Improving Government to Consumer (G2C) E-Governance Models in Kingdom of Saudi Arabia: A Web Usage Mining Based Approach

Hisham Al Saghier

College of Computer and Information Sciences, Majmaah University, P.O Box 66,
11952 Al Majmaah, Kingdom of Saudi Arabia

Abstract: At present, most of the countries are using the E-governance model to provide better services to the citizens/business. The services provided to the citizens are always create a greater impact as this serves as a backbone for other kind of E-services. These services are referred as Government-to-Consumer (G2C) models in E-business. The Kingdom of Saudi Arabia uses G2C Models for most of their operations and provide greater services to all the citizens and residents. Some example services provided by the kingdom are communication and technology, training and education and culture, housing and municipal services, labour and employment, etc. The data generated by this information systems are very huge and it is the appropriate time to the government to facilitate more and better services. This research paper proposed an architecture which can be used with the available information systems for improving the G2C services. As all the services by the kingdom are web-enabled, this research work uses web-mining models in the architecture. The web mining models mine the web usage data available in the server and provide the output in the form of knowledge. Web-usage mining also will help the government in setting up the strategic plans, predicting the citizen/resident behaviour. The government decision makers to improve the given services then use the knowledge generated by the proposed model.

Key words: E-Governance, web-mining, plans, labour, service

INTRODUCTION

In the past decade, the development of electronic government shows a rapid growth. Due to increase in the speed of the present generation networks, most of the nations would like to provide all the services to the citizens. The citizens will get a greater benefit by accessing all the needed services online. E-governance is one of the best application of data mining due to the voluminous of data generated out of various online services provided by the government. Using data mining in E-governance helps the government in providing more valuable and customizable services to the end users. Web mining is a branch of data mining which deals with the extraction of hidden and interesting knowledge from the large volume of web documents and records (Kamber and Han, 2006). Web mining is divided into three types: web content mining, web structure mining and web usage mining. Web content mining focuses on the raw information available on Web pages. Web structure mining focuses on the structure of Web site including intra-page and inter-page structural information presented on web pages. Web usage mining deals with the extraction of knowledge on users, access patterns and

user behavior from data collected from the main sources such as web servers, proxy servers and web clients using some kind of data mining techniques. The sources are also called as web log files.

The Kingdom of Saudi Arabia provides various services to the citizens as listed in the study. As of now, the database will be voluminous and contains more useful information such as user navigation patterns, user usage patterns and user account details, etc. This is the appropriate time for the government to make more useful decisions by referring the historical databases. The proposed research work uses web usage mining in these databases to extract useful knowledge. The generated knowledge will be of more use to the government for improving the services and introducing new value added services.

Services provided by the kingdom of Saudi Arabia and need of using weblog data: The kingdom of Saudi Arabia provides various services to the citizens/residents. Saudi Arabia is one of the biggest countries in the Gulf region and is in process for a transition to E-government where Internet and Communication Technology (ICT) is playing an important role in Saudi Arabia (Alshehri and Drew,

```
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /headerhome.js - 200 0 939 283
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /wcstyles.css - 200 0 10213 282
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /images/27pct-banner.gif - 200
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /images/subway2.jpg - 200 0 114
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /images/ligature_spacer6.gif -
:09:13 24.161.88.191 - W3SVC175 WSH120 80 GET /images/wclogov3t.gif - 200 0 2
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /whatwedo.html - 200 0 7073 382
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /header.js - 200 0 1145 292 0 H
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /wcstyles.css - 200 0 10213 295
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /images/ligature_spacer6.gif -
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /images/whitespacer.gif - 200 0
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /images/wclogo100w.gif - 200 0
:09:17 24.161.88.191 - W3SVC175 WSH120 80 GET /footer.js - 200 0 1083 292 0 H
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /wcjournal.html - 200 0 3990 39
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /header.js - 200 0 1145 293 0 H
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /wcstyles.css - 200 0 10213 296
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /images/whitespacer.gif - 200 0
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /images/ligature_spacer6.gif -
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /images/btn-subscribe.gif - 200
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /images/wc-journal-logo-250x50.
:09:26 24.161.88.191 - W3SVC175 WSH120 80 GET /footer.js - 200 0 1083 293 0 H
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /whoweare.html - 200 0 3292 396
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /header.js - 200 0 1145 292 0 H
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /wcstyles.css - 200 0 10213 295
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /images/whitespacer.gif - 200 0
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /images/ligature_spacer6.gif -
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /images/wclogo100w.gif - 200 0
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /images/weidman2-100x125-bw.jpg
:09:32 24.161.88.191 - W3SVC175 WSH120 80 GET /footer.js - 200 0 1083 292 0 H
:09:43 24.161.88.191 - W3SVC175 WSH120 80 GET /contactus.html - 200 0 2406 39
:09:43 24.161.88.191 - W3SVC175 WSH120 80 GET /header.js - 200 0 1145 293 0 H
:09:43 24.161.88.191 - W3SVC175 WSH120 80 GET /wcstyles.css - 200 0 10213 296
```

Fig. 1: Sample of a web-log file

2010). To implement the E-governance services, the Ministry of Communication and Information Technology (MCIT) established an E-government program namely 'Yesser'. The ministry of finance and Communication and Information Technology Commission (CITC) also partners with MCIT to make the implementation more easy and reachable. The Yesser program implemented by the kingdom provided the following services to the benefit of all citizens/end-users:

- Communication and technology
- Training and education and culture
- Housing and municipal services
- Labour and employment
- Transport
- Economy and business
- Social life
- Islamic affairs
- Utilities (electricity, water, sewage services, etc.)
- Personal documents
- Insurance and pension
- Travel and tourism
- Health and environment
- Traffic and safety

The data collected by these systems represent the navigational patterns of the users. This data will be considered as a primary source for web mining in this proposed model. Whenever, the user enters in any of the

above systems, the server generates a hit and maintains a log. Each log may contain the user name, IP address, type of browser, operating system used, webpages visited etc. This log files will be collected from all the above-mentioned services and web-mining models will be applied to discover the hidden knowledge. This discovered knowledge thus applied to systems to improve the present provided services. For example, Fig. 1 provides a sample of a weblog file generate for a chosen date.

Literature review: Ali Al-Ahmary explain the integration of various E-services in the Kingdom of Saudi Arabia and proposing a framework for the same. Ali S Al-Soma presents the key achievements of the first and second phase of the Saudi E-government implementation. The researcher discusses E-government infrastructure, change management, transformation indicators, positive change indicators in detail.

Al-Mushayt *et al.* (2012) discusses the tangible benefits of E-governance in the Kingdom of Saudi Arabia and discusses the components of E-government framework.

Basamh *et al.* (2014) addresses the current practices, challenges and obstacles that affect the implementation of E-services from the perspective of a society. Researchers concludes that the government needs to out more effort in providing better e-services and reduce the digital divide.

14

Alateyah *et al.* (2013) presents the factors affecting the citizen's intention to adopt E-government in the Kingdom of Saudi Arabia. Various factors such as quality if service, diffusion of innovation, computer and information literacy, culture, lack of awareness, technical infrastructure, website design and security are identified and discussed in detail (Alateyah *et al.*, 2013).

Qwaider and Al-Shafi (2013) elaborates the factors currently impeding the development and implementation of e-Government in the Kingdom of Saudi Arabia (Tamrakar, 2012).

Borges and Levene (2005) present a new web usage mining model which uses higher-order Markov model in clustering. This model also takes into account higher-order conditional probabilities.

Miguel Dario Dussan-Sarria and Elizabeth Leon Guzman proposes a recommendation system to be used by the university community. This system contains an offline and online model. Data pre-processing is also concentrated in this study.

Poongothai *et al.* (2011) discusses about building of a robust model for web usage mining based on clustering. The researcher optimizes the usage framework with fuzzy C means clustering technique and also proposes an evolutionary clustering algorithm.

Sain and Tamrakar (2012) proposes a web usage mining technique based on HMM (Hidden Markov Model) and fuzzy clustering. This model measures the similarity efficiently among the users on the basis of their navigational patterns (Qwaider and Al-Shafi, 2013).

It is clear from the above reviewed literatures that any attempt is not being made by a researcher to apply the web usage mining on the E-government data. The literature also justifies the usage of various web usage mining models in the proposed research work.

MATERIALS AND METHODS

Web usage mining architecture: The web usage data available with Yaseer is converted into a data warehouse. This warehouse contains all the web log data from all the E-services. The data available in this data warehouse may contain missing data, erroneous data or inconsistent data. A data cleaner tool removes these initially. The end user will supply he necessary input to this tool to get the accurate data for mining (Fig. 2).

Key elements of data pre-processing in web server log files: Data pre-processing is very important in the web-server logs. As the proposed architecture collects
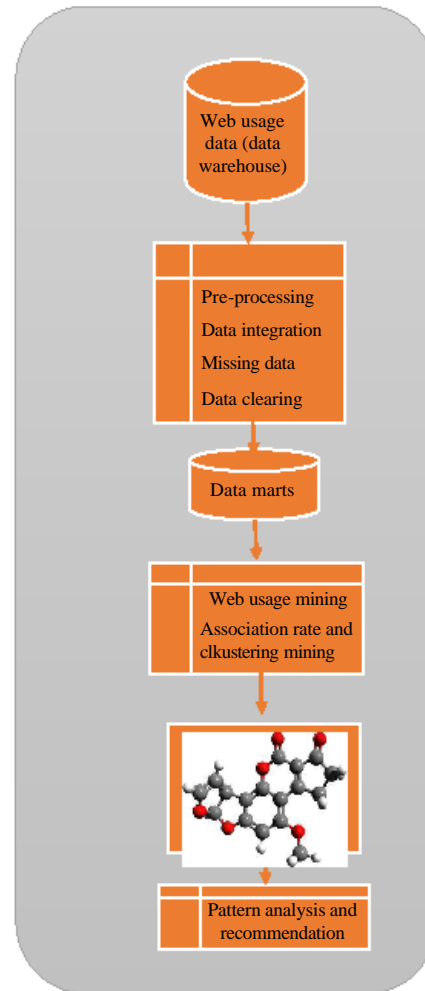

Fig. 2: Proposed web usage mining architecture

the weblog data from various sources, it needs to be pre-processed and presented to the web usage mining models. As described by Mobasher (2006), the following essential pre-processing stages will be used in this research work:

- Synchronization of weblog data from multiple services
- Data cleaning
- Page view identification
- User identification
- Session identification
- Episode identification

Synchronization refers the process of merging the weblog data from the listed services in study. The web logs from multiple servers may contain few redundant

data. This strep is very essential as it serves as a base for web usage mining. Data cleaning refers to removal of data not needed for mining. The number of data fields needs to be removed depends on the type of service. Page view identification is an aggregation model used to collect the web objects which contains the webpages visited by the user. User identification is the process of referring the user activity record which contains the logged activities of a same user. Sessionization is the process of grouping the user activity record into many sessions each represents a single visit to the particular service. Episode refers a subset of a page view that can be performed to ben more focused on the functional behaviour of a particular user. This is the last step in the data pro-processing.

Once the pre-processing is over, the data belong to a particular service is separated as a data mart. This is carried out in order to be more focused on a particular E-service. The user can select the E-service of his choice to proceed further. The data marts are subject-oriented. Once the data mart is ready, web usage mining models such as association rule mining and/or clustering techniques are applied in the data mart. This web usage mining models extracts the knowledge patterns from the data mart and presents it to the user/expert. The expert analysis this pattern and select the best recommendation to be adopted for the particular E-service.

RESULTS AND DISCUSSION

Web usage mining models: In order to achieve the defined objective, the following types of analysis are:

- Cluster analysis
- Association and correlation analysis

Clustering a data mining techniques which groups together all the relevant data having same characteristics. This is the most frequently used model in the web usage mining to identify the interesting group of user characteristics. This is used to study the navigation pattern of the users. The familiar clustering technique called K-means will be used in the proposed model initially. Efforts will also be made to use the Markov models with the K-means model to improve the cluster quality.

Association analysis in data mining is a yet another interesting model to find out the correlation between two items in a database. The familiar association rule mining technique Apriori will be used in this research work. The concept of frequent items sets (Fig. 3) which generates all the frequent item sets from the given web log files. Mining of association rules in web logs will bring out few interesting useful knowledge for providing better services to the end-users (Mobasher, 2006). The generated knowledge will act as a recommendation for the administrators.
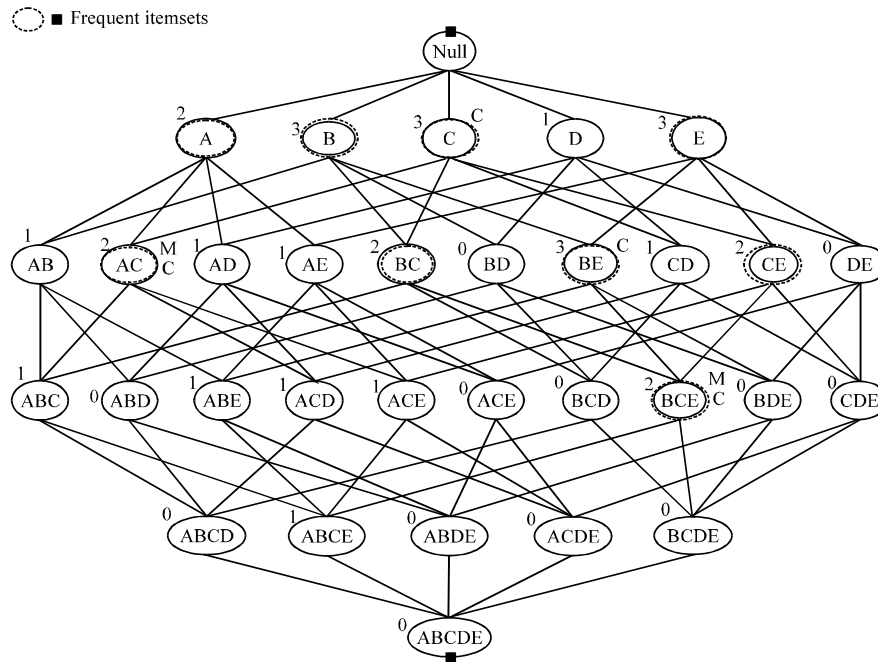

Fig. 3: Frequent itemset mining model

Table 1: The benefits of applying web-usage mining

Service	Expected benefits
Communication and technology	Increase in use of better value added services provided by Saudi post; providing easy way for DNS registration
Training and education and culture	Provide better academic monitoring of a student by universities for course registration; add/Drop of courses, maintenance of student record etc.
Housing and municipal services	Providing revised housing and municipal services; introducing a new licensing system to open a new shop to approve new systems for building plan approvals
Labour and employment	Provide flexible services for job seekers; apply for a work visa based automatically based on navigational history; provide customized summer training to the students
Transport	Provide a customized travel plan for an air passenger based on history; create and maintain new plan to improve the transport across the kingdom
Economy and business	Providing customizable business solution to an end user; better complaint monitoring across industry practices management of trademark and copyright
Social life	Providing more meaningful and appropriate services to the residents based upon the events happening in the family; automatic renewal of eligible modules upon the registration of births; automatic closure of module and settlements to the family upon the registration of a death
Islamic affairs	Providing more customizable travel plan during Haj and Umrah; automatic alert of a missing pilgrimage/passenger
Utilities	Providing a customizable water supply plan for a family; providing a customizable electricity supply plan for a family
Personal documents	Better monitoring of civil details like license,national id, insurance etc; provide a customizable VISA for a foreign worker based on the history
Insurance and pension	Automatic settlement to pensioners based on retirement date and other conditions; showing irregular patterns in case of deviation in deduction of contribution to pension plans
Travel and tourism	Provide a customizable EXIT RE-ENTRY VISA for a Foreign worker based on the history; Introduce a new service (train/bus) based on the history
Health and environment	Presenting the reason for a particular decease in a region or for a specific set of peoples; predicting the likelihood or impact of a particular decease spreads from other countries
Traffic and safety	Predicting the accident in a place by tracking a individual violator; automatic settlement to the victim by the insurance company

Expected beneits of using web usage mining: The benefits of applying web-usage mining for each services are described in Table 1.

CONCLUSION

As of now, the application of web usage mining in the E-government is very small. The impact of using the web usage mining on the E-government data as listed the study contain only few observation by the researchers. The detailed and expanded list will be made available in the future. Upon using the proposed web usage mining architecture and models researchers is confident that the benefits will surely create a very big impact on the society.

REFERENCES

Al-Mushayt, O.S., Y. Perwej and K. Haq, 2012. Electronic-government in Saudi Arabia: A positive revolution in the peninsula. Int. Trans. Appl. Sci., 1: 87-98.

Alateyah, S.A., R.M. Crowder and G.W. Bills, 2013. Factors affecting the citizen's intention to adopt e-government in Saudi Arabi Int. J. Soc. Manage. Econ. Bus. Eng., 7: 1287-1292.

Alshehri, M. and S. Drew, 2010. Challenges of e-government services adoption in Saudi Arabia from an e-ready citizen perspective. World Acad. Sci. Eng. Technol., 66: 1053-1059.

Basamh, S.S., H.A. Qudaih and M.A. Suhaimi, 2014. E-government implementation in the Kingdom of Saudi Arabia: An exploratory study on current practices, obstacles and challenges. Int. J. Humanities Soc. Sci., 4: 296-300.

Borges, J. and M. Levene, 2005. Generating Dynamic Higher-Order Markov Models in Web Usage Mining. In: Knowledge Discovery in Databases: PKDD 2005. Jorge, A. L. Torgo, P. Brazdil, R. Camacho and J. Gama (Eds.). Springer Berlin Heidelberg, Berlin, Germany, ISBN: 978-3-540-31665-7, pp: 34-45.

Kamber, M. and J.W. Han, 2006. Data Mining: Concepts and Techniques. 2nd Edn., Elsevier Science Inc., Philadelphia, PA., USA., ISBN-13: 9781558609013, pp: 252-272.

Mobasher, B., 2006. Web usage mining web data mining: Exploring hyperlinks, contents and usage data. Chapter, 12: 449-483.

Poongothai, K., M. Parimala and D.S. Sathiyabama, 2011. Efficient web usage mining with clustering. IJCSI. Int. J. Comput. Sci., 8: 203-209.

Qwaider, W.Q. and K.A. Al-Shafi, 2013. Understanding factors influencing e-government in Saudi Arabia. Int. J. Comput. Appl., 81: 37-41.

Sain, N. and S. Tamrakar, 2012. Web usage mining and pre-fetching based on hidden markov model and fuzzy clustering. Int. J. Comput. Sci. Inf. Technol., 3: 4874-4877.