EARTH OBSERVATION AND GEOMATICS ENGINEERING
website: https://eoge.ut.ac.ir

A new approach based on the RFS method for evaluating popularity of top open-source GIS software packages using user comments in forums

Mostafa Ghodousi, Abolghasem Sadeghi-Niaraki*, Bahram Saeidian

Department of GIS, Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran, Iran

Article history:
Received: 15 January 2017, Received in revised form: 15 April 2017, Accepted: 15 May 2017

ABSTRACT

Open-source software packages have various functions. An important factor is choosing the right package. On the other hand, no organization is responsible for these packages. In this regard, using the users' experience can be a good option or method in selecting the right package. There are many online forums which users discuss their experiences about the open-source software packages. Analyzing the data from these forums can help users to select the appropriate packages. This paper evaluates three top open-source software packages, namely QGIS, GRASS, and gvSIG, based on online forums covering spatial issues. In addition, the paper compares software packages and forums based on users' behaviors through a new approach, namely a Recency-Frequency-Satisfaction (RFS) method based on Recency-Frequency-Monetary (RFM) in Customer relationship management (CRM). Finally, the paper analyzes the procedure for using software packages by users' comments by year. The results show that QGIS was used more than the other two and that the procedure was in ascending order for years.

KEYWORDS

GIS software
Open source
Software selection
RFS
Forums

1. Introduction

The growth of Free and Open-Source Software (FOSS) packages has gone through huge development recently (Steiniger & Bocher, 2009). The popularity and dependence of FOSS has increased (Chen et al, 2010). Being dependent on the vendor, or the so called "vendor-lock", is an obvious flaw of the closed-source code. So the user cannot make its basic function different by extending the program. There exist some prominent advantages of open-source software, such as vendor independence, cost savings, and open standards. The main significance of open-source software is, according to its free access, distribution, publishing, and use and the ability to modify it (Istvan, 2012). The growth of this type of software has caught attention from the GIS community regarding the advantages of open-source software. Due to the high costs and the license requirements for the specific tools, the commercial GIS software imposes some other restrictions for more studies. On the other hand, the FOSS GIS tools can equally solve the problems in a similar way without any cost, license, and access restrictions for the full source code, but the tools are recently-developed in general (Wikipedia, 2015). Free and open-source GIS software (FOSSS4G) is known to be a reliable choice for lots of users, especially for the people who are in research

* Corresponding author
E-mail addresses: mostafaghodosi1@gmail.com (M. Ghodousi); a.sadeghi@kntu.ac.ir (A. Sadeghi-Niaraki); bahram_saeidian@yahoo.com (B. Saeidian)

DOI: 10.22059/eoge.2017.237260.1009

institutions, nonprofit organizations, and small businesses. The reason is that limited funds will not afford the cost of installation and maintenance of the commercial software. The development of the FOSS4G community gives evidence on its accomplishment to satisfy the needs and demands of the users (Mohammed, 2014). Since there are a lot of different open-source software packages, it is cumbersome to detect the most suitable alternative for a specific need. Selecting the appropriate GIS software package is important for the investment accomplishment (Eldrandaly, 2007). Choosing the most suitable GIS software package for a specific GIS project is not a well-defined or structured decision problem and hence, solving this problem, needs careful examination of a general set of factors and equalization of many objectives to detect the appropriateness of a specific software package for making a defined GIS application (Eldrandaly, 2007). For this reason, this paper develops a new and easier model for evaluating the open-source software packages that reflects simplicity and integrity. The other problem concerning open-source software is that it does not have a trustee, so there is not any organization or individual responsible for the software operation. One way to appreciate the open-source software packages and launch issues about the possible software problems in projects is using online forums such that the other users can find the answers to their questions in the responses given by the other users. In addressing these problems, such issues as limited knowledge could be argued in these online forums (Spinellis & Giannikas, 2012). About different topics ranging from products, software packages and services to politics and world events, users talk and exchange their ideas and experiences (Dellarocas, 2006). There are several characteristics of the information from online forums that make them a unique platform for comparing software packages. The credibility of information taken from online forum is more than what the software companies and developers offer, this exchanged information is more related to the users. This information is very precious and a lot of data on software could be gathered by users through them which includes: data on satisfaction of a given software package among target users, data on desirable or undesirable features for future software versions, data on the level of popularity, prices (if a software package is commercial) and data on changes in users' attitudes over time (Pitta & Fowler, 2005). Software forums possess a wide range of knowledge for exchanging ideas and solving the problems which were posed by different users and developers. Therefore, mining such contents may be ideal and considerable (Gottipati et al., 2011). This study uses useful information from the forums by the importance it gives to the selection of the suitable open-source GIS software packages. To do so, this study uses some open-source software packages from eight mature open-source software packages for the desktop said in the forum, namely (1) GRASS GIS, (2) Quantum GIS (3) ILWIS/ILWIS Open, (4) uDig, (5) SAGA, (6) OpenJUMP, (7) MapWindow GIS, and (8) gvSIG (Steiniger & Hay, 2009). In this paper, the capability of the most widely used open-source GIS software packages from the software that is mentioned by using questions and answers in the most important GIS forums is assessed. In addition, three software packages in forums are compared, and the capacity of online forums is evaluated using the RFM model. The RFM analysis is a marketing technique to analyze quantitatively the best customers by examining the last time they have purchased (recency), the frequency they purchase (frequency), and the money they spend (monetary) (Techtarget, 2015). This study is the first to evaluate software in forums based on the RFM model. Here an innovative approach using RFM, a CRM method, is taken, and the factors are changed to alter RFM to RFS. The RFS method possesses a lot of benefits, including examination of users' behaviors and easily make these behaviors quantified and rank the software packages (Miglautsch, 2000), predicting users' responses and increasing the profitability of software companies and providing important insight for the developer to solve the problems in daily activities in order to develop effective strategies to satisfy a wider range of users' needs in the short term, summarizing the popularity of the software package by using a limited number of variables (Wei t al., 2010). The reason for using RFS is to help future users of software packages select the most suitable packages based on forum users' opinions on GIS software packages. Therefore, users do not have to conduct their own surveys and random tests of packages to identify the suitable ones. The RFS model estimates GIS software from the user's view. This paper evaluates software packages by using data from forums between 2013 and 2014, but the using trend evaluation has been done until 2013. Four important problems should be addressed to better compare software's popularity and performance. The first is the initial assessment to select the software and forums. The second is the development of a model to compare the software. The third is the evaluation and comparison of the software based on the developed model. Finally, the forth is to study the popularity temporal trend. The rest of this paper is organized as follows: Section 2 provides a review of previous research on open-source GIS software packages, software evaluations, forums, and the RFM model. Section 3 discusses the methodology. Section 4 provides an initial comparison to select three open-source software packages and some forums based on their users number and online popularity and presents the features of some expert GIS forums and public forums covering GIS software. Section 5 describes the proposed model. Section 6 assesses the forums and software packages in two stages: the first stage for forums on three software packages, and the second stage for three software packages. Section 7 discusses the procedure for determining the number of questions raised

by users at one-year and six-month intervals to investigate the software development procedure and popularity of new versions of software packages. Section 8 concludes the limitations of this paper and suggestions for future research.

2. Background

Previous studies of open-source GIS software packages have identified mature GIS software packages, and many studies have discussed and compared FOSS4G software packages for particular applications such as ecology, education, and hydrology. Steiniger and Hay (2009) introduced free and open-source GIS software for landscape ecologist and surveyed the available tools, identifying eight mature desktop GIS packages: (1) GRASS GIS, (2) Quantum GIS, (3) ILWIS/ILWIS Open, (4) uDig, (5) SAGA, (6) OpenJUMP, (7) MapWindow GIS, and (8) gvSIG (Steiniger & Hay, 2009). This paper investigates these eight software packages as main software packages. Ramsey (2007) offered a yearly overview on open-source GIS packages and argued that "open source (GIS) software can give us a feature-complete alternative to proprietary software in most system designs" (Ramsey, 2007). Steiniger and Bocher (2009) argued about the state of free desktop GIS package stand and drew their usefulness for the GIS science research compared with the propriety software (Steiniger & Bocher, 2009). Steiniger and Hunter (2013) analyzed the most important factors affecting the choice of open-source software packages in business, also the studies pave the way for the coming developments (Steiniger & Hunter, 2013). Istvan analyzed the most famous open-source GIS systems and geographic information tools which could be applied to landscape ecology. They offer a comparative analysis of the most used open-source GIS application (Istvan, 2012). Donnelly (2010) made a comparison among many open-source desktop GIS packages to the propriety ArcGIS software package through examining their capability to make thematic maps in a library environment to argue each FOSS GIS package had its own specific pros and cons and to point out that no single package could adapt with all the functionalities (Donnelly, 2010). Steiniger and Hunter (2010) evaluated open-source software packages for GIS teaching and claimed that open-source software packages have grown to a point in which they can replace the propriety desktop GIS software regarding the use of open-source GIS software packages for teaching (Steiniger & Hunter, 2010). Hengl et al. (2009) examined the performance of two GIS software packages (GRASS and SAGA) to review the elevation data through concentrating on DEM generation, extracting hydrological features (stream networks), and extracting gridded DEM derivatives (Heng et al., 2009). Chen et al. (2009) evaluated open-source GIS software packages for water resource management in the developing countries (Chen et al., 2009). Singhai and Saxena assessed FOSS GIS tools to extract and classify drainage network to analyze how reliable and general open-source GIS software packages are and to demonstrate how capable open-source geospatial tools are in drainage network extraction (Singhai & Saxena, 2012). Previous studies evaluating GIS software packages have employed one or more of the following methods: the AHP (Analytical Hierarchy Process), GQM (Gold-Question-Metric), and ISO 9126. Eldrandaly (2007) used the MCDM technique and the AHP to assist system developers to choose the most suitable GIS software package for a specific application (Eldrandaly, 2007). Al-Qutaish et al. (2009) integrated the AHP with ISO 9126 and employed them for choosing the open-source software products (Al-Qutaish et al., 2009). Eldrandaly and Naguib (2013) suggested a smart decision support system which integrated the Expert System (ES) and the MCDM (AHP) and offered an advisory system to help the system developers during GIS software procedures for justifying and choosing (Eldrandaly & Naguib, 2013). Dobesova's approach to assess GIS programs in cartography, CartoEvaluation, was according to the Gold-Question-Metric method and was for assessing some desktop GIS software packages (Dobesova, 2013). Dobesova and Kusendová (2009) assessed the functionality of cartography through using the Goal-Question-Metric (GQM) method for evaluating the software packages for several purposes (Dobešová & Kusendová, 2009). Wawer et al. (2008) assessed the open-source GIS and RS software packages for the applications in the environment for the geospatial end-users through integration of GQM technique with ISO 9126 (Wawer et al., 2008). The aforementioned studies paid little attention to the role of users. One important issue to be considered is the importance of this role in assessing software packages. If it is possible to evaluate the advantages and disadvantages of software packages based on an extensive range of users, then assessments and comparisons can be better produced for some software packages. Software users are end users, which include an extensive range from expert users to general users who use software packages for various projects. Therefore, different parameters and capacities should be considered in any comparison of parameters by analyzing users' comments in forums. For this aim, this paper suggests an RFS model offering simple framework to make the user's behaviors in forums quantified (Birant, 2010). This study, more specifically, uses a CRM method to talk about the users' needs and preferences and increase the ability of users in the GIS software community (Thompson & Sims, 2002). As far as the authors are concerned, this paper is the pioneer in analyzing the assessment of software packages in forums according to the RFM model. However, the previous studies have reviewed online forums and their content. Krauss et al. Introduced a web-mining approach which mixes the social network examination and automatic sentiment analysis method, and to examine their approach, they carried out two

experiments to analyze the talks in the online forums on the Internet Movie Database (IMDb) through analyzing the correlation between the social network structure and the external metrics such as box office revenues and Oscar Awards (Krauss et al., 2008). Woo et al. (2013) gathered and examined the content of online forums through employing web-, text-, and data-mining techniques and had a proposal for the foresight support system for the medical industry by a better diffusion model for online medical forums (Woo et al., 2013). Hung et al. (2013) used data-mining techniques including statistical analyses, clustering, association rules, and sequential pattern discovery to the mining of web information from log data to analyze the patterns of self-care behavior in elder adults (Hung et al., 2013). The RFM model measures when people purchase, the frequency they purchase, and how much they purchase. Past purchases of customers can anticipate their coming purchase behaviors (Colombo & Jiang, 1999). The mix of data mining and RFM can result in efficient information (Wei et al., 2010). Hsieh suggested an integrated data-mining and behavior-scoring model to anticipate the profitable customers based on the repayment behaviors and RFM-based behavior-scoring predictors (Hsieh, 2004). Cheng and Chen had a proposal for a procedure that mixes RFM attributes and K-means algorithm into a rough set theory to draw meaningful rules (Cheng & Chen, 2009). Cho and moon proposed a weighted mining frequent pattern according to the customer's RFM score for personalized u-commerce recommendation system (Cho & moon, 2013). The RFM model has been widely applied in many practical areas, including government agencies and services (King, 2007; Akhondzadeh-Noughabi et al., 2013), nonprofit and financial organizations (Sohrabi & Khanlari, 2007), marketing companies (Chang & Tsai, 2011), and ubiquitous-commerce (Cho et al., 2012 ; Cho & Moon, 2013), but no study has compared the software packages.

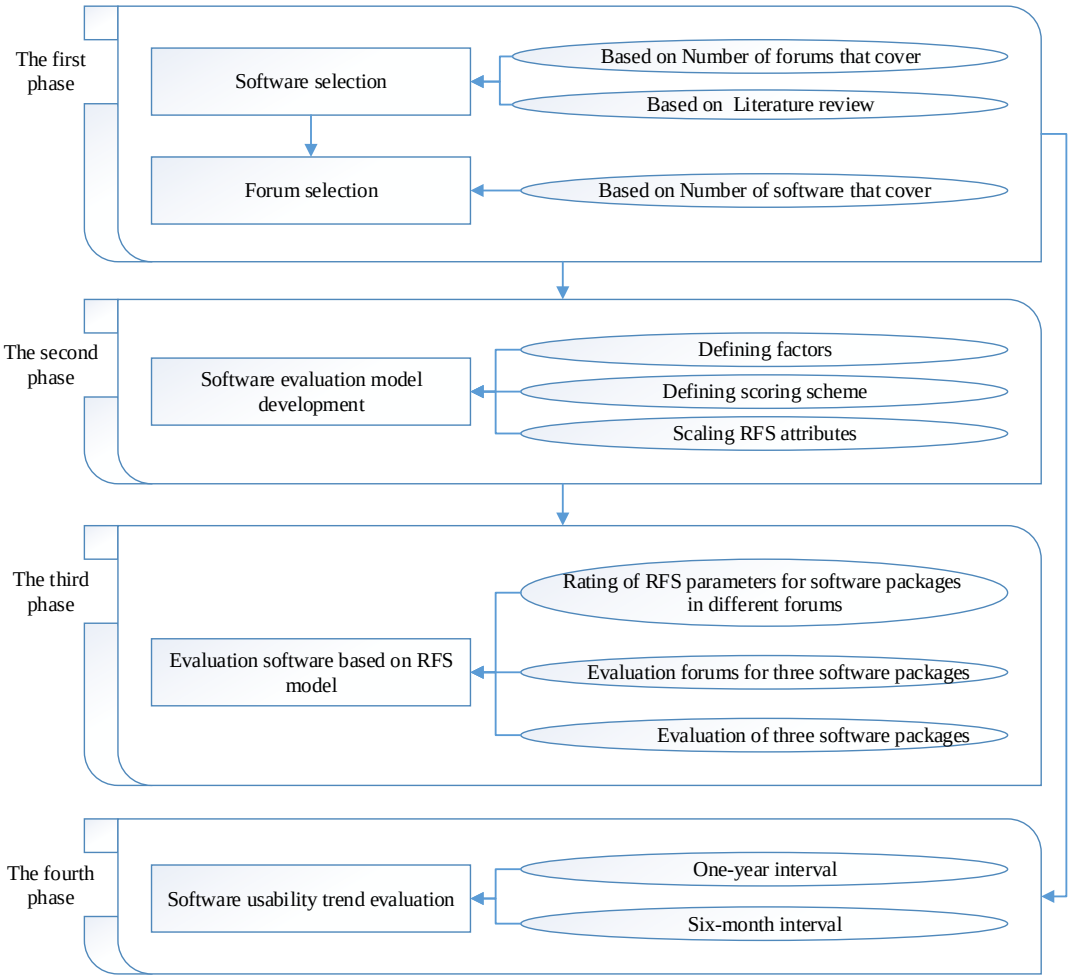


Figure 1. The research method

3. Methodology

Figure 1 shows a general flow chart of this study. As shown in Figure 1, the study was conducted in four phases. In the first phase, mature software packages were selected based on a literature review, the most widely used software packages among the mature software packages were identified based on the number of forums covering them, and some forums were selected based on the number of their users numbers and online popularity. In the second phase, the software evaluation model (RFS) was developed and model soring scheme was defined. In the third phase, the capability of the selected software packages was assessed and compared using the RFS. In the final phase, the usability of the software packages was evaluated over time by using one-year and six-month interval data to investigate the software development procedure and the popularity of the new versions in comparison to the previous versions and the advantages and disadvantages of different versions of software packages.

4. Initial assessment to select software and forums

This section provides an initial comparison of software packages to select three open-source software packages from the eight desktop open-source software packages (Steiniger & Hay, 2009) based on the number of forums covering them and some forums selected based on their user numbers and online popularity. Table 1 shows internet addresses of some important GIS forums and software packages covered by them. As shown in the figure, QGIS, GRASS, and gvSIG were covered more than the other packages, and therefore these were selected for a comparison based on the question-and-answer method. In addition, stackexchange and OSgeo were popular forums and were the only ones to cover all these software packages, and therefore these two forums were selected to evaluate software packages (Section 6) and examine the number of questions by users concerning these packages at one-year intervals (Section 7).

5. Model development

The research proposes a model according to the RFM (Recency, Frequency, and Monetary) model, in which its application has been applied to many practical areas, especially to direct marketing. The valuable customers could be detected and the marketing strategies could be enhanced by the decision makers through choosing the RFM model (Wei et al., 2010). In this paper, software projects were rated according to the behavior of users in the online forum instead of rating the customers, and then the anticipation is regarded in online forums (databases) based on the users' behaviors (Yeh et al., 2009). For this reason, the RFM model was used. Some studies quantified the behavior of users such as number of downloads and user rating in Google Play store and used that for evaluating the software quality (Taba et al., 2014; Tian et al., 2015), but they didn't develop a structured model. Rating software packages based on behaviors and attitudes of users in forums is not dependent on financial issues, and instead, users' satisfaction and active participation in the subjects related to software packages may be used. When we concentrate on the needs of people in segment, the RFM model should be changed to the RFS model, with respect to the citizen relationship management. Financial matters are not related, so the satisfaction of citizens could be employed instead of the financial factors (Akhondzadeh-Noughabi et al., 2013 ; Ghodousi et al., 2016). In this regard, the RFS model, where S implies users' satisfaction and active participation in online forums, was developed. The RFS model has the following factors:

1- Frequency (F): The number of users' questions about software in 2014.

Table 1. Internet addresses of important GIS forums and software packages covered by them

Forum	Software packages covered by forums								Internet addresses
	QGIS	GRASS	gvSIG	MapWindow	ILWLS	SAGA	uDig	Open-JUMP	
stackexchange	●	●	●	●	●	●	●	●	www.gis.stackexchange.com
cyburbia	●		●						www.cyburbia.org/forums
OSgeo	●	●	●	●	●	●	●	●	www.OSgeo-org.1560.x6.nabble.com
malaysiagis		●							www.malaysiagis.com/forum
forum.qgis	●	●			●		●	●	www.forum.qgis.org
qgisforum	●	●							www.qgisforum.org/forum
forum.quantum-gis	●	●	●	●		●			http://forum.quantum-gis.pl/ http://forum.grass-gis.pl/

2- Time interval or "Recency (R)": The time interval between the first question on a software package and the last one in 2014.

3- User satisfaction (S): This implies users' satisfaction and active participation. To measure these two factors,"the average number of responses to e ch question" and "the average number of observations per question" were used.

A one-year (2014) period was used to rank software packages because some software packages had high frequency scores as a result of their long development history, whereas new the ones were not as popular among users. These three variables are for behavioral variables for the popularity of the software package and they are employed to separate variables by watching the views of users (customers) toward software packages according to the talks in online forums (databases) (Wei et al, 2010).

5.1 Definition and scoring scheme of the model

This section discusses how software packages were scored for a better evaluation of those packages. To begin, the previous studies containing scores based on the RFM model were reviewed, and then a scoring scheme was selected based on the review. Miglautsch (2000) suggested an RFM-based scoring method, named the customer quintile method, which sets customers in the values with descending order. This method has the advantage of quantifying customers and yielding equal numbers of customers in each group, but it encounters some scoring challenges in the area of Frequency and is relatively sensitive. (Miglautsch, 2000). Some papers have this suggestion that the mix of RFM could be gotten through assigning or based on comparing the average R (F, M) value of a cluster with the overall averages (Sohrabi & Khanlari, 2007). Based on the sayings of Hughes (1994), the RFM measure possesses the same weight when a composite score is calculated. For example, the composite score occupying cell (4, 1, and 2) is 7 (4+1+2) (Hughes, 1994). Tsai and Chiu (2004) used WR, WF and WM for presenting the importance of R, F and M criteria. They discussed that the sum of weights for each RFM measure should be equal to 1 (Tsai & Chiu, 2004). Liu and Shih (2005) for detecting the relative weights of RFM variables, used the analytic hierarchy process (Liu & Shih, 2005). Wu et al. (2010) proposed fuzzy weighted RFM which shows that allocating different weights bring about more flexibility (Wu et al., 2010). In this study, the composite value of RFS was obtained by multiplying the normalized RFS values for each software package and the weight of RFS variables in each forum. All RFM variables had an equal weight of 1/3 such that their sum was equal to 1. In addition, S in RFM consisted of two parts (the average number of responses to each question and the average number of observations per question) such that each constituted half the weight of S. Then, the scaling of RFS attributes was applied, as shown in Table 2. The scaling is considered based on the distribution of values.

6. Evaluation of open-source software packages based on RFS

This section statistically analyzes the software packages by considering calculated values for the RFS parameters and investigates and compares forums and selected software packages in two sections, including the assessment of forums regarding three software packages, and the general assessment of these packages. Table 3 takes into account QGIS, GRASS, and gvSIG in two forums and the three RFS parameters, which were used to obtain values for the overall evaluation of the three software packages, and also the assessment forum based on these software packages. Table 3 shows the number of questions and answers and that of the observations of each question for each forum. According to the results, the number of questions and the average number of answers to questions in OSgeo exceeded. The users of OSgeo were mainly software developers, and this may explain why the number of answers to each question increased and that of the observations of each question decreased (because general users were more willing to observe general questions about the software). According to Table 3, QGIS was better questioned by users, followed by GRASS and gvSIG, in that order. This indicates that QGIS was used more frequently than the other two packages. GRASS was discussed more frequently in OSgeo than in other forums, indicating its popularity among the expert users of software packages. However, there was no significant difference in QGIS. One major reason for QGIS's popularity among general users was its powerful graphic interface, which exceeded that of GRASS. One major reason for GRASS popularity among the expert users was the 3D capacity of GRASS exceeded that of QGIS. The importance of QGIS was evidenced by the fact that about 37% of questions about GRASS were related to QGIS (e.g., GRASS plugins in QGIS). The second largest number of questions was related to rasters, indicating that most users employed GRASS in raster operations, which is not inconsistent with expectations. To statistically evaluate the software packages, the three RFS parameters from Tables 3 were considered by assigning each, a value from 0 to 10 based on Table 4.

6.1 Evaluation forums for three software packages

Users may choose their software packages and then try to find forums that actively discuss their chosen packages. This section examines this topic. The coefficients for three software packages were calculated for three forums. The total score for each forum for QGIS is shown in column 2 of Table 5; that for GRASS, in column 3; and that for gvSIG, in column 4.

Table 2. Real scaling of RFS attributes in forums

Scaling	Scaling name	R - Recency	F - Frequency	S - Satisfaction	
				Average number of responses	Average number of observations
10	Extremely high	360-365	Over 1000	Over3	Over190
9	Very, very high	350-360	800-1000	2-3	180-190
8	Very high	335-350	500-800	1-2	170-180
7	High	315-335	300-500	0.970-1.000	160-170
6	Moderately high	290-315	100-300	0.940-0.970	150-160
5	Moderate	250-290	60-100	0.900-0.940	100-150
4	Moderately low	180-250	40-60	0.600-0.900	50-100
3	Low	110-180	20-40	0.400-0.600	30-50
2	Very low	50-110	5-20	0.200-0.400	15-30
1	Very, very low	0-50	1-5	0.100-0.200	1-15
0	Extremely low	0	0	0.000-.0100	0

Table 3. Ratings of RFS parameters for software packages in different forums

Forum	**R - Recency**	**F - Frequency**	S - Satisfaction	
			Average number of responses	**Average number of observations**
	QGIS			
stackexchange	365	3197	0.974	178.783
OSgeo	365	1613	2.122	38.579
	gvSIG			
stackexchange	138	5	0.400	197.400
OSgeo	355	64	1.890	14.547
	GRASS			
stackexchange	357	277	0.946	154.563
OSgeo	364	511	3.004	19.066

Table 4. Software characteristics

Forum	Software characteristics		
	QGIS	GRASS	gvSIG
	F (frequency)		
stackexchange	10	6	1
OSgeo	10	7	5
	R (recency)		
stackexchange	10	10	3
OSgeo	10	9	9
	S1 (average number of responses)		
stackexchange	7	6	3
OSgeo	9	10	8
	S2 (average number of observations)		
stackexchange	8	6	10
OSgeo	3	2	1

Table 5. Total scores for three parameters

Forum	QGIS (30)	GRASS (30)	gvSIG (30)
stackexchange	27.5	22	10.5
OSgeo	26	22	18.5

Table 6. Total score for each software package

Software	Total score (30)
QGIS	26.75
GRASS	22
gvSIG	14.5

Here the weight of all coefficients (RFS factors) is equal. Total scores for each forum for three software packages are derived by Eq. (1):

$$T_{ij}=F_{ij}+R_{ij}+\frac{1}{2}\times\left(S1_{ij}+S2_{ij}\right) \quad (1)$$

In which, T is the total scores for each forum for each software (based on the values of column 2-4 in Table 4, F is the frequency, R is the recency, $S1$ is the average number of responses, $S2$ is the average number of observations, i is the number of forums and j is the number of software packages. The results in Table 5 show that the stackexchange had a higher score for QGIS. OSgeo had higher scores for gvSIG and the results of GRASS for two forums are the same.

6.2 Evaluation of three software packages

This section evaluates the software packages across two forums for three RFS parameters. For this, the total score for QGIS is shown in row 2 of Table 6; that for GRASS, in row 3; and that for gvSIG, in row 4. Total score for each software package (S) is derived from Eq. (2).

$$S_j=\frac{1}{2}\times\sum_{i=1}^{2}\left(F_{ij}+R_{ij}+\frac{1}{2}\times(S1_{ij}+S2_{ij})\right) \quad (2)$$

In which, S is the total scores for each software package (based on the values of column 2-4 in Table 4), F is the frequency, R is the recency, $S1$ is the average number of responses, $S2$ is the average number of observations, i is the number of forums and j is the number of software packages. The results in Table 6 indicate that QGIS had the highest score based on users, followed by GRASS. QGIS was significantly more likely to be used, implying its superiority over the other open-source GIS software packages.

7. Numbers of questions raised by users about three software packages at different intervals

This section discusses the procedure for determining the number of questions raised by users at one-year and six-month intervals to investigate the software development procedure and popularity of new versions of the software packages in comparison to the existing ones and evaluate the advantages and disadvantages of different versions of software packages. Figure 2 shows the number of questions in the stackexchange at six-month intervals. As shown in Figure 2, GRASS and QGIS showed a generally upward trend, whereas gvSIG showed a fluctuating trend. By mid-2011, the number of questions about QGIS exceeded that about GRASS. The number of questions about gvSIG after 2010 was below that for the other two software packages. The release of gvSIG version 1.9 in 2010 reduced the number of questions, but the release of version 1.10 in 2011 again increased the number of questions. The release of version 1.12 in 2012 again increased the number of questions, and that of version 2.0 in the first half of 2013 slightly reduced the number of questions. For GRASS, the release of version 6.4.0 in the first half of 2010 increased the number of questions and that of version 6.4.1 in the second half of 2011 reduced this number. However, the release of version 6.4.2 in 2012 again increased the number of questions. Figure 3 shows the results for the stackexchange in terms of the number of questions at one-year intervals. As shown in Figure 3, all three software packages showed an ascending trend. However, QGIS had the sharpest slope, followed by GRASS and gvSIG, in that order. This implies that users were more inclined to use QGIS. From mid-2011, the number of questions about QGIS exceeded that about GRASS. Figure 4 shows the number of questions at six-month intervals for OSgeo. As shown in Figure 4 by the GRASS diagram, the release of version 4.1 and subsequent updates of the version increased the number of questions. Between 1995 and 1997, the number of questions decreased because of some uncertainty about the software package (about the coverage and development organization). After 1997, however, the international GRASS development group started to manage relevant documents, versions, and codes, increasing the number of questions. In addition, the establishment of a website to recruit developers from various places was effective. Between 1997 and 2006, the number of questions increased except for the first half of 2004, when version 5.3 was released.

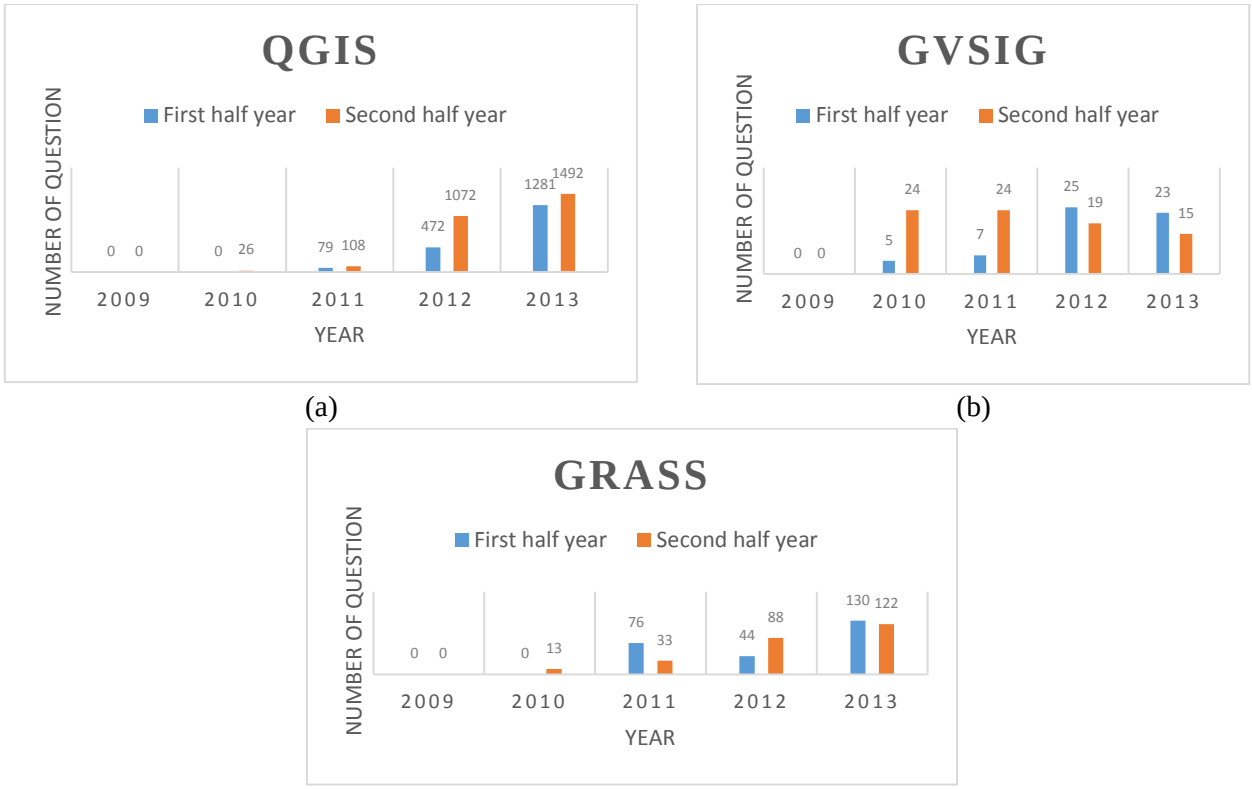


Figure 2. Numbers of questions at six-month intervals in the stackexchange forum ((a) QGIS, (b) gvSIG, and (c) GRASS))

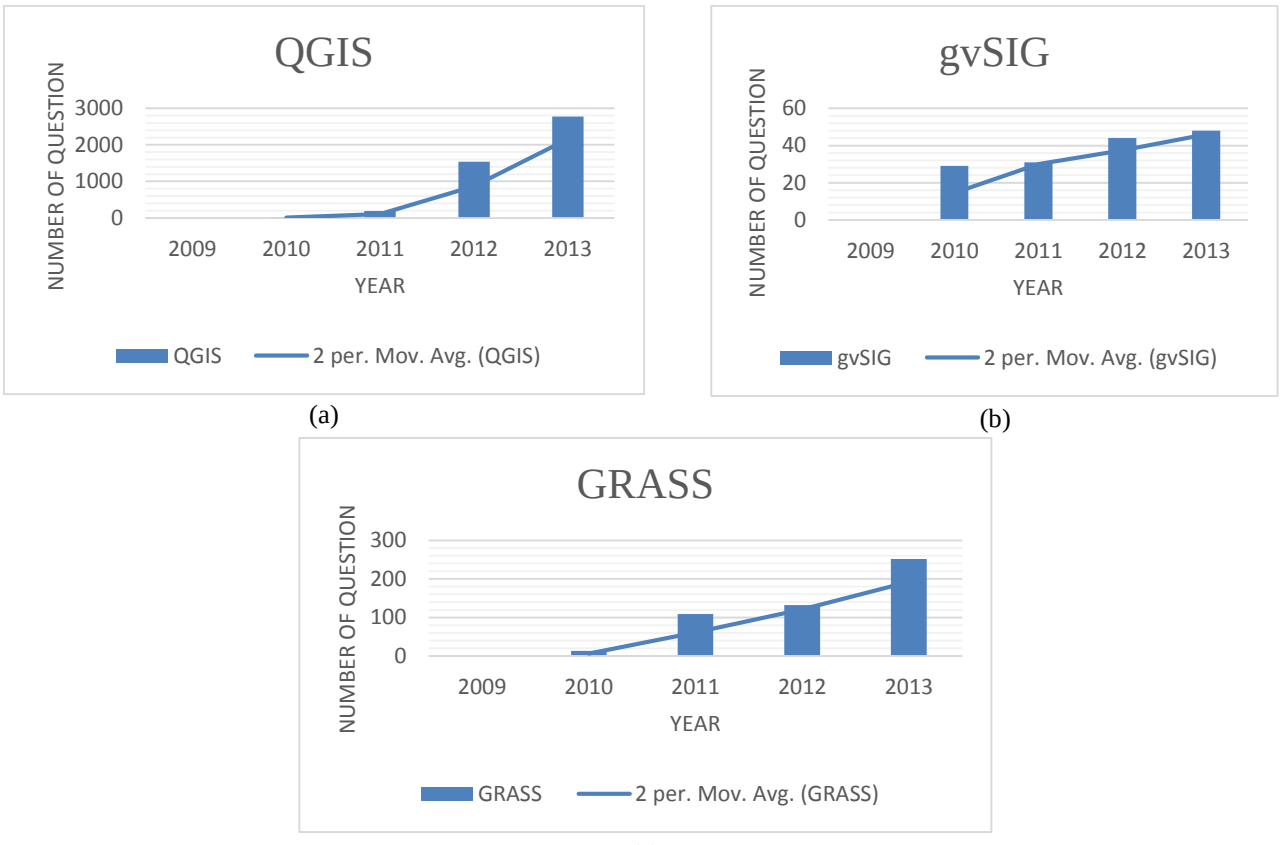
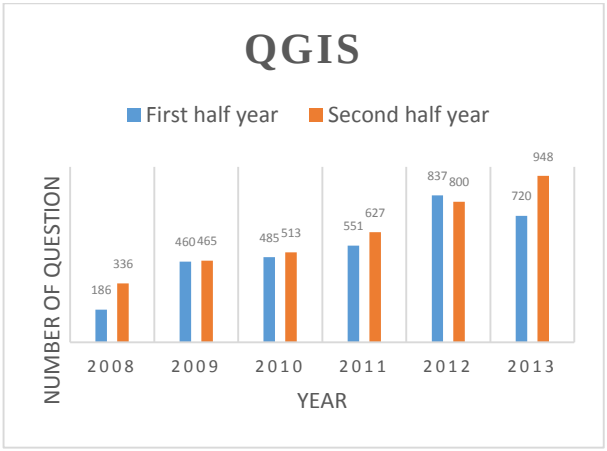
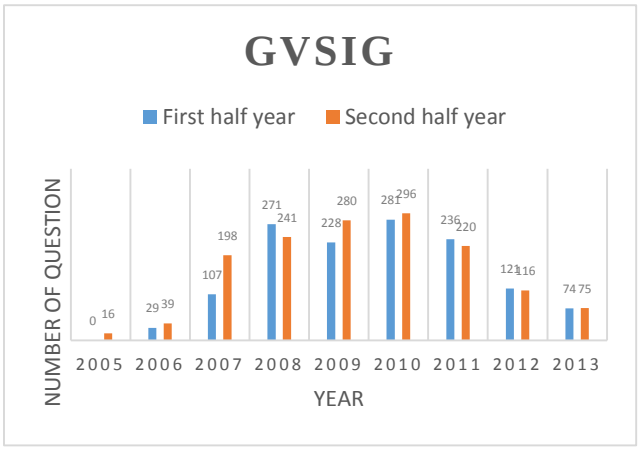


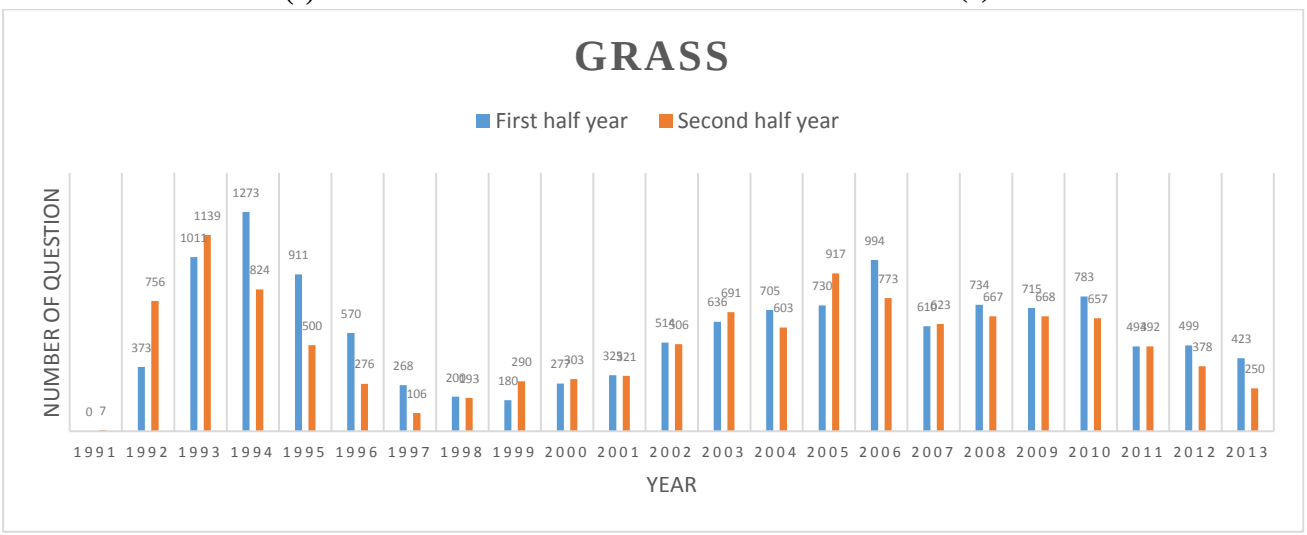
Figure 3. Numbers of questions at one-year intervals in the stackexchange forum ((a) QGIS, (b) gvSIG, and (c) GRASS)



(a)



(b)



(c)

Figure 4. Numbers of questions at one-year intervals in the OSgeo forum ((a) QGIS, (b) gvSIG, and (c) GRASS)

Then, in the second half of 2004, the number of questions increased with the release of version 5.4.0, and in 2006, before OSgeo, the number of questions peaked. After 2006, however, when OSgeo was built to develop the open-source geospatial technology, the number of questions increased, which reflected the release of version 6.0.2. In the second half of 2006, the release of version 6.1.0 increased the number of questions, and since then, the number fluctuated with the release of different versions.

For the QGIS diagram, the number of questions showed consistent increases except for the second half of 2011, when version 1.7.0 was released which reduced the speed at which the number of questions increased. Since the second half of 2012, the number of questions increased with the release of version 1.8.0. For the gvSIG diagram, the number of questions increased until the first half of 2008, after which the number of questions remained stable until the first half of 2010, when version 1.9 was released, resulting in a decrease in the number of questions. In Figure 5 the QGIS diagram shows an ascending trend. The gvSIG diagram shows an ascending trend until 2010, after which there is a descending trend.

The GRASS diagram from 1991 to 1994 and from 1997 to 2006 shows an increasing trend, and that from 1994 to 1997 and from 2006 to 2013 shows a decreasing trend. After 2010, the number of QGIS-related questions exceeded that of the GRASS-related ones, which is consistent with the results in Figure 3. The number of gvSIG questions in Figure 5 was below that for the other two software packages across all years.

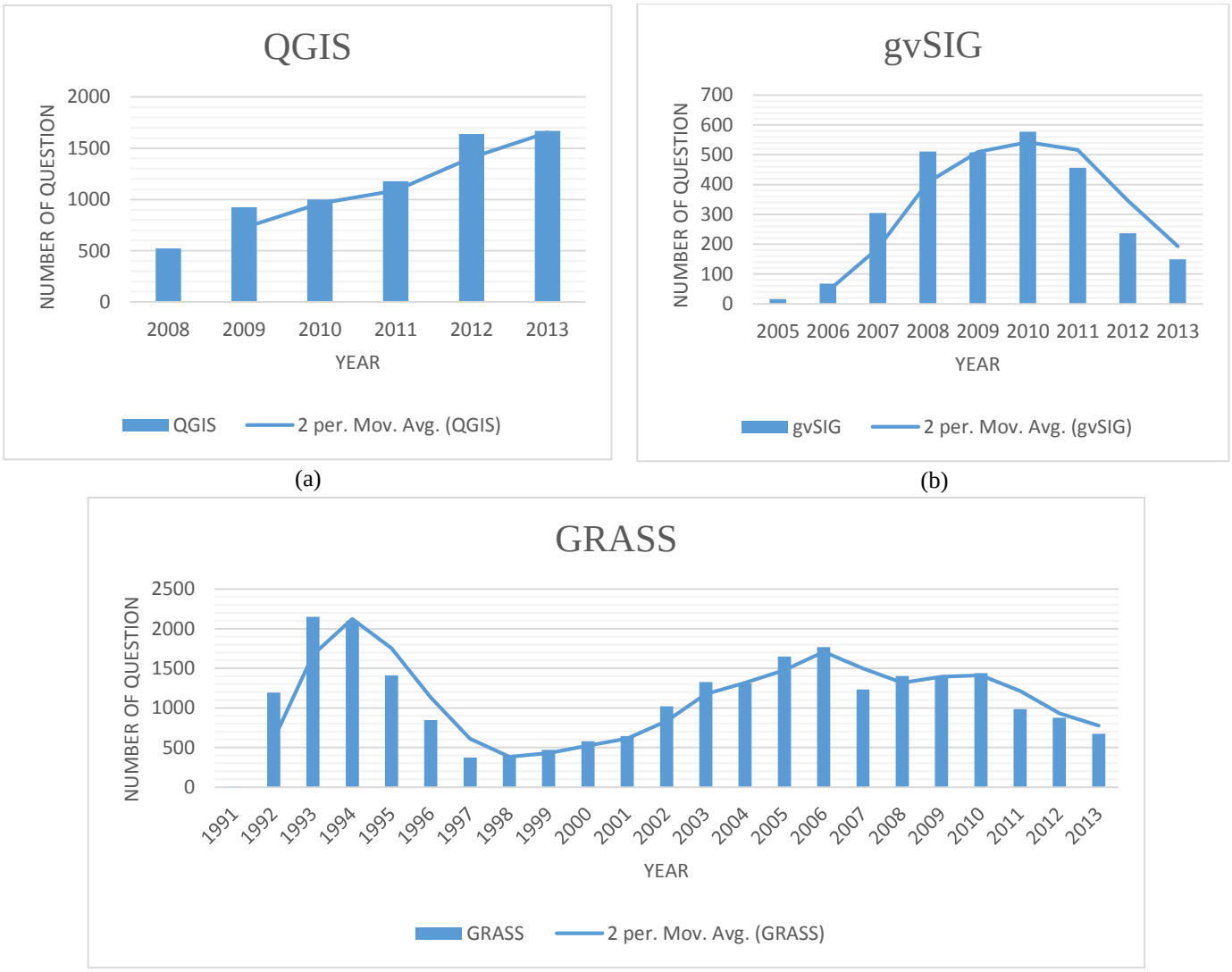


Figure 5. Numbers of questions at one-year intervals in the OSgeo forum ((a) QGIS, (b) gvSIG, and (c) GRASS)

8. Conclusions, Limitations and suggestions for future research

This paper examines three open-source software packages based on user comments in two forums. Based on the discussion, the features and capacity of QGIS were the main reasons behind its choice and popularity. This suggests that those who want to use this software package professionally and communicate with software developers should be aware of the OSgeo project and make use of it. This paper proposes a new RFM method for comparing and evaluating the performance of top open-source GIS software packages based on data from online forums. The results provide several reasons why the proposed model is suitable for ranking software packages. Given that the rating of software packages based on users' behaviors and attitudes in online forums is not strongly dependent on financial issues, users' satisfaction and active participation in subjects related to given software packages may be used instead of monetary factors, and it is for this reason that the present paper modifies the RFM factors to RFS, where S indicates user satisfaction. According to the results, software ratings based on the RFS method were 26.75 for QGIS, 22 for gvSIG, and 14.5 for GRASS. Figures 2 to 5 highlight the development of QGIS. In addition, the software produced improved versions over time, allowing the users to form good relationships with it. The results suggest that various companies that need to equip their systems and cut costs should use the software. In this paper, questions and answers by all members were considered to be equivalent for ratings, whereas in many other forums, members have rankings based on various factors such as the number of questions, the number of answers, forum hall membership, and the length of membership. However, members have different levels of expertise and experience based on their technical knowledge and work experience with respect to given software packages. In this regard, each factor can be given a point, and the investigation may be based on such a point-based system.

This paper assumes that the larger the number of questions for a given software package, the more popular is that package, but this assumption may be problematic if this issue shows high problems of the software. In this regard, future research should verify this assumption. Future research should integrate the RFS method with metaheuristic algorithms such as Bees, Genetic, PSO and other data mining methods to provide more useful insights for addressing the problems associated with software packages. Also integration of artificial neural network and RFS can be useful for predicting software popularity in future.

References

Akhondzadeh-Noughabi, E., Alizadeh, S., Ahmadvand, A. M., & Minaei-Bidgoli, B. (2013). FTiS: A new model for effective urban management: A case study of urban systems in Iran. Cities, 31, 394-403.

Anguix, A., & Carrión, G. (2005). gvSIG: Open Source Solutions in spatial technologies. GISPLANET, Estoril, Portugal.

Al-Qutaish, R. E., Muhairat, M. I., & Al-Kasasbeh, B. M. 2009. The analytical hierarchy process as a tool to select open source software. In Proceedings of the 8th WSEAS International Conference on Software Engineering, Parallel and Distributed systems, Cambridge, UK (pp. 39-44).

Birant, D. (2011). Data Mining Using RFM Analysis. In Knowledge-oriented applications in data mining. InTech.

Chang, H. C., & Tsai, H. P., 2011. Group RFM analysis as a novel framework to discover better customer consumption behavior. EXPERT SYST APPL. 38(12): 14499-14513.

Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert systems with applications, 36(3), 4176-4184.

Chen, D., Shams, S., Carmona-Moreno, C., & Leone, A. (2010). Assessment of open source GIS software for water resources management in developing countries. Journal of Hydro-environment Research, 4(3), 253-264.

Cho, Y. S., Moon, S. C., Noh, S. C., & Ryu, K. H. (2012, June). Implementation of Personalized recommendation System using k-means Clustering of Item Category based on RFM. In Management of Innovation and Technology (ICMIT), 2012 IEEE International Conference on (pp. 378-383). IEEE.

Cho, Y. S., & Moon, S. C. (2013). Weighted mining frequent pattern based customer's RFM score for personalized u-commerce recommendation system. Journal of Convergence, 4(4).

Colombo, R., & Jiang, W. (1999). A stochastic RFM model. Journal of Interactive Marketing, 13(3), 2-12.

Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. MANAGE SCI Science. 52(10): 1577-1593.

Dobesova, Z. (2013). CartoEvaluation method for assessment of GIS software. GEODESY CARTOGR. 39(4): 164-170.

Dobešová, Z., & Kusendová, D. (2009). Goal-Question-Metric method for evaluation of cartographic functionality in GIS software. In Proceedings Symposium GIS Ostrava (pp. 115-120).

Donnelly, F. P. (2010). Evaluating open source GIS for libraries. LIBR HI TECH. 28(1): 131-151.

Eldrandaly, K. 2007. GIS software selection: A multicriteria decision making approach. APPL GIS. 3(5): 1-17.

Eldrandaly, K., & Naguib, S. (2013). A knowledge-based system for GIS software selection. INT ARAB J INF TECHN. 10(2): 152-159.

Ghodousi, M., Alesheikh, A. A., & Saeidian, B. (2016). Analyzing public participant data to evaluate citizen satisfaction and to prioritize their needs via K-means, FCM and ICA. CITIES. 55: 70-81.

Gottipati, S., Lo, D., & Jiang, J. (2011). Finding relevant answers in software forums. In Proceedings of the 2011 26th IEEE/ACM International Conference on Automated Software Engineering (pp. 323-332). IEEE Computer Society.

Hengl, T., Grohmann, C. H., Bivand, R. S., Conrad, O., & Lobo, A. 2009. SAGA vs GRASS: a comparative analysis of the two open source desktop GIS for the automated analysis of elevation data. Proceedings of Geomorphometry.

Hsieh, N. C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. EXPERT SYST APPL. 27(4): 623-633.

Hughes, A. M. (1994). Strategic database marketing. McGraw-Hill Pub. Co.

Hung, Y. S., Chen, K. L. B., Yang, C. T., & Deng, G. F. (2013). Web usage mining for analysing elder self-care behavior patterns. EXPERT SYST APPL. 40(2): 775-783.

Istvan, S. (2012). Comparison of the most popular open-source GIS software in the field of landscape ecology. ACTA GEOGR DEBRECINA. Landscape & Environment Series. 6(2).

King, S. F. (2007). Citizens as customers: Exploring the future of CRM in UK local government. GOV INFORM Q. 24(1): 47-63.

Krauss, J., Nann, S., Simon, D., Gloor, P. A., & Fischbach, K. (2008). Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis. In ECIS (pp. 2026-2037).

Liu, D. R., & Shih, Y. Y. (2005). Hybrid approaches to product recommendation based on customer lifetime

value and purchase preferences. J SYST SOFTWARE. 77(2): 181-191.

Miglautsch, J. R. (2000). Thoughts on RFM scoring. J DATABASE MARKET. 8(1): 67-72.

Mohammed, D. W. E. (2014). Free and Open Source GIS: An Overview on the Recent Evolution of Projects, Standards and Communities, The 9th National GIS Symposium in Saudi Arabia, At Dammam, KSA.

Neteler, M., Bowman, M. H., Landa, M., & Metz, M. (2012). GRASS GIS: A multi-purpose open source GIS. ENVIRON MODELL SOFTW. 31: 124-130.

Olaya, V. (2008). SEXTANTE, a free platform for geospatial analysis. OSGeo J. 6: 32-39.

Pitta, D. A., & Fowler, D. (2005). Internet community forums: an untapped resource for consumer marketers. J CONSUM MARK. 22(5): 265-274.

Ramsey, P. (2007). The state of open source GIS. Refractions Research Inc.

Singhai A and Saxena A. (2012) An Evaluation of FOSS GIS Tools in Drainage Network Extraction and Classification using Cartosat DEM. INT J ADV SCI RES TECH. (2): 433-441.

Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. IRAN ACCOUNT AUDITING REV. 14(47): 7-20.

Spinellis, D., & Giannikas, V. (2012). Organizational adoption of open source software. J SYST SOFTWARE. 85(3): 666-682.

Steiniger, S., & Bocher, E. (2009). An overview on current free and open source desktop GIS developments. INT J GEOGR INF SCI. 23(10): 1345-1370.

Steiniger, S., & Hay, G. J. (2009). Free and open source geographic information tools for landscape ecology. ECOL INFORM. 4(4): 183-195.

Steiniger, S., & Hunter, A. J. S. (2010). Teaching GIScience with Free and Open Source Software?–A first Assessment. In 6th International Conference of GIScience, Zurich, Switzerland.

Steiniger, S., & Hunter, A. J. (2013). The 2012 free and open source GIS software map–A guide to facilitate, development, and adoption. COMPUT ENVIRON URBAN SYST. 39: 136-150.

Taba, S. E. S., Keivanloo, I., Zou, Y., Ng, J., & Ng, T. (2014). An exploratory study on the relation between user interface complexity and the perceived quality. In International Conference on Web Engineering (pp. 370-379). Springer International Publishing.

Techtarget. (2015). RFM Analysis (Recency, Frequency, Monetary) Definition.. Available from: http://searchdatamanagement.techtarget.com/definition/RFM-analysis.

Thompson, B., & Sims, D. (2002). CRM improving demand chain intelligence for competitive advantage. Business Week. 3804, 75-82.

Tian, Y., Nagappan, M., Lo, D., & Hassan, A. E. (2015, September). What are the characteristics of high-rated apps? a case study on free android applications. In Software Maintenance and Evolution (ICSME), 2015 IEEE International Conference on (pp. 301-310). IEEE.

Tsai, C. Y., & Chiu, C. C. (2004). A purchase-based market segmentation methodology. EXPERT SYST APPL. 27(2): 265-276.

Wawer, R., Steenberghen, T., Rusztecka, M., Podolczak, A., & Orlitowa, E. 2008. Evaluation of FOSS4G software projects for environmental applications. Evaluation of gvSIG. Status: published.

Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). A review of the application of RFM model. AFR J BUS MANAG. 4(19): 4199.

Wikipedia. (2015). Open source software. Available from: https://en.wikipedia.org/wiki/Open-source_software

Woo, J., Lee, M. J., Ku, Y., & Chen, H. (2013). Modeling the dynamics of medical information through web forums in medical industry. TECHNOL FORECAST SOC.

Wu, H. C., Chen, T., & Chiu, M. H. (2010). Analyzing customer sales data with a fuzzy set approach. INT REV COMPUT SOFTWARE. 5(5): 536-539.

Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. EXPERT SYST APPL. 36(3): 5866-5871.