet allows people to express freely and publish whatever they want, there is no guarantee that information presented in the internet is true, valid or safe. Usually the concern about validity and trustfulness of information is found in research and academic field where information that has no trustful source will be rejected. However; regular people could not access to tools and resources that available in research and academic field, but they need a mean to verify information posted on web. People may need to verify whether the internet sources they allow their children to browse is safe, educational and helpful to raise children correctly or make it worse.

Families need not just for a tools that filter websites to protect children against unethical content but may need a tools that help them in decision making process to determine which children website is the most appropriate and helpful in terms of education, raising and provides good materials that merge between entertainment and education; as well as it is better to provide colorful content and safe games. Those features may be used with some of those features that most literatures agreed on: ease of use, security/privacy, visual appearance, and information quality (Chiou, Lin, & Perng, 2010; Vela'squez, Dujovne, & L'Huillier, 2011). The website information will be investigated according to keywords stored in database. Those keywords are considered among the most important keywords that should be available in any children websites.

### *ii. Literature review*

The number of kids browsing web and social networks increase, so it is recommended to utilize analysis tools to help families to select the appropriate websites for children (Behzadi & Noghabi, 2010). So the Internet is considering amazing tools for children. Because they can use the internet for searching about allot of things which that may related to the study matter such as doing a school reports, as well as using the social network to communicate with their teachers, or with their friends. Internet is considering main tools for entertainment and kids usually spend time for playing some interactive games. Parents have to control the searching mechanisms for their kids because sometimes children might do online searching for example for the world "Lego." But they may just change one letter mistakenly and the word become "Legs"

Accordingly, in this case the child will directed to wrong websites where they focus on the word leg and usually these websites contains some pornography materials. Therefore, for this reason parents have to aware about what their kids see and access on the internet pages. Behzadi & Noghabi, (2010) analyzed kid websites in terms of design and content and two other features to test suitability of the content, design and easy to use with children. However, features regarding to education and manners had not been investigated.

Processing text and trying to retrieve semantic knowledge is as old as newspaper, content analysis started manually to retrieve semantic knowledge from newspapers and magazines (Hellemans & Govers, 2005). Content analysis is used by different industries to analysis their websites to improve the quality of the services and increase customer intimacy (Hellemans & Govers, 2005; Petch, 2004). According to Pooch & Lefond, (2001) content of the website is a significant factor of website evaluation measurments. Usually, one way to evaluate website and retrieve semantic knowledge is through searching for frequent patterns where the terms that are considered significant are used to measure certain quality, It has been used to analyse customer shopping cart in order to do good marketing (Han, Cheng, Xin, & Yan, 2007). However; this frequent can be utilized in this paper to find out the existence of certain activities, games, educational topics etc. That may be considered as a good feature or bad in websites for kids. Similarly web semantic came for the same purpose with more improved tools to find semantic knowledge from website (Janev & Vraneš, 2011).

Semantic technology helps to split the meanings from the data, so the compute can understand the documents meaning. The semantic web is actually provides a structure which allow data to be shared and reused across applications. Semantic web provides a great opportunity to retrieve some knowledge about website and gives intelligence analysis for website which could help web master to improve the quality of their websites (Vela'squez, Dujovne, & L'Huillier, 2011). We can think of the Semantic Web as a resourceful way to embody data on the World Wide Web, or as a database that is internationally linked, in a manner comprehensible by machines, to the content of documents on the Web. Semantic technologies signify meaning with ontology and offer reasoning through the relations, regulations, reason, and circumstances represented in the ontology.

Moreover, web semantic needs ontology because ontology could define structure and pattern of knowledge that suppose to be retrieved using semantic web (Guo & Zhang, 2009; Vela'squez, Dujovne, & L'Huillier, 2011). It is agreed that the requirments of information quality varies from website to another (Chiou, Lin, & Perng, 2010), in this paper- uses childern website- the consideration is how to be ensured that childern website deliver save and educational information, in the same time deliver it in pleasure way. Many proposals have analzed the contents of different sectors such as tourism website, healthcare and e-commerce using web semantic and ontology, the investigation may start by information quality and ends by visual features such as font type, color and size (Hellemans & Govers, 2005; Petch, 2004). The main challenge that is need to be solved is how to construct ontology i.e. what are the main domains, relationships and keywords that are supposed to be evaluated and how to measure the quality of web semantic results. Many studies first surveyed experts and others to collect the viewpoints of the type and quality of information that suppose to be featured the content of the websites (Chiou, Lin, & Perng, 2010; Mustapasa, Karahoca, Karahoca, Yücel, & Uzunboylu, 2010; Pooch & Lefond, 2001; Petch, 2004). Then they use data mining algorithms to scan website content and find out the significant information (Vela'squez, Dujovne, & L'Huillier, 2011).

### **iii. Methodology**

Among methods used to analyse data is seeking significant frequent patterns that meet minimum threshold to evaluate the data, methods such as Apriori, FP-growth and Eclat. Those methods process database, which is contains structured data, and generate frequent patterns and their existence in the database (Han, Cheng, Xin, & Yan, 2007). However the case in this paper is different because website content does not have structured pattern and needs different method that may employ crawler method to retrieve frequent patterns from website content as well as frequent patterns are predetermined because it is difficult to find association among patterns in the database where data is mostly stored in raw form (Han, Cheng, Xin, & Yan, 2007). Apart, data mining methods analyse multidimensional data (Han, Cheng, Xin, & Yan, 2007), this may be dissimilar with website as data space which may need a different methodology such as web content mining that could use ontology which provide flexible, structured and meaningful representation of the patterns (or web objects) that makes searching for semantic knowledge easier which may lead eventually to significant evaluation. Web content management is knowledge discovery (Han, Cheng, Xin, & Yan, 2007).

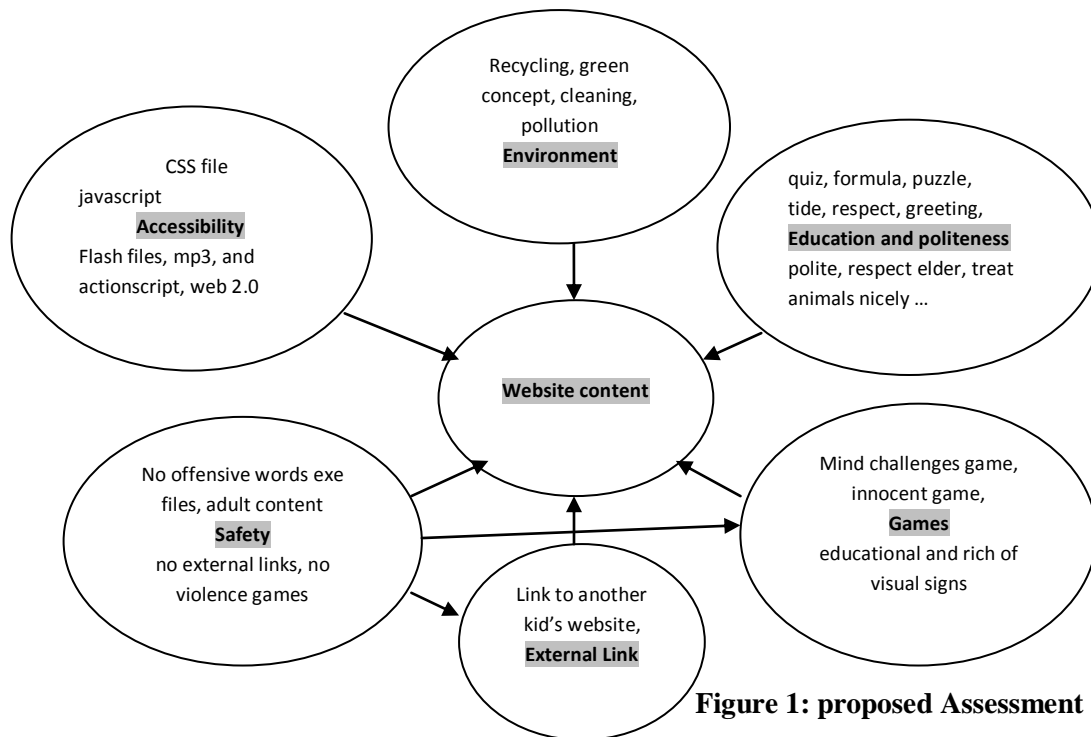
Certain literatures use ontology to define and create the web object- the structure and pattern that knowledge in the website could be shaped; then extract and compare objects (Vela'squez, Dujovne, & L'Huillier, 2011). However; some certain content analysis techniques can analysis images for semantic purposes (Hellemans & Govers).

In order to make easy text semantic Zhixing, Zhongyang, Yufang, Chunyong, & Kuan, (2010) proposed interpreting words and text fragments in the space of concepts. Space of concept contains the main keywords related to specific concepts which make the process of retrieving semantic knowledge easier, however; F. Noy & L. McGuinness recommended adding synonyms and speech words to the keywords list. However text semantic involves expensive CPU processing for text (Zhixing, Zhongyang, Yufang, Chunyong, & Kuan, 2010). While Web semantic uses ontology as a tool to represent knowledge which makes knowledge inference easier and smarter, ( Joo, 2010; (Mustapasa, Karahoca, Karahoca, Yücel, & Uzunboylu, 2010). Ontology uses the same concept of object-oriented programming however the major concern is how to represent attributes and data (F. Noy & L. McGuinness). Then ontology is significant to web semantic (Joo, 2010; Barros, Silva, Costa, Bittencourt, Holanda, & Sales, 2011), w3c recommends many technologies that build ontology graphs and make semantic web applicable (Janev & Vraneš, 2011).

Hence, ontology is responsible to represent the structure (create objects with attributes with appropriate relationship with each other) of the knowledge and information that suppose to be available in the website. Web semantic takes this presentation as input, starts process the website and produces summary report. This report is comprehensive and summarizes the main keywords and context that researcher seeks.

**iv. Categories of children website’s content**

The proposed framework to assess suitability of the website for kids can be depicted diagram in figure 1, the expected keywords categories that suppose to be available in website that intended to be directed to the kids. This diagram illustrates the categories that may involve to assess website, the main consideration is suitability; safety category is substantial it will be used to check external links, games, educational content to ensure suitability for kids. Other categories may determine how elegant and comprehensive this website. Every category may contain list of the main keywords that should be available in the website.



**Figure 1: proposed Assessment Framework**

**v. Conclusion**

The process of collecting and classifying keywords into categories then establish relationship among those categories may need more knowledge and concern; consequently will provide a clear idea about the suitability and usefulness of the website for kids. Ontology helps in classifying keywords and creating relationships, this classification will be fed into web discovery tool to look for and assess those keywords within the content of the website. However; performance is a critical consideration when knowledge discovery tools are used because it consumes substantial CPU cycles, and memory space

## References

- Barros, H., Silva, A., Costa, E., Bittencourt, I. I., Holanda, O., & Sales, L. (2011). Steps, techniques, and technologies for the development of intelligent applications based on Semantic Web Services: A case study in e-learning systems. *Engineering Applications of Artificial Intelligence*, 1355–1367.
- Behzadi, H., & Noghabi, M. Z. (2010). Persian websites for children and teenagers in Iran. *LIBRARY HI TECH NEWS*, 20-22.
- Chiou, W.-C., Lin, C.-C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management*, 282–290.
- F. Noy, N., & L. McGuinness, D. (n.d.). *Ontology Development 101: A Guide to Creating Your First Ontology*. Retrieved 10 10, 2011, from <http://protege.stanford.edu>:  
[http://protege.stanford.edu/publications/ontology\\_development/ontology101-noy-mcguinness.html](http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html)
- Guo, Q., & Zhang, M. (2009). Question answering based on pervasive agent ontology and Semantic Web. *Knowledge-Based Systems*, 443–448.
- Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: current status and future directions. *Data Mining Knowledge Discovery*, 55–86.
- Hellemans, K., & Govers, R. (2005). European Tourism Online: Comparative Content Analysis of the ETC Website and Corresponding National NTO Websites. *Information Technology & Tourism*.
- Janev, V., & Vraneš, S. (2011). Applicability assessment of Semantic Web technologies. *Information Processing and Management*, 507–517.
- Joo, J. (2010). Adoption of Semantic Web from the perspective of technology innovation: A grounded theory approach. *Human-Computer Studies*, 139–154.
- Mustapasa, O., Karahoca, D., Karahoca, A., Yücel, A., & Uzunboylu, H. (2010). Implementation of Semantic Web Mining on E-Learning. *Procedia Social and Behavioral Sciences*, 5820–5823.
- Petch, T. (2004). content analysis of selected health information websites. *action for help*.
- Poock, M., & Lefond, D. (2001). how college-bound prospects perceive university websites: findings, implications, and turning browsers into applicants. *c & u journal*, 15-21.
- Vela'squez, J. D., Dujovne, L. E., & L'Huillier, G. (2011). Extracting significant Website Key Objects: A Semantic Web mining approach. *Engineering Applications of Artificial Intelligence*.
- Zhixing, I., Zhongyang, X., Yufang, Z., Chunyong, L., & Kuan, L. (2010). Fast text categorization using concise semantic analysis. *Pattern Recognition Letters*, 441–448.