International Journal of Information Management 38 (2018) 201–216

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Assessing web sites quality: A systematic literature review by text and association rules mining

Rim Rekik[a,*], Ilhem Kallel[a,c], Jorge Casillas[b], Adel M. Alimi[a]

[a] REGIM-Lab: Research Groups in Intelligent Machines, University of Sfax, ENIS, BP 1173, Sfax, 3038, Tunisia
[b] Department of Computer Science and Artificial Intelligence, University of Granada and the Research Center on Information and Communications Technology (CITIC-UGR), Granada, E-18071, Spain
[c] ISIMS: Higher Institute of Computer Science and Multimedia of Sfax, University of Sfax, Tunisia

ARTICLE INFO

Keywords:
Assessing web sites
Multiple criteria decision making
Text mining
Association rules mining
Systematic literature review

ABSTRACT

Nowadays society is deeply affected by web content. A web site, regardless of its category, can provide or not for users their needs. To identify its strengths and weaknesses, a process of analyzing and assessing its quality, via some criteria, is necessary. Assessing web sites is considered as a Multiple Criteria Decision Making problem (MCDM), with a massive number of criteria; a reduction phase is needed. This paper presents, firstly a Systematic Literature Review (SLR) to identify the purposes of recent researches from the assessment and determine the affected categories; secondly, it proposes a process of collecting and extracting data (criteria featuring web sites) from a list of studies. Text mining is applied for this SLR to construct a dataset. Then, a method based on Apriori algorithm is assigned and implemented to find association rules between criteria and the category of the web site, and to get a set of frequent criteria. This paper also presents a review on soft computing assessing methods. It aims to help the research community to have a scope in existing research and to derive future developments. The obtained results motivate us to further probe datasets and association rule mining.

1. Introduction

Over the last decade, the circulating online data in the World Wide Web has been rapidly increasing in terms of amount and diversity. In addition, most of the internet visitors used those data without knowing their credibility. In order to guarantee the reliability and credibility of online data, several assessment solutions have emerged.

In the field of web assessment, evaluating the quality of web sites is a key to filtering interesting sites for surfing or not. A web site is regarded not only as a showcase for commercial goals but also as a scientific showcase for considering an institution or represented organization's reliability. Assessing fields is complex, due to the panoply of web sites such as E-commerce, education, entertainment, health, etc. and also the diversity of criteria qualifying a web site. In some cases, the most common strategy adopted by some approaches is to know the type of web site to propose an interesting set of criteria in the field of assessment. For example, a commercial web site that provides payments mode (Chen, Rungruengsamrit, Rajkumar, & Yen, 2013) will not be evaluated with the same factors as an educational one (Violante and Vezzetti, 2015). In other cases, some studies (Ozmen-Ertekin and Ozbay, 2012) proposed a general model to assess any type of web site by choosing common characteristics. Consequently, the problem is one of multiple criteria decision making (MCDM). Existing methods are mainly based on making a hierarchy to divide high level criteria, sub-level criteria and alternatives. In fact, each criterion is weighed but there is no assumption about any association between them.

1.1. Research purpose and motivation

One such topic of web sites' quality assessment is a critical problem in practice and research. Therefore, the first motivation is to identify the current state of the art by applying a systematic literature review (SLR) as introduced by Kitchenham et al., (2010) and Kitchenham and Charters (2007), and recently applied (Aleti, Buhnova, Grunske, Koziolek, & Meedeniya, 2013; Yagüe et al., 2014). Its aim is to find the goals of assessment in this field and to classify studies according to web sites' category in order to determine the different and important domains concerned by the evaluation. In general, it gives a good overview for exploring the existing research and expanding new horizons in the area. Moreover, text mining for SLR induces us to collect data from studies and apply association rules mining (see Appendix A for more details about this concept). The purpose is to search for criteria

* Corresponding author.
E-mail addresses: rim.rekik@ieee.org (R. Rekik), ilhem.kallel@ieee.org (I. Kallel), casillas@decsai.ugr.es (J. Casillas), adel.alimi@ieee.org (A.M. Alimi).

http://dx.doi.org/10.1016/j.ijinfomgt.2017.06.007
Received 26 December 2015; Received in revised form 19 April 2017; Accepted 16 June 2017

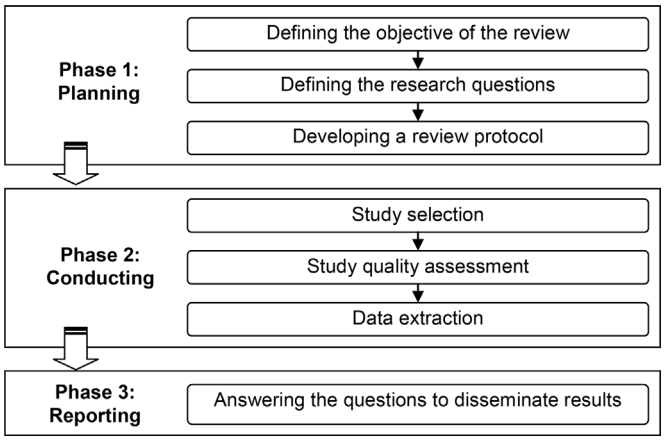

Fig. 1. Summary of the SLR process phases according to (Kitchenham and Charters, 2007).

dependencies with the category of the web site and to know the most reliable ones. Moreover, another motivation is to search for approaches that use soft computing techniques to resolve the problem.

1.2. Research method

In this work, we exercise the SLR approach according to Kitchenham and Charters' guidelines (Kitchenham and Charters, 2007) supervised by Fig. 1. The process is composed of three phases: planning, conducting and reporting the review. The aim of the first phase "planning the review" is to define the objective of the SLR and a clear review protocol. It specifies the main raised research questions, the adopted search strategy and a set of established inclusion and exclusion criteria to select a publication. The second phase "conducting the review" is for executing the protocol. In the third phase, we report the obtained results.

1.3. Content overview

The remaining part of the paper is organized as follows: Section 2 introduces a detailed description of the systematic literature review protocol. Section 3 presents conducting the review when applying the developed protocol. In Section 4, a report of the extracted data and related works from the previous review is discussed. Finally, some findings are drawn in Section 5 for developing future work and the main conclusions are given in Section 6.

2. Planning the review

In planning the review phase, the objective of the SLR and the research questions are defined. Moreover, a clear review protocol is developed. It consists of defining a search process strategy and the inclusion/exclusion criteria considered in the research.

```
(assess && "web site") OR (assess AND website) OR (evaluat* AND "web site")
OR (evaluat* AND website)
```

Fig. 3. Example of a Lucene search expression.

2.1. Objective and research questions

The objective of the SLR consists of exploring different aspects of the assessment of web sites quality existing in the literature. According to the objective and motivation of the review as described in section 1.1, we formulate a set of six research questions as follows:

RQ1: What are the purposes of the recent research from the assessment of web sites quality?

RQ2: What are the most common categories of web sites considered for the assessment?

RQ3: What are the criteria that characterize a web site? What are their semantic groups?

RQ4: Can we extract association rules between the criteria? Which ones?

RQ5: What are the frequent criteria considered in an assessment process?

RQ6: Due to the subjectivity and imprecision of that MCDM problem, are there studies that performed the assessment using soft computing or hybrid methods? If so, in which phase of the evaluation and what is the motivation of applying such intelligent methods?

2.2. Search process

It is important to follow a search strategy in order to ensure a convincing review conducted in phase 2 (see Fig. 1). In fact, a phase is needed for exploring scientific publications from related journals and conferences in relevant electronic sources such as Elsevier's Scopus, Elsevier's ScienceDirect, IEEExplore, ACM Digital Library, SpringerLink or Google Scholar. It is necessary to define some key concepts as selection words. Indeed, we consider some words such as "assessment", "assessing", "evaluation", "quality", "web sites", etc. A combination of these terms should be made to enlarge the scope of searching for better results.

2.3. Inclusion and exclusion criteria

Since we cannot include all collected papers, we introduce some inclusion and exclusion criteria. In order to select the most relevant ones, we determine the criteria that specify whether a study will be included or excluded. The first inclusion criterion based on terms which appeared in the titles, abstracts and keywords in studies by browsing the computer science discipline; an identification of relevant ones was established. However, papers published before 2009 and non-English written studies have to be excluded. In addition, we exclude some subdisciplines such as "Web services" not related exactly to the topic of assessment. Fig. 2 presents a boolean expression query performed using the Scopus database.

After obtaining a large set of papers, a step to eliminate short ones (up to 4 pages) and to make a fine filter on the set of papers is required

```
TITLE-ABS-KEY ( assessment quality WEBSITE ) OR TITLE-ABS-KEY ( assess quality WEBSITE )
OR TITLE-ABS-KEY ( assessing quality WEBSITE ) OR TITLE-ABS-KEY ( evaluation quality
WEBSITE ) OR TITLE-ABS-KEY ( evaluate quality WEBSITE ) OR TITLE-ABS-KEY ( evaluating
quality WEBSITE ) AND ( LIMIT-TO ( PUBYEAR , 2015 ) OR LIMIT-TO ( PUBYEAR , 2014 ) OR
LIMIT-TO ( PUBYEAR , 2013 ) OR LIMIT-TO ( PUBYEAR , 2012 ) OR LIMIT-TO ( PUBYEAR , 2011
) OR LIMIT-TO ( PUBYEAR , 2010 ) OR LIMIT-TO ( PUBYEAR , 2009 ) ) AND ( LIMIT-TO (
SUBJAREA , "COMP" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) ) AND ( EXCLUDE (
EXACTKEYWORD , "Web services" ))
```

Fig. 2. The performed search query in Scopus.

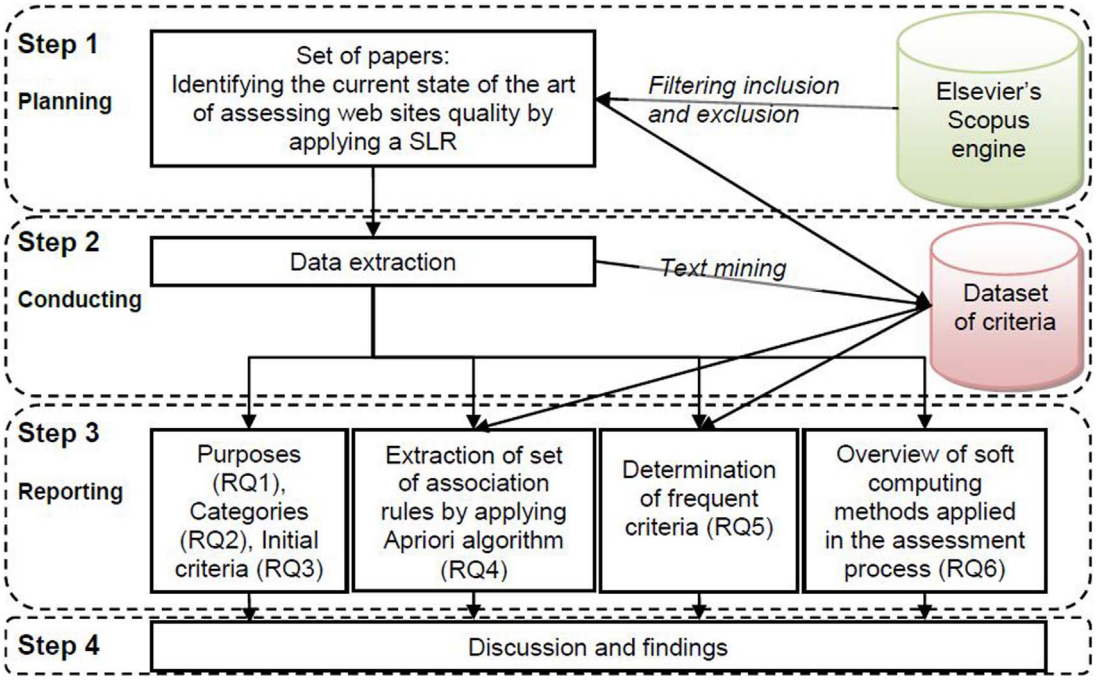

Fig. 4. Flowchart of our review methodology.

using Lucene search engine described in the next section (see Appendix B for more details). Fig. 3 presents one example of a Lucene search expression.

3. Conducting the review

Fig. 4 depicts a flowchart methodology containing all steps followed in this review.

3.1. Study selection

Following the above search strategy, we conducted a paper selection process illustrated by Fig. 5. Indeed, 4321 document results are returned in the first search on Scopus database by applying the initial inclusion/exclusion criteria based on terms which appeared in article title, abstract and keywords. Then, in step 2 we limited the search to the subject area "Computer Science", to studies written in English and published between 2009 and 2015. It returns 1041 articles. Step 3 also includes a refinement by excluding keywords that do not respond to the topic such as "Web services"; the number was reduced to 758. Step 4 is based on accessible PDF documents PDF (578). In step 5; we parse all these documents taking into consideration the other inclusion/exclusion criteria without short papers and with applying a fine filter using Lucene by searching the frequency of defined terms in articles. Finally; 532 papers is the total number of distinct papers included in the SLR.

3.2. Study quality assessment

The evaluation concerns each study in order to be sure that it satisfies the objective of the research and all inclusion/exclusion criteria. A step of skimming and reading parts from the articles is necessary. In this step, we also refer to figures, tables, appendixes and other indices for analyzing them.

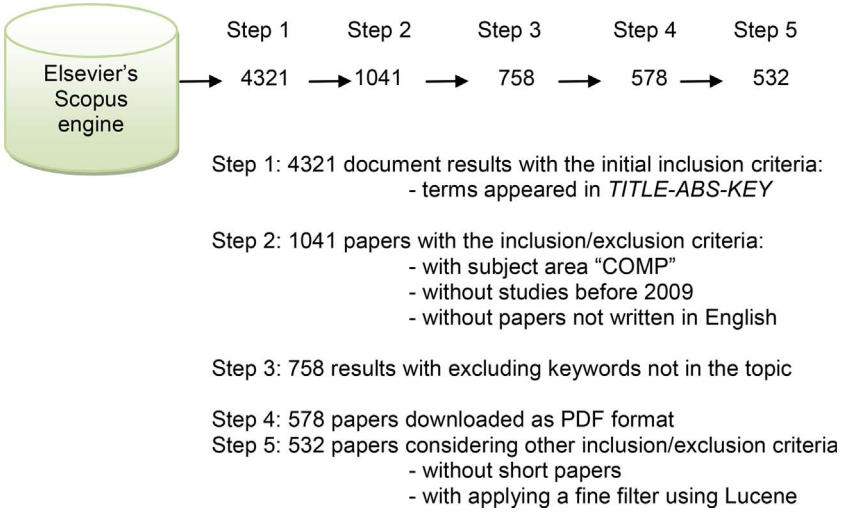

Fig. 5. Our paper selection process by applying the search strategy filtering criteria.

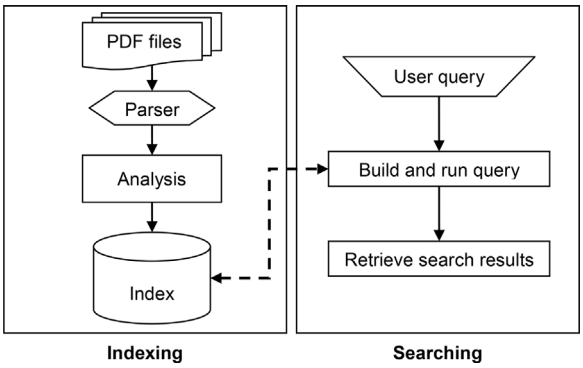

Fig. 6. The indexing and the searching steps with Lucene adapted from (McCandless, Hatcher, & Gospodnetic, 2010).

3.3. Data extraction

The data extraction process was elaborated in order to publish and disseminate results in the reporting phase of the SLR.

3.3.1. Text mining for papers included in the SLR to build a dataset

Text mining is applied to the 532 papers included in the SLR. It is the solution to parse a huge set of papers to extract needed information and build the dataset of criteria.

Lucene is an open-source Java API and a powerful text search engine tool (https://lucene.apache.org/). Its purpose is to search for documents with specified terms.

So, the followed steps are:

1. All papers are downloaded in PDF format.
2. Apache Lucene does not have the ability to extract text from PDF files. All of them are parsed by Apache Tika (https://tika.apache.org/) which is a library in Java that extracts text from different file types. Then, a library of documents is built in Text format.
3. Lucene looks for string only, it analyzes Text files and so creates an index from them. Then, it enables us to query against the indices to retrieve the matching results. Indexing and searching steps are presented in Fig. 6 and the process is detailed in Appendix B.
4. To browse the contents of Lucene index, we used the Luke GUI tool written in Java (https://github.com/DmitryKey/luke/releases). It helps in running a search expression by criterion and presenting results.

For each criterion, Luke is used to search the documents that have a term and build the transaction for the dataset with the set of terms included in each document. Then, to decide if a document has a term, instead of considering just the use of the word once, we used the scoring formula in Eq. (1) (available in Lucene):

$$score(q,\ d) = coord(q,\ d) \cdot queryNorm(q) \cdot \sum_{t\ in\ q} (tf(t\ in\ d) \cdot idf(t)^2 \cdot t.getBoost() \cdot norm(t,\ d)) \quad (1)$$

where:

- *tf(t in d)* is term's frequency t in document d. Documents that have more occurrences of a given term receive a higher score.
- *idf(t)* is Inverse Document Frequency.

$$idf(t) = 1 + \log\left(\frac{numDocs}{docFreq + 1}\right) \quad (2)$$

```
0.2921  weight (contents:usability in 414)
[DefaultSimilarity], result of:
  0.2921  fieldWeight in 414, product of:
    11.0905  tf(freq=123.0), with freq of:
      123.0000  termFreq=123.0
    1.6857  idf(docFreq=267, maxDocs=532)
    0.0156  fieldNorm(doc=414)
```

Fig. 7. Score explanation of "usability" term in Doc. Id 414.

where *numDocs* is the total number of documents and *docFreq* is the number of documents with the term.

- *coord(q,d)* is a coordination factor used to reward documents that contain a higher percentage of the query terms.
- *queryNorm(q)* is a query normalization factor used to normalize a query so scores; It plays the role of a weighting factor.
- *t.getBoost()* is a search time boost of term t in the query q, or a given high score to some particular thing. The higher the boost factor is, the more relevant the term will be, and therefore the higher the corresponding documents score.
- *norm(t,d)* is a combination of the length factor with the indexing time.

An example of a scoring explanation for a document considering "usability" criterion as a term through a printout from Luke GUI is given in Fig. 7. The final score in Doc. Id 414 is 0.2921 which is the product of tf, idf and fieldNorm.

Another example of a scoring explanation for a document when considering multiple queries combined using Boolean operators as "content OR information" is given in Fig. 8.

We fixed a threshold to consider if the term is relevant enough in the document as to consider that the paper uses this criterion.

3.3.2. Collection of data

Some examples of collected data are shown in Table 1 according to the purpose of assessment, category and references from studies included in the SLR.

According to Table 1, we notice different categories are implied in different purposes of assessment. All extracted categories are presented in the results' dissemination of RQ2.

```
0.2306  sum of:
  0.1329  weight(contents:content in 446) [DefaultSimilarity], result of:
    0.1329  score(doc=446,freq=183.0), product of:
      0.7471  queryWeight, product of:
        1.1218  idf(docFreq=470, maxDocs=532)
        0.6660  queryNorm
      0.1778  fieldWeight in 446, product of:
        13.5277  tf(freq=183.0), with freq of:
          183.0000  termFreq=183.0
        1.1218  idf(docFreq=470, maxDocs=532)
        0.0117  fieldNorm(doc=446)
  0.0977  weight(contents:information in 446) [DefaultSimilarity], result of:
    0.0977  score(doc=446,freq=158.0), product of:
      0.6647  queryWeight, product of:
        0.9981  idf(docFreq=532, maxDocs=532)
        0.6660  queryNorm
      0.1470  fieldWeight in 446, product of:
        12.5698  tf(freq=158.0), with freq of:
          158.0000  termFreq=158.0
        0.9981  idf(docFreq=532, maxDocs=532)
        0.0117  fieldNorm(doc=446)
```

Fig. 8. Score explanation of "content OR information" terms in Doc. Id 446.

Table 1
Some examples of data extraction from selected studies in the SLR.

Purpose of assessment	Category	References
Provide models or results and recommendations for administrators and/or users/customers and/or developers/enterprises for improvement of the web site quality and/or classifying it as good or bad.	Institutional Educational Health	(Rekik and Kallel, 2011) (Silambannan and Srinath, 2013) (Esteves and Lopez, 2010; Esteban, Porcel, Moral-Muñoz, & Herrera-Viedma, 2014)
Evaluate the electronic service to attract and increase the customers for making transactions or accessing to needed information.	E-commerce E-government	(Hsu et al., 2012; Lin, 2011) (Alanezi, Mahmood, & Basri, 2012)
Implement a set of criteria to insure the assessment.	Any type E-commerce	(Rekik, Kallel, & Alimi, 2014) (Hernández, Jiménez, & Martín, 2009; Hsu et al., 2012)
Search about the quality of online information and analyze it.	Health Any type Social media	(Yang et al., 2015) (Kotenko, Chechulin, Shorov, & Komashinsky, 2014) (Vosecky, Leung, & Ng, 2012)

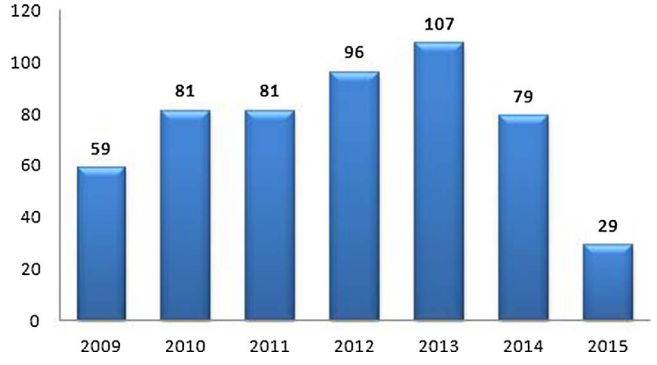

Fig. 9. The number of selected publications per year.

4. Reporting the review

The final phase of the SLR is a result report of the research questions. Indeed, the findings are based on studies retrieved by the conducted review phase and based on the research questions set out previously. Fig. 9 recalls and graphically presents the number of selected publications per year.

4.1. RQ1: what is the aim of the recent research from the assessment of web sites quality?

In the following, we focus on identifying the aim of the selected works to assess a web site. As presented in Table 1, examples of studies are regrouped according to their purpose; they provide models, results or recommendations to readers (i.e. administrators, users, customers, developers, enterprises) in terms of improvement of a web site quality, with eventually its classification as good or bad.

Moreover, the aim of other studies is to evaluate the electronic service to attract customers and enhance the number of transactions. For example, the E-commerce category is assessed for electronic service quality (Hsu, Hung, & Tang, 2012) then an interdependence perspective between multiple criteria and sub-criteria is studied. Another purpose is to search about the quality analysis of online information. For example, in healthcare web sites patients are searching for credible information to select physicians (Yang, Guo, Wu, & Ju, 2015). In addition, some researches implement a set of criteria for the evaluation process. A multitude of assessing purposes exists according to users' needs; we are here focusing on the well known ones.

4.2. RQ2: what are the most common categories of web sites considered for the assessment?

In order to extract the assessed category from studies, text mining is applied for the extraction and analysis of this information. Fig. 10 puts on show the distribution of web sites categories according to the selected studies. It also indicates a scope of interest in evaluating categories such as Any type, Social Media (i.e. social networking, social bookmarking, forums, micro-blogging, wikis, etc.), E-commerce, Health, Educational, E-government, Service (i.e. travel, hotel, tourism, airlines, etc.), Institutional and Corporate.

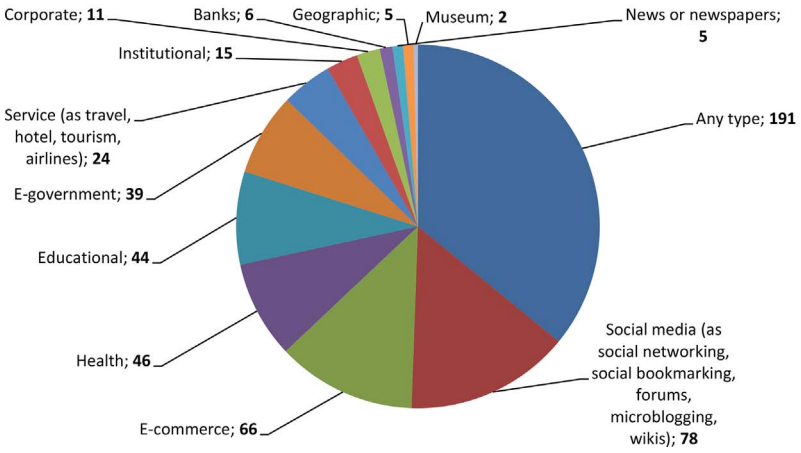

Fig. 10. The distribution of web sites categories considered for the assessment.

Table 2
Initial criteria and attributes.

Criteria or attributes	Number of studies	Examples of references
Accessibility	92	(De Lima et al., 2009; Eidaroos and Alkraiji, 2015)
Accuracy/Correctness/Trustworthiness	62	(Leite, Gonçalves, Teixeira, & Rocha, 2014; Ozmen-Ertekin and Ozbay, 2012)
Adequacy	9	(Schäfer, Kummer, & Günther, 2011)
Advertising	27	(Vatankhah et al., 2014)
Aesthetics/Visual appeal	54	(Pengnate and Antonenko, 2013)
Animation	14	(Vatankhah et al., 2014)
Attractiveness	23	(Leite et al., 2014)
Audio/Sound	28	(Cha, 2014)
Authority	19	(Rafe and Monfaredzadeh, 2012)
Availability	31	(Chen et al., 2013)
Clarity/Simplicity	22	(Silambannan and Srinath, 2013)
Color	42	(Rafe and Monfaredzadeh, 2012)
Communication	39	(Rafe and Monfaredzadeh, 2012)
Compatibility/Interoperability	17	(Chatzopoulos and Economides, 2009)
Completeness	30	(Rafe and Monfaredzadeh, 2012)
Comprehensiveness	13	(Leite et al., 2014)
Conciseness	11	(Leite et al., 2014)
Consistency/Coherence	30	(Eidaroos and Alkraiji, 2015; Leite et al., 2014)
Content/Information	320	(Vosecky et al., 2012)
Credibility/Believability	44	(Leite et al., 2014)
Currency	26	(Leite et al., 2014)
Customer support/Supportability	14	(Luo, Ba, & Zhang, 2012)
Design/Layout/Organization/Structure	111	(Eidaroos and Alkraiji, 2015; Pengnate and Antonenko, 2013)
Ease of use/User-friendliness/Ease of operation/Operability	60	(Cha, 2014)
Effectiveness	30	(Vatankhah et al., 2014)
Efficiency	55	(Fogli and Guida, 2013; Vatankhah et al., 2014)
Feedback	56	(Mavlanova, Benbunan-Fich, & Koufaris, 2012)
Form	14	(Chatzopoulos and Economides, 2009; Eidaroos and Alkraiji, 2015)
Functionality	34	(Fogli and Guida, 2013)
Image/Graphic	65	(Rafe and Monfaredzadeh, 2012)
Interactivity	72	(Silambannan and Srinath, 2013)
Language	32	(Rafe and Monfaredzadeh, 2012)
Learnability	20	(Fogli and Guida, 2013)
Link	76	(Eidaroos and Alkraiji, 2015; Vosecky et al., 2012)
Loyalty/Trustworthiness	15	(Liu and Wang, 2013)
Memorability	7	(Fogli and Guida, 2013)
Navigability/Browsing	89	(Eidaroos and Alkraiji, 2015; Vosecky et al., 2012)
Objectivity	15	(Leite et al., 2014)
Openness	6	(Chatzopoulos and Economides, 2009)
Payment	11	(Mavlanova et al., 2012)
Performance	27	(Rafe and Monfaredzadeh, 2012)
Personalization/Customization	36	(Rafe and Monfaredzadeh, 2012)
Playfulness/Enjoyment/Entertainment	20	(Chen et al., 2013; Lin, 2010)
Price/Cost	27	(Luo et al., 2012; Murakata and Matsuo, 2011)
Privacy	55	(Eidaroos and Alkraiji, 2015; Muñoz-Leiva, Luque-Martínez, & Sánchez-Fernández, 2010)
Product	57	(Luo et al., 2012; Mavlanova et al., 2012)
Purchase intention/Purchase	18	(Mavlanova et al., 2012)
Readability	24	(Vatankhah et al., 2014)
Reading	20	(Vatankhah et al., 2014)
Relevancy	85	(Demeester, Nguyen, Trieschnigg, Develder, & Hiemstra, 2012)
Reliability/Trustworthiness	64	(Fogli and Guida, 2013; Pengnate and Antonenko, 2013)
Reputation	28	(Rafe and Monfaredzadeh, 2012)
Reservation/Booking	7	(Murakata and Matsuo, 2011)
Responsibility	5	(Chen, Tzeng, & Chang, 2015; Fuertes-Callén, Cuellar-Fernández, & Pelayo-Velázquez, 2014)
Responsiveness	31	(Rafe and Monfaredzadeh, 2012)
Satisfaction/Fulfillment	69	(Chen et al., 2013; Luo et al., 2012)
Search	122	(Demeester et al., 2012; Vosecky et al., 2012)
Security	75	(Eidaroos and Alkraiji, 2015; Muñoz-Leiva et al., 2010)
Service	106	(Chen et al., 2015; Muñoz-Leiva et al., 2010)
Sociability/Sociality	4	(Liu and Wang, 2013)
Speed (of loading and/or download)	43	(Cha, 2014; Chen et al., 2015)
Technology	25	(Silambannan and Srinath, 2013)
	60	(Rafe and Monfaredzadeh, 2012)
Timeliness	22	(Ozmen-Ertekin and Ozbay, 2012)
Traceability	7	(Leite et al., 2014)
Transaction	25	(Mavlanova et al., 2012)
Trust	41	(Muñoz-Leiva et al., 2010; Pengnate and Antonenko, 2013)
Understandability/Comprehension	28	(Fogli and Guida, 2013)
Uniqueness/Value added	19	(Ozmen-Ertekin and Ozbay, 2012; Silambannan and Srinath, 2013)
Up-to-date	48	(Rafe and Monfaredzadeh, 2012)
Usability	131	(Fogli and Guida, 2013)
Usefulness	34	(Cha, 2014)
Validity	34	(Leite et al., 2014)
Video	42	(Cha, 2014)
Visibility	13	(Fuertes-Callén et al., 2014)

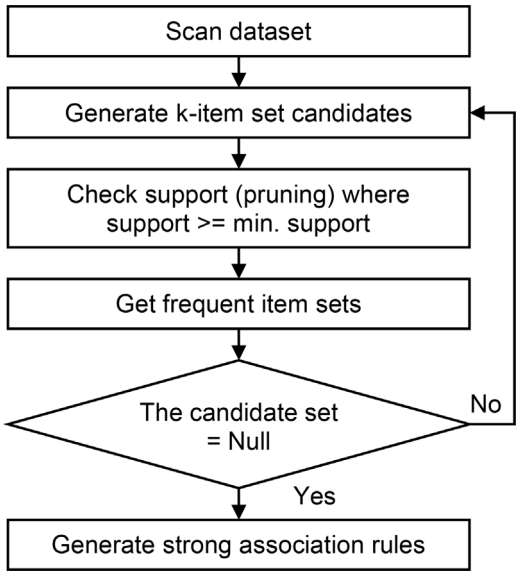

Fig. 11. The Apriori algorithm diagram (Agrawal and Srikant, 1994).

Indeed, this distribution shows the most important categories included in the evaluation. This reflects that Social media and E-commerce categories have competitive environments compared to other ones and should provide a good showcase to satisfy users' needs. The biggest part is also devoted to any type category that reflects the interest of users with usable web sites in general.

4.3. RQ3: what are the criteria that characterize a web site? what are their semantic groups?

Any process of assessing web site quality goes through the step of specifying certain criteria. It is a decision problem characterized by multiple criteria that can be solved by MCDM methods. Some criteria present a conflict with others as the confusion of meaning. Trustworthiness, for example, can mean the reliability or the accuracy of a web site or the loyalty of customers. Some factors are semantically similar and are regrouped together such as Content/Information, Design/Layout/Organization/Structure, Satisfaction/Fulfillment, Ease of use/User-friendliness/Ease of operation/Operability, credibility/believability, Aesthetics/Visual appeal, Personalization/Customization, Understandability/Comprehension, Playfulness/Enjoyment/Entertainment, etc. In the data extraction step, we collected intital different criteria and attributes from studies included in the SLR as described in Table 2.

The number of initial criteria and attributes is high. In order to reduce them, the next contribution is to study the relations' importance between them and to search for their dependencies with the category of the web site using association rules mining.

4.4. RQ4: can we extract association rules between the criteria? Which ones?

By applying a text mining process for sudies included in the SLR presented in the data extraction step, a dataset of criteria is implemented for the assessment. Extracting a set of association rules to discover interdependencies between criteria and their importance is the major purpose of this contribution. Before answering RQ4, it is necessary to elucidate some notions related to association rules such as confidence, support, lift and the Apriori algorithm applied to generate the strong association rules (Orriols-Puig and Casillas, 2010). Appendix A presents some fundamental notions.

4.4.1. Apriori algorithm

Apriori algorithm was introduced more than twenty years ago by (Agrawal and Srikant, 1994). It is considered as the most powerful association rule miner. It performs in two steps illustrated by Fig. 11.

- Discover all frequent item sets
  - Scan the dataset to find frequent items with an occurrence which is greater than or equal to the support threshold defined by the user. It is called the minimum support in the literature.
  - Generate candidates from frequent items and then find the frequent item sets.
- Generate a set of strong association rules from frequent item sets

When applying this algorithm, a study is considered as an item set; it represents the set of criteria for the assessment of web sites quality. Consequently, an item is considered as a criterion. All collected criteria from different studies are stored in a dataset.

4.4.2. Generation of association rules

Totally, 2054 is the number of generated association rules with minimum support 5% and confidence more than 25%. A filter is then applied by fixing a threshold to the lift of a rule greater than or equal to 2.15. The number is reduced to 1405. These rules are reliable and strong enough as the lift superior to 1. Finally, we analyze them to search for the most useful association rules. The analysis generates 632 rules. A presentation of relations between them is given by a network graph as shown in Fig. 12 using Gephi software (http://gephi.github.io/). The nodes characterize the antecedent or the consequent of rule. A step of partitioning the nodes into different groups is necessary. The filter modularity class (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) is applied to identify relations between nodes; it determines sets of vertices strongly connected between them. The partition in this graph has 6 classes with different colours. Another step is to weigh the nodes in order to resize them according to their degree of connection with each other calculated based on the number of incoming and outgoing links. Arrows are coloured with the same colour of source node and their thicknesses present their weights which are the lifts of rules.

Among the purposes of the research is to reduce the massive number of criteria. Consequently, we extract a useful set of association rules according to the web site category. Criteria that influence the category are determined. In Table 3, an example of extracted association rules useful for E-commerce category. They are sorted by rule lift. The higher its value the more the rule is interesting.

From this table, a detailed explanation of interesting association rules' results can be as follows:

- {E-commerce, Product} → {Purchase intention}: If the web site is E-commerce and the evaluation concerns a product, then it is highly probable it influences the purchase intention of customers. Its item set support is 13 with rule confidence 36.1% and lift 10.69.
- {Content, Product} → {Purchase intention}: The content describing a product influences the purchase intention of customers. The support of the rule is 11, confidence 31.4% and lift 9.31.

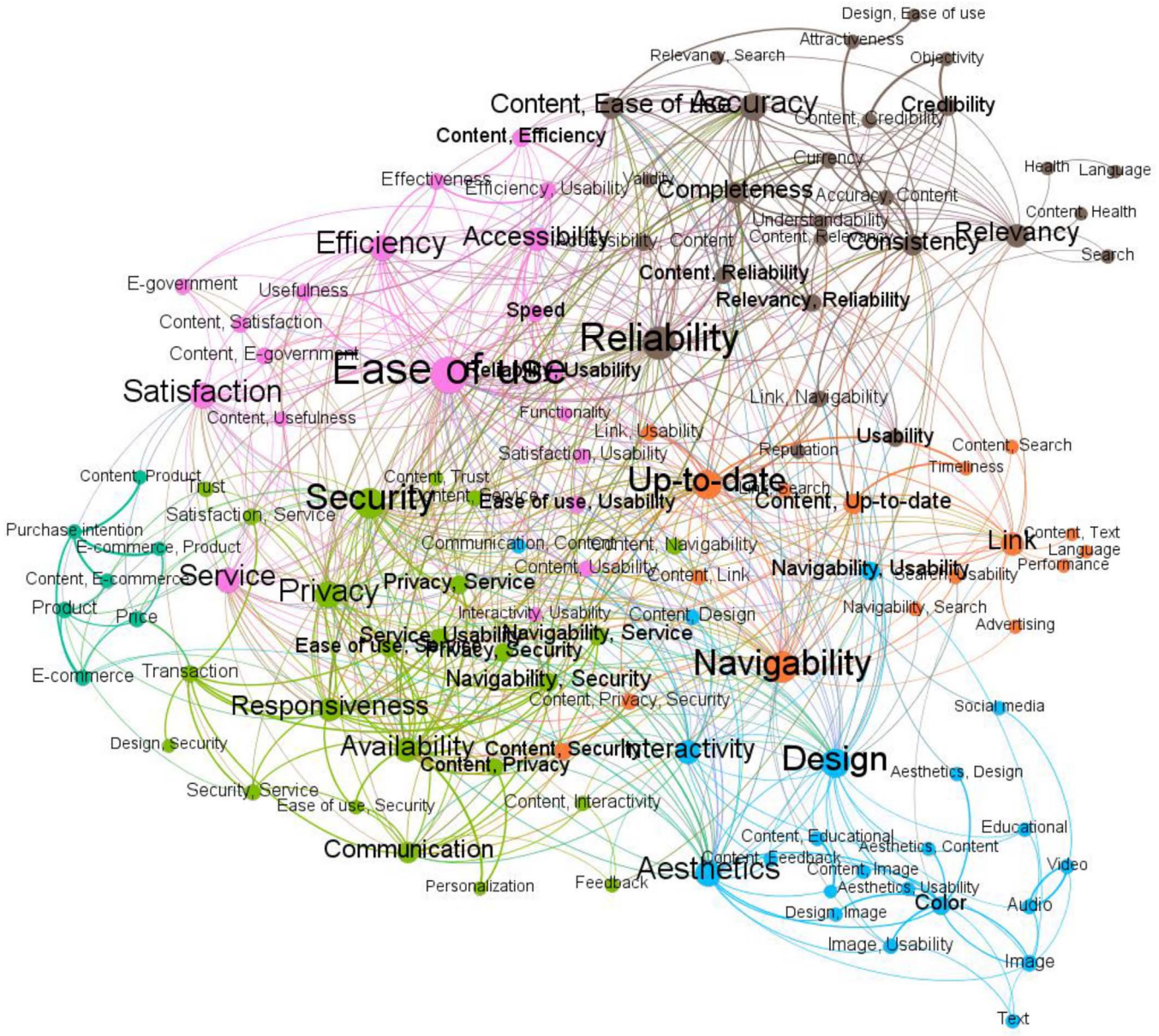

Fig. 12. Network graph to visualize the relations between the top useful 632 association rules.

- {Price} → {Transaction}: The price influences making transactions or not. The rule has 8 as item set support, 29.6% confidence and 6.32 lift.
- {Responsiveness} → {Transaction}: The web site should be responsive to allow making transactions easily. The support of the rule is 9, confidence 29% and lift 6.19.
- {Design, Security} → {Transaction}: A well designed web site and security are very important factors for making a transaction according to the rule lift 5.50. The corresponding support and confidence are 8 and 25.8%.
- {E-commerce} → {Security}: If the web site is E-commerce, then evaluating security is needed. The rule has 20 studies that support it with 30.3% confidence and 2.15 lift.

For deeper analysis of the relations with the E-commerce category, a dependency wheel graph (http://circos.ca/) is used as depicted by Fig. 13. It shows not only an important dependence of content and product with purchase intention but also it exposes the factors in relation with making a transaction which are price, satisfaction of users, the services provided, security, design and responsiveness of the web site. The E-commerce category depends on other criteria such as aesthetics and trust.

Criteria relations for other categories are also determined. Figs. 14–17 represent dependency wheel graphs respectively to Educational, E-government, Health and Social media categories.

The graph in Fig. 14 shows a high interdependence between the evaluation of content for Educational web sites and learnability. The site should provide means in order to assist learning to users. Moreover,

Table 3
Association rules related to E-commerce category.

Consequent	Antecedent	ItemSet Support	Rule Confidence %	Rule Lift
{Purchase intention}	{E-commerce, Product}	13	36.1	10.69
{Purchase intention}	{Content, Product}	11	31.4	9.31
{Purchase intention}	{Content, E-commerce}	12	30.8	9.11
{Purchase intention}	{E-commerce}	17	25.8	7.63
{Transaction}	{Security, Service}	13	31.0	6.60
{Transaction}	{Satisfaction, Service}	10	30.3	6.46
{Transaction}	{Price}	8	29.6	6.32
{Transaction}	{Responsiveness}	9	29.0	6.19
{Price}	{E-commerce, Product}	11	30.6	6.03
{Product}	{Content, E-commerce}	23	59.0	5.51
{Transaction}	{Design, Security}	8	25.8	5.50
{Responsiveness}	{Security, Service}	13	31.0	5.32
{Responsiveness}	{Satisfaction, Service}	10	30.3	5.21
{Product}	{E-commerce}	36	54.5	5.10
{Product}	{Price}	14	51.9	4.85
{Price}	{E-commerce}	15	55.6	4.49
{Service}	{Responsiveness}	24	77.4	3.89
{Security}	{Responsiveness}	16	51.6	3.67
{Aesthetics}	{Responsiveness}	10	32.3	3.18
{Security}	{Satisfaction, Service}	14	42.4	3.01
{Trust}	{E-commerce}	15	36.6	2.95
{Security}	{Price}	11	40.7	2.90
{Satisfaction}	{Price}	10	37.0	2.86
{Product}	{Satisfaction, Service}	10	30.3	2.83
{Security}	{Service}	42	39.6	2.82
{Product}	{Responsiveness}	9	29.0	2.71
{Satisfaction}	{Content, Product}	12	34.3	2.65
{Service}	{Price}	14	51.9	2.61
{Security}	{Trust}	15	36.6	2.60
{Satisfaction}	{Security, Service}	14	33.3	2.57
{Product}	{Security, Service}	11	26.2	2.45
{Satisfaction}	{Service}	33	31.1	2.40
{Responsiveness}	{E-commerce}	9	29.0	2.34
{Product}	{Content, Satisfaction}	12	25.0	2.34
{Satisfaction}	{Product}	17	29.8	2.30
{Service}	{Content, Satisfaction}	22	45.8	2.30
{Satisfaction}	{Trust}	12	29.3	2.26
{Aesthetics}	{E-commerce}	15	27.8	2.24
{Security}	{Content, E-commerce}	12	30.8	2.19
{Security}	{E-commerce, Product}	11	30.6	2.17
{Security}	{E-commerce}	20	30.3	2.15

the assessment of this category is based on evaluating interactivity, design and audio. A well designed course web site is related to easy navigation which leads to interactivity. In order to support interactive content for users, communication is highly needed. Audio is correlated to the quality of video and image. The feedback of learners is also essential for continuous improvement.

The E-government category can be assessed according to the following criteria accessibility, privacy, service furnished to citizens, efficiency, security, satisfaction and ease of use. The circle graph in Fig. 15 indicates an important link between content, privacy and personalization. The web site should be customized for users to ensure personalization. It can offer registration, configuration of services to individual user requirements, etc.

The Health category is mostly evaluated according to the relevancy and language criteria. The web sites should provide relevant content to users and different languages. Relevant content should be credible. A very important dependence between credible content and objectivity is presented in Fig. 16. It means that credible content should be unbiased, accurate, believable, complete and reliable. Accurate content is intended to be consistent, credible, complete, up-to-date and relevant according to the provided results in the circle graph.

The assessment of Social media web sites is characterized by evaluating video and reputation according to the associations in Fig. 17. Video is related to the quality of audio and image. Reputation is related to reliability, relevancy and security criteria. Responsiveness and accessibility are among important criteria associated to reliability. The social media circle graph shows that the search results should be relevant to users' needs.

4.5. RQ5: what are the frequent criteria considered in an assessment process?

Criteria diversity and multiplicity involve many issues in information space rendering, and then in the assessment process. That is why one of the purposes of this SLR is to identify the frequent criteria used for the assessment of web sites from the selected studies. Moreover, identifying the frequency of criteria can be considered as searching the frequent item sets from a large set of items (Orriols-Puig and Casillas, 2010).

Given a support threshold 54 means item sets that appear in at least 54 studies are called frequent item sets. To conclude, the plot in Fig. 18 shows some frequent criteria referred by the different studies; they are frequent item sets with size equal to 1.

RQ6: Due to the subjectivity and imprecision of that MCDM problem, are there studies that performed the assessment using soft computing or hybrid methods? If so, in which phase of the evaluation and what is the motivation of applying such intelligent methods?

Table 4 presents a synthetic overview of soft computing and hybrid methods in the field of web sites assessment when focusing on the motivations of applying methods based on intelligent techniques.

Several studies devote interest in establishing the evaluation with soft computing techniques. The fuzzy technique was largely used in different phases in the evaluation process. This can be explained by subjectivity and imprecision of such decision making problems. In fact, fuzzy reasoning is applied to obtain a web sites ranking (Rekik and Kallel, 2011) using linguistic terms such as {poor, average, good, excellent}. It also helps decision makers to distribute weights for criteria and to limit subjective human judgment (Lin, 2010; Xing, 2010).

5. Discussion and findings

The adopted strategy and obtained results by the SLR are compared to a review of literature (Chiou, Lin, & Perng, 2010) that proposes a strategic framework for web site evaluation in Table 5.

This section also highlights the main findings in applying the systematic literature review in the assessment of web sites quality. Its main objective is to explore studies in the domain, by filtering the essential ones and extracting data needed from them. Moreover, this SLR provides a clear answer for some fundamental considerations, essentially, what to be assessed, and how to assess a web site?

First, by answering the research questions in reporting the review phase we notice that selecting and gathering criteria was a critical and

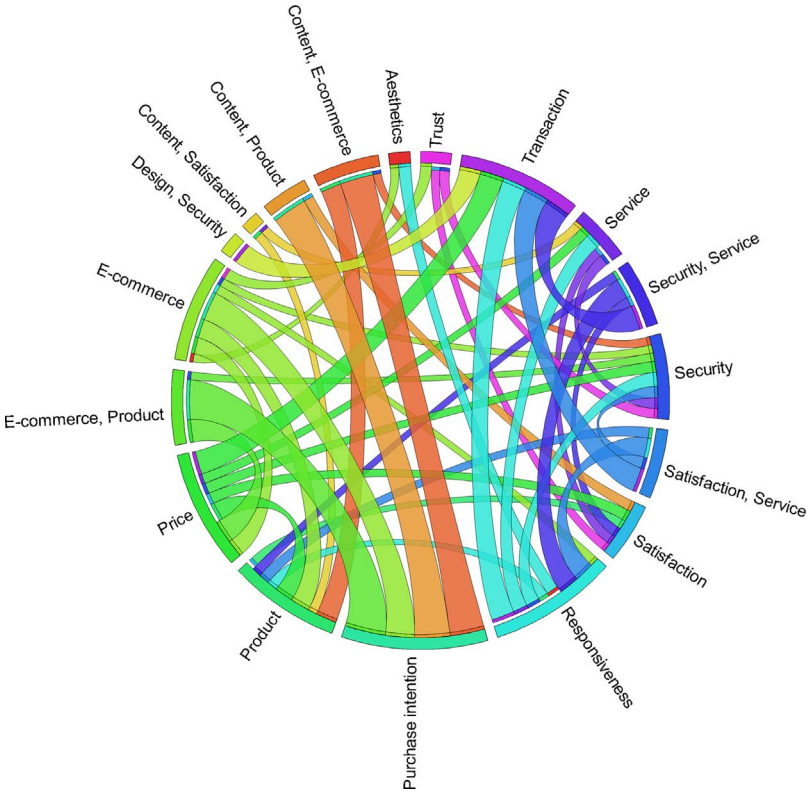

Fig. 13. Visualization of relations for E-commerce category.

essential phase in the process of evaluation. In fact, another finding is the set of association rules that reveals the interaction and relation between criteria and their importance. Another issue is extracting a set of frequent criteria as a phase of reducing the massive number of criteria.

Among the most relevant findings related to the evaluation methods and techniques is how to assess the quality from the selected criteria. Two ways have been applied, which are qualitative and quantitative.

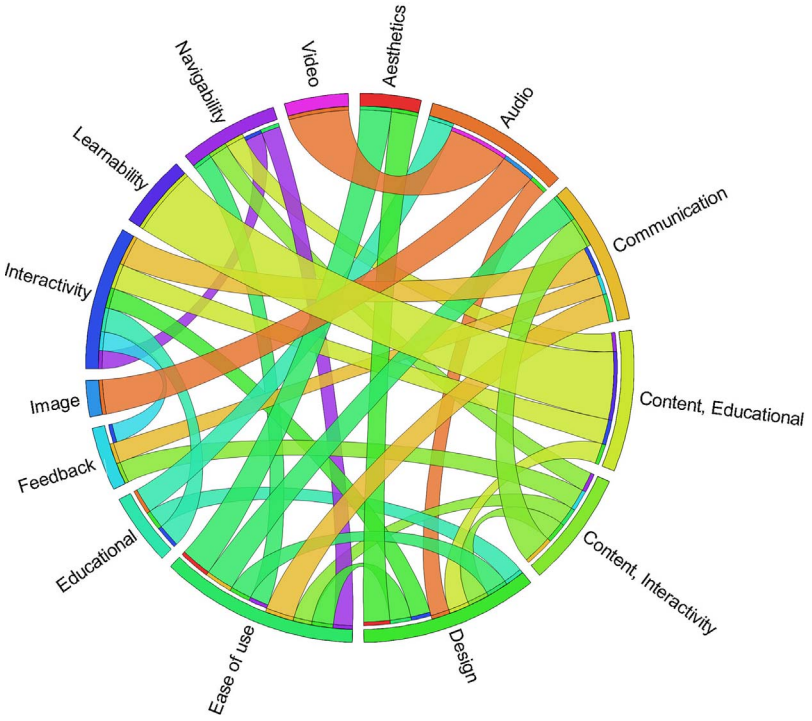

Fig. 14. Visualization of relations for Educational category.

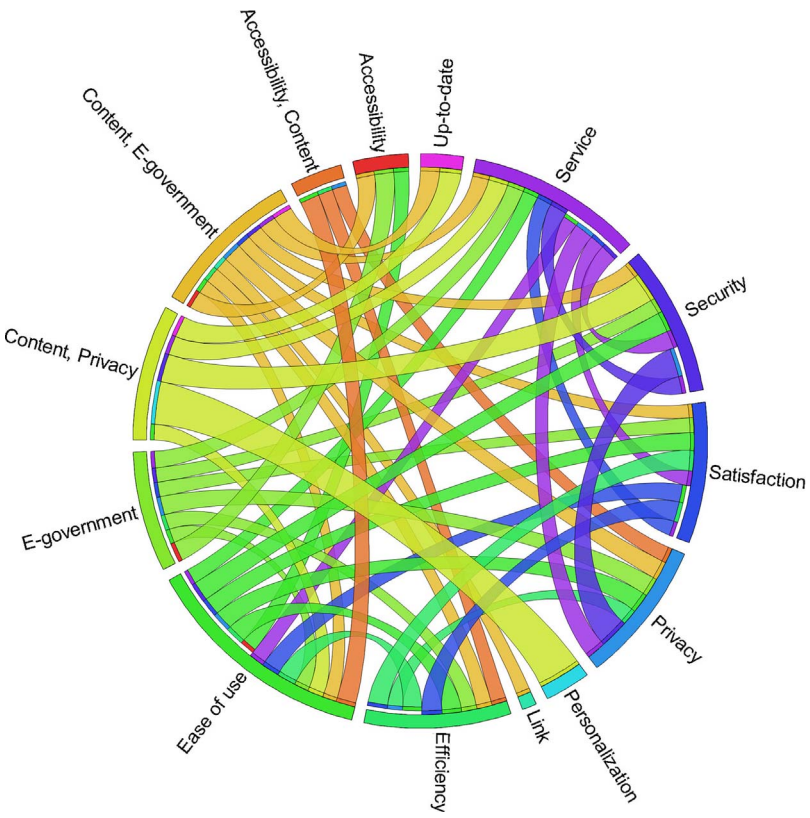

Fig. 15. Visualization of relations for E-government category.

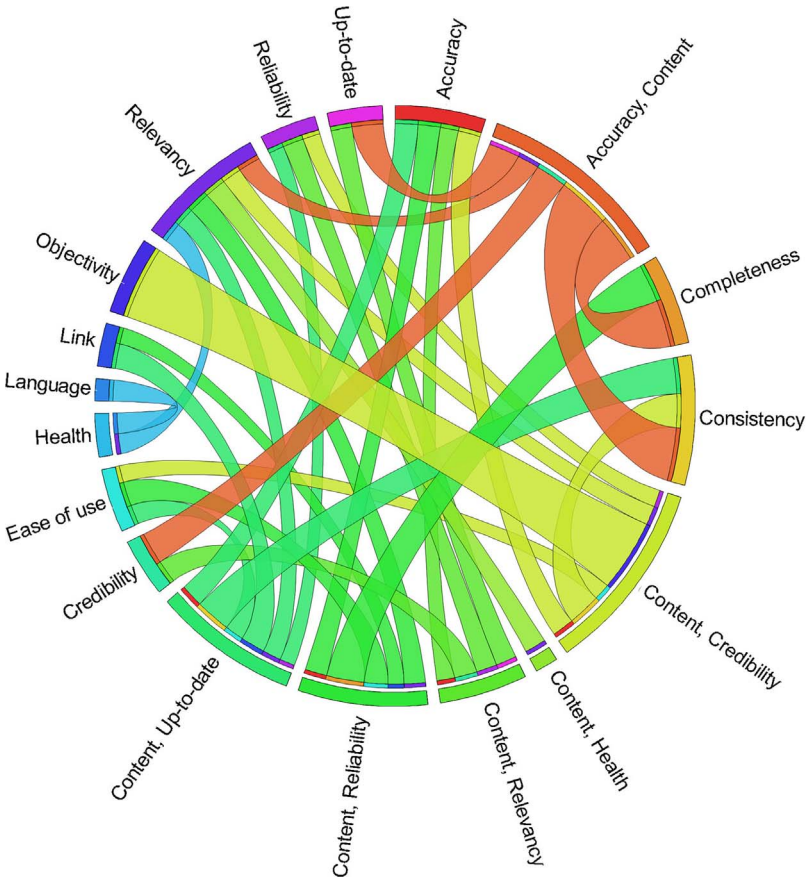

Fig. 16. Visualization of relations for Health category.

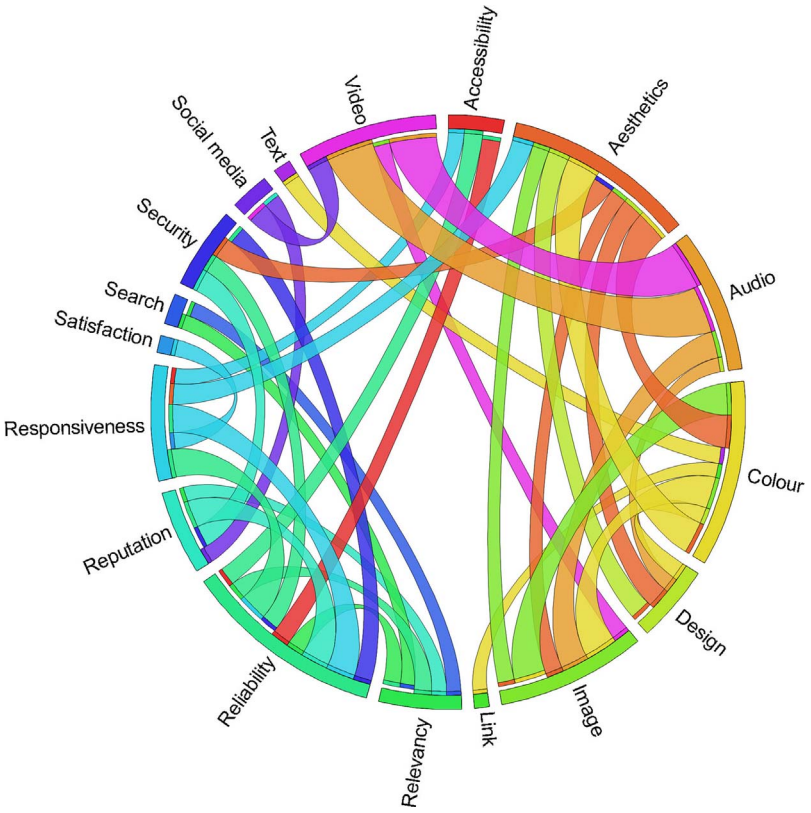

Fig. 17. Visualization of relations for Social media category.

The qualitative approach is characterized by observing user interactions from gathering user explanations and opinions. Some studies such as (Heradio, Cabrerizo, Fernández-Amorós, Herrera, & Herrera-Viedma, 2013; Hsu et al., 2012; Vatankhah, Wei, & Letchmunan, 2014) consider qualitative criteria from the user's judgments through questionnaires while the quantitative approach is based on measuring criteria by evaluation tools (De Lima, Lima, & De Oliveira, 2009; Rekik and Kallel, 2013).

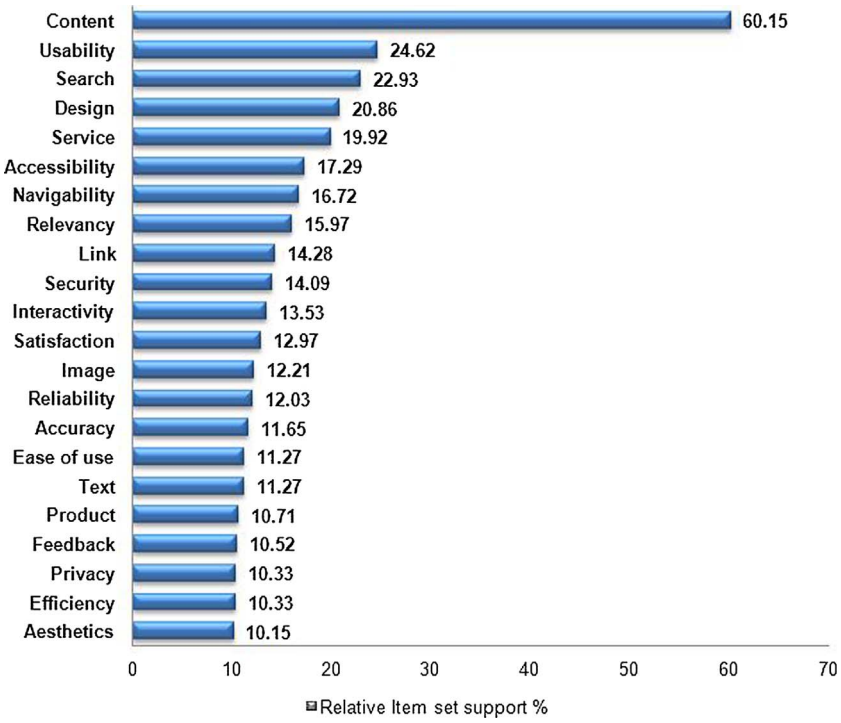

Fig. 18. Determination of frequent criteria.

Table 4
Soft computing methods in the field of the quality assessment of web sites.

Soft computing technique	Soft computing method	Concerned phase of the evaluation process	Motivation	References
Fuzzy	Fuzzy logic	In performing the fuzzy computation when the measured criteria are presented as inputs for the fuzzy system	Rank a web site	(Rekik and Kallel, 2011)
	Fuzzy Analytic Hierarchy process	Representation of users opinions	Prioritize and weigh criteria to rectify the limitation of subjective human judgment	(Lin, 2010; Xing, 2010)
	Fuzzy linguistic		Evaluate the services of library 2.0 Recommendation to users in health domain	(Heradio et al., 2013) (Esteban et al., 2014)
	Fuzzy c-means	Defining the number of clusters	Clustering results and determining the best quality solution for web search	(Cobos, Mendoza, Manic, León, & Herrera-Viedma, 2013)
Bayesian network	Bayesian Information Criterion	Evaluating the best solution and the number of clusters		
	Bayes net	Prediction of the best model for the evaluation of web page quality	Predict the quality of web sites	(Dhiman and Anjali, 2014)
Fuzzy neural	Adaptive neural fuzzy inference system (ANFIS)	In training and testing dataset		(Liu and Krasnoproshin, 2014)
Support Vector Machine (SVM)	SVM	Training tweet dataset	Filtering and ranking tweets by quality	(Vosecky et al., 2012)
Evolotuionary: Genetic Algorithm (GA)	GA based learning method	In determining the connection weights for a given hierarchical network by minimizing the root-mean-square error	Identify critical criteria for the evaluation	(Hu, 2009)

Table 5
Studies' comparison.

Study (Chiou et al., 2010)	The current study
– Review of literature 1995–2006	– Systematic literature review 2009–2015
– Collection of data from 83 selected studies	– Collection of data from 532 selected studies
– Category assessed: E-commerce	– Category assessed: Any type, Social media, E-commerce, Health, Educational, E-government, Service, Institutional, Corporate, etc.
– Comparison between frameworks	
– Proposition of a strategic framework that deals with:	
●Web site strategy	- Text mining applied for studies to construct a dataset of criteria
●Strategy consistency	- Association rules mining applied to the dataset and determination of interdependencies between criteria
●Evaluation factors and criteria most frequently used	- Providing a set of frequent criteria generally assessed.

Finally, soft computing methods are largely applied in the last decade in some steps of the assessment process as concluded in answering RQ6.

6. Conclusion

By following a SLR process, it was seen as a suitable strategy to define the objective and the questions in this research. Many purposes for assessing web sites are identified as providing recommendations for improving the quality, collecting a set of criteria and weighing them to ensure the assessment and ranking web sites.

The SLR also enables one to methodically collect a set of papers in the scope of web sites quality assessment. Firstly, initial criteria to perform the evaluation are defined and regrouped semantically. Then, text mining is applied to extract useful information from papers and create a data set of criteria. The same methodology of extracting and analyzing information is used to classify studies according to the assessed category.

Reliable association rules are obtained to study the interdependencies between criteria. They are presented by a network graph to show these relations clearly and to highlight the most important ones. We find that some categories depend on a specific set of criteria for example the E-commerce category is related to purchase intention, product, satisfaction, service, security, aesthetics, etc. The most important relations between the category and criteria are analyzed and clearly represented by dependency wheel graphs for E-commerce, Educational, E-government, Health and Social media categories.

Determining frequent criteria followed by current studies is made. It is an important phase in the absence of standards to follow.

With regard to future work we are interested in Multiple Criteria Decision Making methods existing in the literature for weighing and decomposing criteria (Rekik, Kallel, & Alimi, 2015, Rekik, Kallel, Casillas, & Alimi,2016) and choosing the suitable ones to prioritize criteria collected from the developed work for some categories of web sites such as the E-commerce category (Rekik, Kallel, Alimi, 2016).

Acknowledgments

The authors express thanks to the Erasmus Mundus Al-Idrisi II for funding the research reported under the Grant Agreement number 2013-2401/001-001-EMA2. The first author would also like to thank Knime community for using their instruments for data monitoring. The

authors acknowledge, too, the financial support of this work by grants from the Tunisian General Direction of Scientific Research (DGRST) under the ARUB program. This work also was partially supported by the Spanish Ministry of Science and Innovation (grant no. TIN2014-57251-P).

Appendix A. Fundamental notions about association rules

A set of association rules can be extracted from a set of data. An association rule is an implication of the form:
$X \rightarrow Y$, where both X and Y are item sets and $X \cap Y = \oslash$.
X can be defined as $\{c_1,c_2,\ldots,c_n\}$ and Y as c_j, the notation of the association rule $\{c_1,c_2,\ldots,c_n\} \rightarrow c_j$ means if X contains all of $c_1,c_2,\ldots,c_n$ then it is likely to contain c_j.
The support for an item set S is the number of item sets in the dataset containing all items in S.

■ Confidence of a rule

The problem of discovering association rules is characterized by If-then rules from the dataset. In order to measure the reliability of a rule, we compute the confidence expressed by Eq. (3).

$$Confidence(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \tag{3}$$

where both *X* and *Y* are item sets and $X \cap Y = \oslash$.
It is the fraction of studies with X that also contains Y.
Example: given 6 item sets from 6 studies with some criteria as follows:
S_1 = {Usability, Content, Aesthetics, E-services};
S_2 = {Usability, Content};
S_3 = {Content, Aesthetics, Efficiency};
S_4 = {Usability, Content, Reliability, Accessibility};
S_5 = {Security, Privacy};
S_6 = {Usability, Content, Aesthetics, Search}
A possible association rule is:
{Usability, Content} → {Aesthetics}.
Its confidence is 2/4 = 50% because there are 4 studies that focus on Usability and Content which are S_1, S_2, S_4 and S_6 but just S_1 and S_6 have Aesthetics.

■ Support of association rule

The support of a rule is an important measure that indicates the frequency of occurring patterns defined by Eq. (4). A rule that has very low support may occur simply by chance.

$$Support(X \rightarrow Y) = \frac{Support(X \cup Y)}{Number\ of\ total\ item\ sets} \tag{4}$$

where both *X* and *Y* are item sets and $X \cap Y = \oslash$.
Note: if $\{c_1,c_2,\ldots,c_n\} \rightarrow c_j$ has high support and confidence, then both $\{c_1,c_2,\ldots,c_n\}$ and $\{c_1,c_2,\ldots,c_n, c_j\}$ will be frequent.

■ Lift of an association rule

The lift of a rule expressed by Eq. (5)

$$Lift(X \rightarrow Y) = \frac{Support(X \cup Y)}{Support(X) \cdot Support(Y)} \tag{5}$$

where both *X* and *Y* are item sets and $X \cap Y = \oslash$.
The more the lift value is high, the more the rule is strong, and vice-versa.
A value greater than 1 indicates that X and Y appear more frequently together than expected; this means that the occurrence of X has a positive effect on that of Y, or that X is positively correlated with Y and vice versa in the case of a value smaller than 1.

Appendix B. The concept of indexing and searching in Apache Lucene

■ Indexing

The concept of indexing is the heart of all search engines in order to facilitate quick search among a large amount of data. So, converting data to a suitable format is the core of indexing and its output is called an *index*.
To create an *index* with Lucene, the first step is to create an *IndexWriter* object. The *IndexWriter* object is used to create the index and to add new index entries (i.e., Documents) to this index. The code to create an *IndexWriter* is presented in Fig. 19.
Note that *IndexWriter* takes two parameters, *dir* and *iwc*, which are *Directory* and *IndexWriterConfig* objects, respectively.

```
Directory dir = FSDirectory.open(Paths.get(indexPath));
Analyzer analyzer = new StandardAnalyzer();
IndexWriterConfig iwc = new IndexWriterConfig(analyzer);
IndexWriter writer = new IndexWriter(dir, iwc);
```

Fig. 19. The creation of an *IndexWriter*.

```
IndexReader reader = DirectoryReader.open(FSDirectory.open(Paths.get(index)));
IndexSearcher searcher = new IndexSearcher(reader);
```

Fig. 20. The opening of an index with *IndexSearcher*.

```
Analyzer analyzer = new StandardAnalyzer();
QueryParser parser = new QueryParser(field, analyzer);
```

Fig. 21. The creation of a *QueryParser*.

```
Query query = parser.parse(line);
```

Fig. 22. The creation of a query.

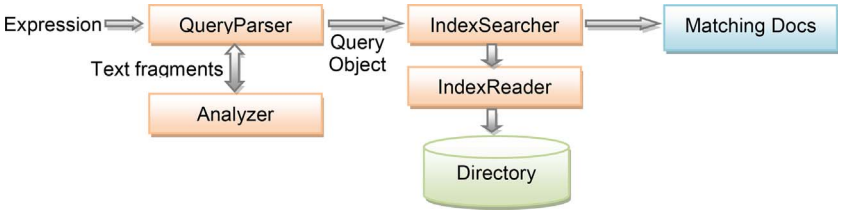

Fig. 23. The searching process.

- *Directory: a path to a directory where the Lucene index is stored.*

IndexWriterConfig: specifies the configuration of the *index* using an *analyser* for indexing data which is *StandardAnalyzer* in this case.

■ Searching

Searching in Lucene is simple and rapid as indexing. Searching is characterized by looking up words in an index to find the most relevant documents where they appear.

To search in an index, the first step is to open the index with an *IndexReader* and *IndexSearcher* (see Fig. 20).

The next step is to run a search in the index (Figs. 21 and 22). There is collaboration with the *IndexSearcher*, *StandardAnalyzer* and a *QueryParser*. A *QueryParser* is created by instantiating it using *StandardAnalyzer* (the same Analyzer that the documents in the index were created with).

Note that the parameter *field* can be a string representing a field name (e.g. title, author, contents).

Once a parser is created, to start search a query is then created by passing a search expression through *QueryParser*. The list of matching documents is finally retrieved by calling the *search()* method of the Lucene *IndexSearcher* object.

Fig. 23 illustrates the searching process using the different classes (https://www.tutorialspoint.com/lucene/lucene_search_operation.htm).

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. *Proceedings of the 20th international conference on very large data bases, VLDB '94* (pp. 487–499).

Alanezi, M. A., Mahmood, A. K., & Basri, S. (2012). A proposed model for assessing E-government service quality: an E-S-QUAL approach. *2012 international conference on computer and information science, ICCIS 2012 – a conference of world engineering, science and technology congress, ESTCON 2012,* 130–135.

Aleti, A., Buhnova, B., Grunske, L., Koziolek, A., & Meedeniya, I. (2013). Software architecture optimization methods: A systematic literature review. *IEEE Transactions on Software Engineering, 39*, 658–683.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*, P10008.

Cha, J. (2014). Usage of video sharing websites: Drivers and barriers. *Telematics and Informatics, 31*, 16–26.

Chatzopoulos, K.-C., & Economides, A. (2009). A holistic evaluation of Greek municipalities' websites. *Electronic Government, 6*, 193–212.

Chen, J. V., Rungruengsamrit, D., Rajkumar, T. M., & Yen, D. C. (2013). Success of electronic commerce Web sites: A comparative study in two countries. *Information and Management, 50*, 344–355.

Chen, F. H., Tzeng, G.-H., & Chang, C. C. (2015). Evaluating the enhancement of corporate social responsibility websites quality based on a new hybrid MADM model. *International Journal of Information Technology and Decision Making, 14*, 697–724.

Chiou, W.-C., Lin, C.-C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995 to 2006. *Information & Management, 47*, 282–290.

Cobos, C., Mendoza, M., Manic, M., León, E., & Herrera-Viedma, E. (2013). Clustering of web search results based on an iterative fuzzy C-means algorithm and bayesian information criterion. *2013 joint IFSA world congress and NAFIPS annual meeting, IFSA/NAFIPS 2013,* 507–512.

De Lima, S. T., Lima, F., & De Oliveira, K. M. (2009). Evaluating the accessibility of websites to define indicators in service level agreements. *Enterprise information systems, lecture notes in business information processing,* 858–869.

Demeester, T., Nguyen, D., Trieschnigg, D., Develder, C., & Hiemstra, D. (2012). What snippets say about pages in federated web search. *Information retrieval technology, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),* 250–261.

Dhiman, P., & Anjali (2014). Empirical validation of website quality using statistical and machine learning methods. *Proceedings of the 5th international conference on confluence 2014: The next generation information technology summit,* 286–291.

Eidaroos, A., & Alkraiji, A. (2015). Evaluating the usability of library websites using an heuristic analysis approach on smart mobile phones: Preliminary findings of a study in Saudi universities. *New contributions in information systems and technologies, advances in intelligent systems and computing,* 1141–1152.

Esteban, B., Tejeda-Lorente Á, Porcel, C., Moral-Muñoz, J. A., & Herrera-Viedma, E. (2014). Aiding in the treatment of low back pain by a fuzzy linguistic Web system. *Rough sets and current trends in computing, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),* 250–261.

Esteves, J., & Lopez, V. W. B. (2010). Comparing the quality of Latin American e-Health National Websites. *16th Americas conference on information systems 2010, AMCIS*

2010, 1418–1426.
Fogli, D., & Guida, G. (2013). Assessing corporate web sites: Quality model and methodology. *Current trends in web engineering, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),* 92–103.
Fuertes-Callén, Y., Cuellar-Fernández, B., & Pelayo-Velázquez, M. (2014). Determinants of online corporate reporting in three Latin American markets: The role of web presence development. *Online Information Review, 38*, 806–831.
Heradio, R., Cabrerizo, F. J., Fernández-Amorós, D., Herrera, M., & Herrera-Viedma, E. (2013). A fuzzy linguistic model to evaluate the quality of Library 2.0 functionalities. *International Journal of Information Management, 33*, 642–654.
Hernández, B., Jiménez, J., & Martín, M. J. (2009). Key website factors in e-business strategy. *International Journal of Information Management, 29*, 362–371.
Hsu, T.-H., Hung, L.-C., & Tang, J.-W. (2012). The multiple criteria and sub-criteria for electronic service quality evaluation: An interdependence perspective. *Online Information Review, 36*, 241–260.
Hu, Y.-C. (2009). Fuzzy multiple-criteria decision making in the determination of critical criteria for assessing service quality of travel websites. *Expert Systems with Applications, 36*, 6439–6445.
Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.
Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., et al. (2010). Systematic literature reviews in software engineering – a tertiary study. *Information and Software Technology, 52*, 792–805.
Kotenko, I., Chechulin, A., Shorov, A., & Komashinsky, D. (2014). Analysis and evaluation of web pages classification techniques for inappropriate content blocking. *Advances in data mining. applications and theoretical aspects, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),* 39–54.
Leite, P., Gonçalves, J., Teixeira, P., & Rocha, A. (2014). Towards a model for the measurement of data quality in websites. *New Review of Hypermedia and Multimedia, 20*, 301–316.
Lin, H.-F. (2010). An application of fuzzy AHP for evaluating course website quality. *Computers and Education, 54*, 877–888.
Lin, C.L. (2011). The improvement strategy of online shopping service based on SIA-NRM approach. In: Smart Innovation, Systems and Technologies. pp. 295–306.
Liu, H., & Krasnoproshin, V. V. (2014). Quality evaluation of E-commerce sites based on adaptive neural fuzzy inference system. *Neural networks and artificial intelligence, communications in computer and information science,* 87–97.
Liu, G., & Wang, Y. (2013). Trust-oriented service provider selection in complex online social networks. In: Advanced Web Services. pp. 363–380.
Luo, J., Ba, S., & Zhang, H. (2012). The effectiveness of online shopping characteristics and well-designed websites on satisfaction. *MIS Quarterly: Management Information Systems, 36*, 1131–1144 A9.
Mavlanova, T., Benbunan-Fich, R., & Koufaris, M. (2012). Signaling theory and information asymmetry in online commerce. *Information and Management, 49*, 240–247.
McCandless, M., Hatcher, E., & Gospodnetic, O. (2010). *Lucene in action, second edition: Covers apache lucene 3.0* (2nd ed.). Greenwich, Conn .u.a: Manning Publications.
Muñoz-Leiva, F., Luque-Martínez, T., & Sánchez-Fernández, J. (2010). How to improve trust toward electronic banking. *Online Information Review, 34*, 907–934.
Murakata, K., & Matsuo, T. (2011). Information disclosure incentives in tourism information systems. *2011 3rd international conference on awareness science and technology, ICAST 2011,* 422–427.
Orriols-Puig, A., & Casillas, J. (2010). Evolution of interesting association rules online with learning classifier systems. In J. Bacardit, W. Browne, J. Drugowitsch, E. Bernadó-Mansilla, & M. V. Butz (Eds.). *Learning classifier systems, lecture notes in computer science* (pp. 21–37). Berlin, Heidelberg: Springer.
Ozmen-Ertekin, D., & Ozbay, K. (2012). Dynamic data maintenance for quality data, quality research. *International Journal of Information Management, 32*, 282–293.
Pengnate, S., & Antonenko, P. (2013). A multimethod evaluation of online trust and its interaction with metacognitive awareness: An emotional design perspective. *International Journal of Human-Computer Interaction, 29*, 582–593.
Rafe, V., & Monfaredzadeh, M. (2012). A qualitative framework to assess hospital/Medical websites. *Journal of Medical Systems, 36*, 2927–2939.
Rekik, R., & Kallel, I. (2011). Fuzzy reduced method for evaluating the quality of institutional web sites. *2011 7th international conference on next generation web services practices (NWeSP),* 296–301.
Rekik, R., & Kallel, I. (2013). Fuzz-Web: A methodology based on fuzzy logic for assessing web sites. *International Journal of Computer Information Systems and Industrial Management Applications, 5*, 126–136.
Rekik, R., Kallel, I., & Alimi, A. M. (2014). Extraction of association rules used for assessing web sites' quality from a set of criteria. *2014 14th international conference on hybrid intelligent systems (HIS),* 291–296.
Rekik, R., Kallel, I., & Alimi, A. M. (2015). Quality evaluation of web sites: A comparative study of some Multiple Criteria Decision Making methods. *2015 15th international conference on intelligent systems design and applications (ISDA),* 585–590.
Rekik, R., Kallel, I., & Alimi, A. M. (2016). Ranking criteria based on fuzzy ANP for assessing E-commerce web sites. *2016 IEEE international conference on systems, man, and cybernetics (SMC),* 3469–3474.
Rekik, R., Kallel, I., Casillas, J., & Alimi, A. M. (2016). Using multiple criteria decision making approaches to assess the quality of web sites. *International Journal of Computer Science and Information Security, 14*, 747–761.
Schäfer, K., Kummer, T.-F., & Günther, O. (2011). Measuring web site information and service quality ? an extended multi-attribute attitude model. *17th americas conference on information systems 2011, AMCIS 2011,* 1360–1369.
Silambannan, R., & Srinath, M. V. (2013). A convictive framework for quality base construction and evaluation of e-learning website. *Journal of Theoretical and Applied Information Technology, 58*, 147–156.
Vatankhah, N., Wei, K. T., & Letchmunan, S. (2014). Usability measurement of Malaysian online tourism websites. *International Journal of Software Engineering and Its Applications, 8*, 1–18.
Violante, M. G., & Vezzetti, E. (2015). Virtual interactive E-learning application: An evaluation of the student satisfaction. *Computer Applications in Engineering Education, 23*, 72–91.
Vosecky, J., Leung, K. W.-T., & Ng, W. (2012). Searching for quality microblog posts: Filtering and ranking based on content analysis and implicit links. *Database systems for advanced applications, lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics),* 397–413.
Xing, H.-H. (2010). Study of evaluating web-based courses based on FAHP. *Fuzzy information and engineering 2010, advances in intelligent and soft computing,* 435–443.
Yagüe, A., Garbajosa, J., Pérez, J., & Díaz, J. (2014). Analyzing Software Product Innovation Assessment by Using a Systematic Literature Review. pp. 5049–5058.
Yang, H., Guo, X., Wu, T., & Ju, X. (2015). Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. *Electronic Commerce Research and Applications, 14*, 192–203.