A NOVEL USER PROFILE-BASED FUZZY APPROACH FOR EVALUATING TRUST IN SEMANTIC WEB

SOMAYEH ASHTARI[1], MALIHE DANESH[2] AND HOSSEIN SHIRGAHI[3]

[1]Islamic Azad University, Science and Research Branch Khuzestan, Ahvaz, Iran.
[2]Faculty of Electrical and Computer Engineering,
University of Science and Technology of Mazandaran, Behshahr, Iran.
[3]Department of Computer Engineering, Jouybar Branch, Islamic Azad University, Jouybar, Iran.

[*]Corresponding author: hossein.shirgahi@gmail.com

(Received: 21st Jan 2019; Accepted: 29th April 2019; Published on-line: 1st June 2019)

https://doi.org/10.31436/iiumej.v20i1.1060

ABSTRACT: As a developed World Wide Web architecture, the Semantic Web collects traditional web contents with a formal and understandable semantic using a machine. The main purpose of the Semantic Web is to increase automation, web information processing, and improve interactions and collaboration among information systems. The subject of trust is one of the main challenges in the semantic web. Since different tools and individuals exist in the semantic web, a certain measure of trust in an entity cannot be used and a central system is responsible for data collection and estimating the reliability. In this study, a fuzzy system is used to evaluate the trust measure in the semantic web. For this purpose, the user profile data including a list of pages, user sessions, and visited pages in each session, and the time of page viewings are used as semantic parameters. After determining the general framework of trust in the semantic web, the effectiveness of the above mentioned semantic parameters on the trust measure is investigated and effective parameters are used for evaluation in the fuzzy system. The experiment results show that the proposed fuzzy method with a mean absolute error of 2.5% and an average precision of 97.5% could achieve the right value of trust in the semantic web.

ABSTRAK: Sebagai senibina World Wide Web, Semantik Web mengumpul kandungan web tradisional bersama semantik formal dan mudah difahami menggunakan mesin. Tujuan utama Semantik Web adalah bagi meningkatkan automasi, pemprosesan maklumat web, dan meningkatkan interaksi dan kerjasama antara sistem maklumat. Kepercayaan adalah salah satu cabaran utama dalam web semantik. Disebabkan perbezaan alatan dan pelbagai individu wujud dalam web semantik, langkah tertentu sebagai entiti dipercayai tidak dapat digunakan dan sistem pusat adalah bertanggungjawab bagi pengumpulan data dan kebolehpercayaan anggaran. Dalam kajian ini, sistem rawak telah digunakan bagi menilai tahap kepercayaan dalam web semantik. Bagi tujuan ini, data profil pengguna termasuk senarai halaman, sesi pengguna, dan halaman yang dikunjungi dalam setiap sesi, dan masa paparan halaman telah digunakan sebagai parameter semantik. Selepas menentukan rangka umum kepercayaan dalam web semantik, keberkesanan parameter semantik yang dinyatakan di atas pada ukuran kepercayaan telah disiasat dan parameter yang berkesan telah digunakan bagi penilaian sistem rawak. Keputusan eksperimen menunjukkan bahawa kaedah rawak yang dicadangkan dengan ralat mutlak purata sebanyak 2.5% dan

ketepatan purata sebanyak 97.5% dapat mencapai nilai kepercayaan yang benar dalam web semantik.

KEYWORDS: *semantic web; trust; user profile; fuzzy system*

1. INTRODUCTION

The time we are in is the age of the ability to choose the best information in the shortest possible time. The World Wide Web is an information environment that provides different people around the world with information storage, provision, and exchange using embedded technologies. In the primary web, non-human factors had no realization of the collected information and could not process them. Through the growth and evolution of the Web and the creation of hypertext links, the possibility of Web exploration and creating links between contents was provided. With the advent of languages like XML, the possibility of adding semantic information was then created in addition to what previously was only for the information format and display. This kind of information helped the machine to understand the information on the web and provided it with the ability to interact with non-human factors where it previously just had the ability to collect information on the web.

The Semantic Web is an extension of the primary Web (WWW) that has been derived from the injection of meaning into the Web, where computers can use, interpret, analyze and process a wide network of data and provide it to the user. The Semantic Web is a network of relations between entities that are represented as a graph, in which concepts and propositions are represented as nodes and edges, respectively. The evolution of Web 3.0 or Semantic Web makes users' searches much faster and easier on the Internet. The ontology is the major reason for the emergence and success of the semantic web. In fact, the ontology determines the relationship between concepts in web documents and the real world, by which the relevant documents are processed and become understandable for the machine and facilitates sharing between agents.

The highest level in the semantic web structure is dedicated to trust management that is associated with the trust in content and reliability [1]. Trust is an abstract concept that enables individuals to act in uncertain conditions without complete certainty. Trust, in fact, reflects the amount of belief a truster has with respect to the reliability of an entity (trustee). When it is said that entity A has the X value of trust to entity B, it means that based on the viewpoint of entity A, entity B could not perform the assigned task correctly with an X probability [2].

Due to the existence of various machines and individuals in the current Web as well as the semantic web, providing a certain measure as a trust value of an entity cannot be taken into consideration because the trust values are different with respect to the personal needs of truster and the trusted content. Therefore, this measure should be used in central systems where a central entity is responsible for collecting information and estimating the reliability.

In this study, trust in the semantic web is evaluated based on user profiles and it is investigated whether the use of semantic parameters, especially user profiles, could help to have a proper precision in evaluating the trust in the semantic web.

The rest of the paper is organized as follows. In section 2, various models and algorithms in the field of trust evaluation as well as previous studies in this scope are presented. Then, in section 3, the proposed framework for the trust evaluation in the

semantic web is described and effective parameters in the user profile are employed to model the proposed fuzzy system. Finally, in section 4, the proposed model is tested, evaluated and compared with previous studies.

2. RELATED WORK

There are different trust models and policies in the semantic web. These models and policies can be classified according to different analytical views and node communications. In general, there are two categories of trust models. The first model comes with a centralized configuration and there is a central manager or node for trust management. The second model has a distributed configuration that tries to directly recognize the request of the trustee entity. Distributed models are divided into two categories: local models and global models. Global models calculate a global trust value for each node in the network. The local trust models calculate the trust value by responding to the request of each node and essentially by calculating the obtained results from each node. In the following we will review previous studies in this scope.

Richardson et al. [3] used a trust network to calculate a user's belief in a phrase. This calculation is based on finding paths from the source to each node and provides an opinion about an expression. The final trust value of the expression is calculated by interpolated trust values and the integration of these values. They provide propagation-based trust prediction methods in the semantic web. Lesani et al. [4] used a fuzzy model to evaluate trust. This algorithm uses fuzzy linguistic words to determine the trust among users and also employs a technique for expressing trust between two nodes that may not be directly related to the graph of the social trust network.

Matsu and Yamamoto [5] empirically measured the two-sided effects of trust and the similarity of product rankings. In their study, they developed a trust prediction model using the support vector machine classifier to identify positive attributes such as similarity of product rankings that are common among users.

In [6], a time-aware prediction method was presented that predicts future trust relations and connections using a supervised learning method. The time-aware prediction method can monitor the effect of dynamics in the trust networks and thus improves the prediction accuracy of future trust connections. In the supervised model, an explicit trust value is necessary to train the trust prediction model as an output variable. In [7] and [8], users' trust was assessed based on a history of ranked errors in interacting with measurable recommendations. The error-based trust model provides a public trust that affects the participation of all users, but does not create a local trust (like private trust).

In [9], a multi factor ontology-based framework was developed to assess the quality of service that resulted in a novel trust model. Nguyen et al. [10] proposed a Bayesian network-based trust model for web service, in which some criteria such as direct opinion of the truster, user rating (subjective viewpoint), and QoS monitoring information (objective viewpoint) were considered to evaluate the trust level of web services.

Yahyavi [11] proposed a game theory based trust model for web services collaboration. The web service provider calculates the trust-based cost that is the consequence of the suggested cost and the trust value of the web service tender. The game winner is the Web service that has the lowest trust-based cost.

Liu [12] merged a trust management module with a standard service-oriented architecture and proposed a trust evaluation model from service entities by transforming a web service network into a network based on trust relationships. In [13], a fuzzy model

was proposed for the reputation reasoning in web services. In their fuzzy system, they estimated the user behavior, as a ranking factor, on web services based on three parameters response time, availability, and efficiency. In their study, they considered response time and availability as an objective aspect of user trust and efficiency as the subjective aspect of user trust for ranking web services. Additionally, ranking user behavior based on the above mentioned parameters was considered as a criterion of reputation and Web services ranking.

In [14], using classical linear regression models, an algorithm was proposed for trust prediction in the semantic web. Shirgahi et al. [15, 16] used parameters such as the social network validity, the validity of the number of links in pages, and semantic validity for the reliability evaluation, and considered the problem space as a cluster of semantic subnetworks due to the broad span of the semantic network space. The trust value of elements in each cluster was calculated locally within the cluster and in order to access the resources outside of each cluster, they calculated the trust of those resources based on their local trust and the inter-cluster trust. Based on the experimental results, the proposed method could provide an acceptable estimation for trust in the trust-based web network.

Shekarpour et al. [17] addressed the modeling and evaluation of trust in the Semantic Web. Their method depends on two algorithms: one for propagation, and another for aggregation. The propagation algorithm uses statistical methods and the aggregation algorithm is based on the weighting mechanism.

In the previous studies, some of the trust management techniques were distributed in which the trust value was calculated locally. The main problem of these studies is related to the low accuracy of the trust calculation and, in some cases, it is not possible to assess local trust due to the lack of information resources [18].

In this study, the aim is to use a distributed model in a global manner to achieve the appropriate precision with the acceptable execution time in evaluating trust in the Semantic Web. In large social networks, the scalability of the trust evaluation will increase to the acceptance level using the distributed model.

3. PROPOSED METHOD

3.1 The Architecture of the Proposed Method

In this study, user profile datasets are used based on semantic parameters. For this purpose, based on user profiles, a fuzzy system is firstly provided for the trust evaluation in the semantic web. Then, the precision of the proposed fuzzy system is investigated for the trust evaluation. The information collection from user profiles and other quantitative information obtained about a user can be helpful for the user trust estimation.

In the semantic web, reputation is the amount of popularity or reliability of each user based on some criteria. For example, if a user has written a lot of documents in the field of computer science, he/she is recognized as a content provider and also considered as a user who has a reputation in this science. Similarly, if a user has been read a lot of documents in the field of computer science, he/she is recognized as a content user. Meanwhile, he/she is a user who has a reputation in this science.

In this study, trust is estimated based on user profiles using the proposed fuzzy system; then, the estimated values are compared with the actual trust values in the dataset. The evaluation metrics such as absolute error, precision, recall, and F-Score are calculated

for the proposed method. The overall architecture of the proposed method is illustrated in Fig. 1.

In this study, the DePaul CTI web server with 5446 users and 20950 sessions was used for obtaining the user profile datasets. Data was extracted from a random sampling of users visiting the site over a two-week period. Files used in the dataset included files such as cti.code. This file contains a list of views related to the pages in the dataset with a page view ID that is in the range of 0 to 682. The total number of records in the dataset was 8583. Statistical information such as the percentage of support sessions, the number of sessions count which increments by 1 per view, and the average duration of page viewing by users. The URL stem is in a file named cti-stats.txt. A part of the file is mentioned in Fig. 2.

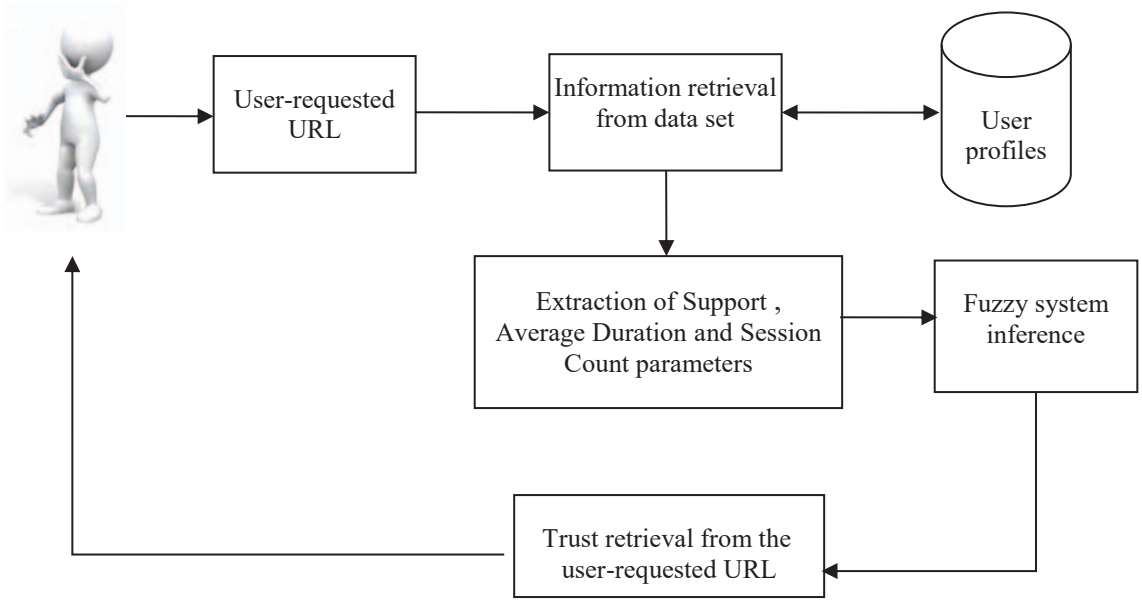

Fig. 1: Architecture of the proposed method.

Support	Session Count	Avgerage Duration	URL Stem
0.054	1136	48.46	/admissions/
0.000	1	11.00	/admissions/_vti_cnf/general.asp
0.000	1	6.00	/admissions/_vti_cnf/requirements.asp
0.000	2	35.00	/admissions/aboutgrad.asp
0.004	91	146.11	/admissions/career.asp
0.001	27	147.42	/admissions/checklist.asp
0.010	210	174.91	/admissions/costs.asp
0.004	86	80.24	/admissions/default.asp
0.010	206	89.30	/admissions/general.asp
0.001	31	41.08	/admissions/helloworld/arabic.asp
0.002	50	53.31	/admissions/helloworld/chinese.asp
0.001	11	43.40	/admissions/helloworld/italian.asp
0.001	13	56.42	/admissions/helloworld/portugese.asp
0.001	22	74.50	/admissions/helloworld/russian.asp
0.000	8	21.78	/admissions/helloworld/spanish.asp
0.001	19	26.20	/admissions/helloworld/thai.asp
0.001	15	111.25	/admissions/i20_support.doc
0.004	92	202.04	/admissions/i20visa.asp
0.002	50	60.70	/admissions/inqinsert.asp

Fig. 2: Part of cti-stats.txt file of the dataset.

Each user visits different pages in each session. The visits are stored in the cti.nav file. : Some of these visits are presented in Fig. 3. Consider the statement SESSION #n (USER_ID = k) in Fig. 3; for example, in the first section, it states that the user with user

ID 1 (USER_ID = 1) at session 1 visited which pages and also indicates the duration of page viewing. Similarly, the pages visited by users with the specific IDs over different sessions are shown in the next sections. The number of page visits is called frequency.

```
SESSION #1  (USER_ID = 1)
10203611        /search/newsearch.asp   /news/default.asp
10203615        /search/newsearch.asp   /search/newsearch.asp
10203624        /search/newsearch.asp   /search/newsearch.asp
10203628        /search/newsearch.asp   /search/newsearch.asp
10203630        /programs/      /search/newsearch.asp
10203632        /programs/2002/gradcs2002.asp   /programs/
10203639        /programs/courses.asp?depcode=21&deptmne=csc&courseid=211        /programs/2002/gradcs2002.asp
10203648        /courses/syllabus.asp?course=211-21-502&q=2&y=2002&id=483       /programs/courses.asp?depcode=
10203658        /courses/syllabus.asp?course=211-21-503&q=2&y=2002&id=363       /programs/courses.asp?depcode=
10203743        /courses/syllabus.asp?course=211-21-503&q=2&y=2002&id=363       /programs/courses.asp?depcode=
10203746        /programs/courses.asp?depcode=21&deptmne=csc&courseid=211        /programs/2002/gradcs2002.asp
10203748        /courses/syllabus.asp?course=211-21-504&q=2&y=2002&id=123       /programs/courses.asp?depcode=
-----------------------------------------
SESSION #2  (USER_ID = 2)
9045150 /news/default.asp       /
9045150 /admissions/    /news/default.asp
9045150 /news/  /news/default.asp
9045150 /advising/      /news/default.asp
9045151 /courses/       /news/default.asp
9045151 /people/        /news/default.asp
9045151 /programs/      /news/default.asp
9045151 /research/      /news/default.asp
9045151 /resources/     /news/default.asp
9069476 /news/default.asp       /
9069476 /news/  /news/default.asp
9069476 /advising/      /news/default.asp
9069476 /admissions/    /news/default.asp
9069476 /courses/       /news/default.asp
9069477 /people/        /news/default.asp
9069477 /programs/      /news/default.asp
9069477 /research/      /news/default.asp
9069477 /resources/     /news/default.asp
9132650 /news/default.asp       /
```

Fig. 3: Part of the cti.nav file of the dataset.

3.2 Distribution Analysis of the Data on the Problem's Input Parameters

The implementation of the fuzzy system of the proposed method was performed using MATLAB software. In order to determine the fuzzy intervals for the input parameters of the system, the histogram analysis of parameters support, session count, and average duration were used. Using this analysis, the number of fuzzy sets and their intervals were determined so that there was a relatively balanced sample distribution in different parts. Two fuzzy sets as no-support and ok-support were considered for the parameter support. For this purpose, different fuzzy sets were implemented and eventually the zmf set was used because it had the lowest output error. This set is based on polynomial functions. As seen in Fig. 4, the area of ok-support starts from the point 0.001.

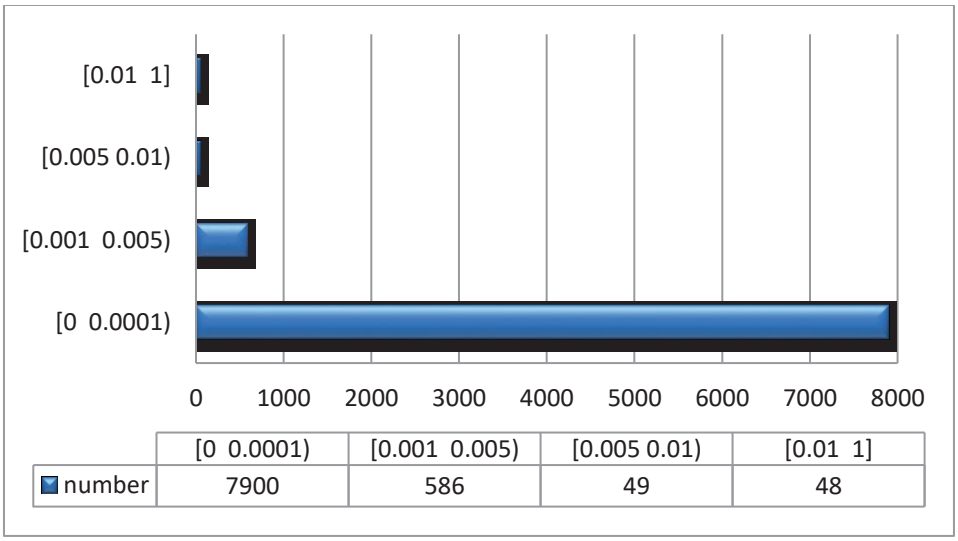

Fig. 4: Histogram of the data distribution for parameter support.

For the parameter session count, three fuzzy sets low, medium, and high were considered. As seen in Fig. 5, there are about 7,000 data samples in the range of 0 to 0.1 and about 800 data samples in the range 0.1 to 0.2. These numbers were used to determine the frequency of data in each numerical range (Fig. 5).

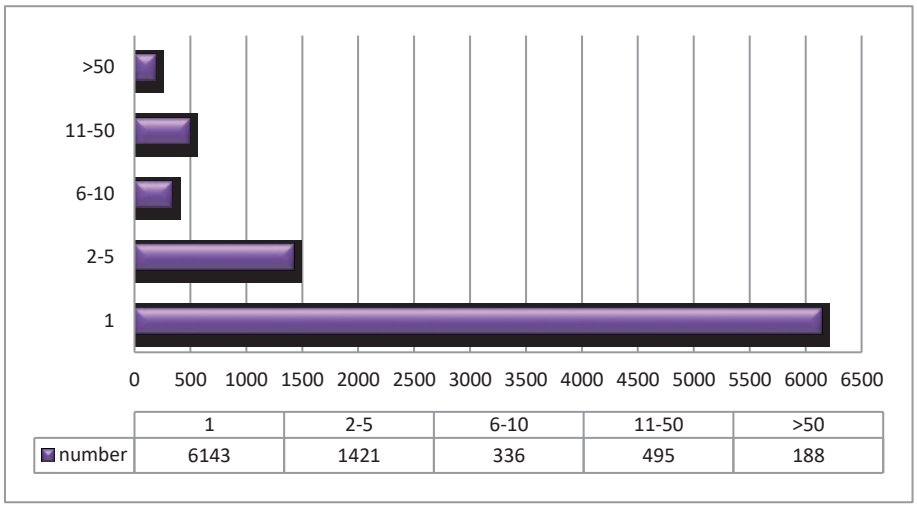

Fig. 5: Histogram of data distribution for the parameter session count.

For the parameter average duration, five sets were considered as follows (Fig. 6):

Time<10 sec: [low]

10< time<30 sec: [low medium]

30< time<60 sec: [medium]

60< time<90 sec: [medium high]

Time> 90 sec: [high]

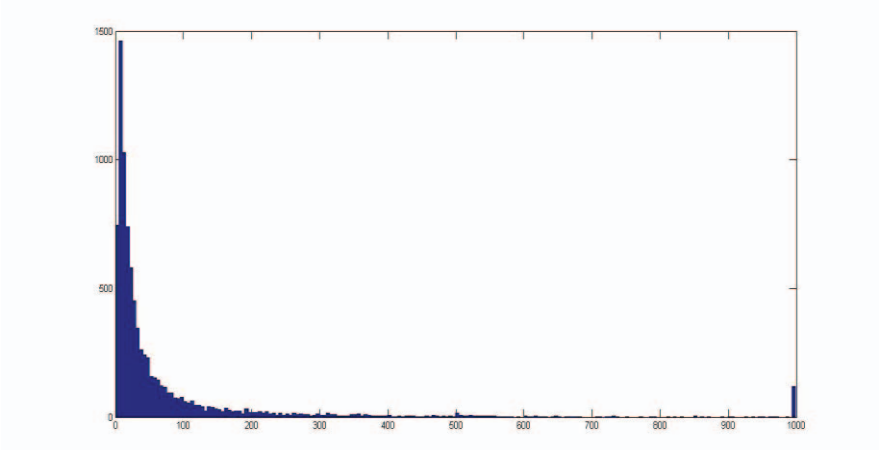

Fig. 6: Histogram of data distribution for the parameter average duration.

3.3 The Structure of the Proposed Fuzzy System

In this section, the fuzzy rules are first extracted based on the histograms obtained from the previous steps, and then three fuzzy sets are considered as following to determine the trust value:

a) Trust

b) Null
c) Distrust

The goal is to determine the trust status for each record (8583 records) in the dataset. Then, the actual trust status suggested by users is compared with the predicted trust value obtained by the proposed method. Based on the comparison, the accuracy of the proposed method can be determined. In this study, the curve membership function and the Max-min fuzzy inference method were used. Finally, in defuzzification phase, the mean center method is used to obtain the final result. The overall structure of the fuzzy system is demonstrated in Fig. 7.

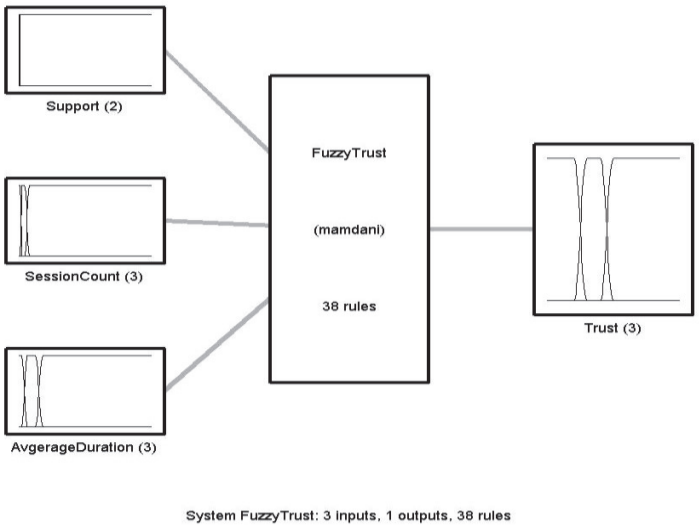

Fig. 7: The overall structure of the fuzzy system.

Based on the histogram data related to the parameter support, the fuzzy sets of the fuzzy input variable support are in the form of a parabola (Quadratic function curve) that is shown in Fig. 8. The no-support fuzzy set is a type of zmf that has a membership=1 for the value=0. There is an overlap between ok-support and no-support in the interval [0, 0.001]. Additionally, the ok-support fuzzy set is a type of smf that has a membership=1 for the value greater than 0.001; refer to Fig. 8.

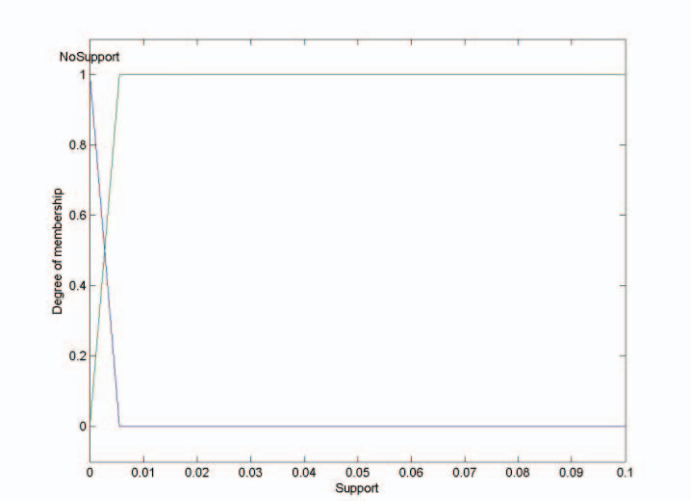

Fig. 8: The fuzzy sets of the fuzzy input variable support.

Based on the histogram data related to the parameter session count, the fuzzy sets of the fuzzy input variable session count are in the form of a parabola curve that is offered in Figure 9. The fuzzy set 'low' is a type of zmf that has a membership=1 for the value lower than or equal to 3. There is an overlap between 'low' and 'medium' in the interval (3 10). The fuzzy set 'medium' is a type of pimf that has a membership=1 for the value in the interval [10 20]. There is an overlap between 'high' and 'medium' in the interval (20 40). The fuzzy set 'high' is a type of smf that has a membership=1 for the value greater than 40 (Fig. 9).

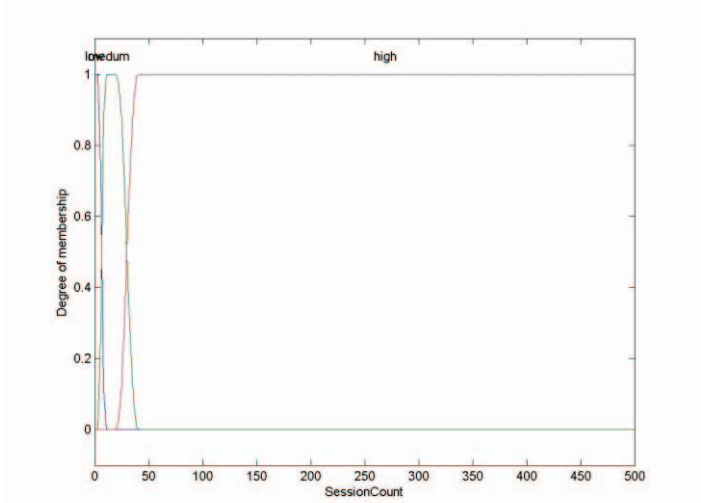

Fig. 9: The fuzzy sets of the fuzzy input variable session count.

Based on the histogram data related to the parameter average duration, the fuzzy sets of the fuzzy input variable average duration are in the form of a parabola curve that is explained in Fig. 10. The fuzzy set 'low' is a type of zmf that has a membership=1 for the value lower than or equal to 10. There is an overlap between 'low' and 'medium' in the interval (10 30). The fuzzy set 'medium' is a type of pimf that has a membership=1 for the value in the interval [30 60]. There is an overlap between 'high' and 'medium' in the interval (60 90). The fuzzy set 'high' is a type of smf that has a membership=1 for the value greater than 60 (Fig. 10).

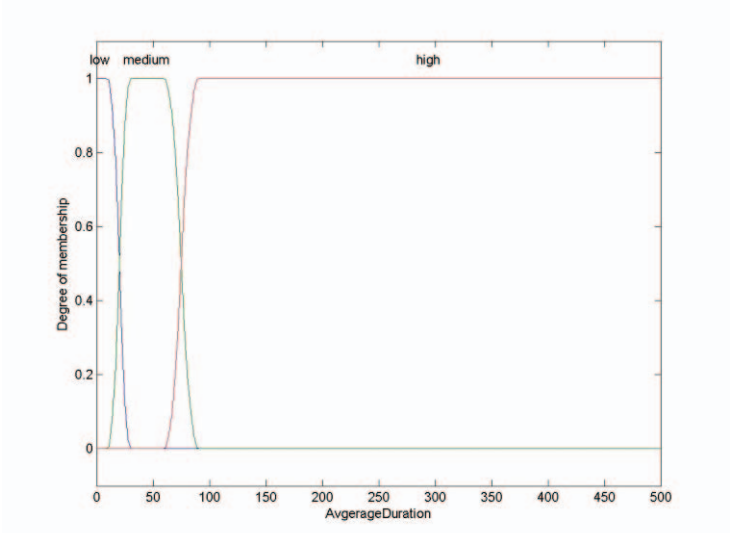

Fig. 10: The fuzzy sets of the fuzzy input variable average duration.

Based on the concepts related to the trust evaluation, the fuzzy sets of fuzzy output variable trust are in the form of a parabola curve that is mentioned in Fig. 11. The fuzzy set 'distrust' is a type of zmf that has a membership=1 for the value lower than or equal to 0.2. There is an overlap between 'distrust' and 'null' in the interval (0.2 0.3). The fuzzy set 'null' is a type of pimf that has a membership=1 for the value in the interval [0.3 0.4]. There is an overlap between 'null' and 'trust' in the interval (0.4 0.5). The fuzzy set 'trust' is a type of smf that has a membership=1 for the value greater than 0.5; as pointed out in Fig. 11.

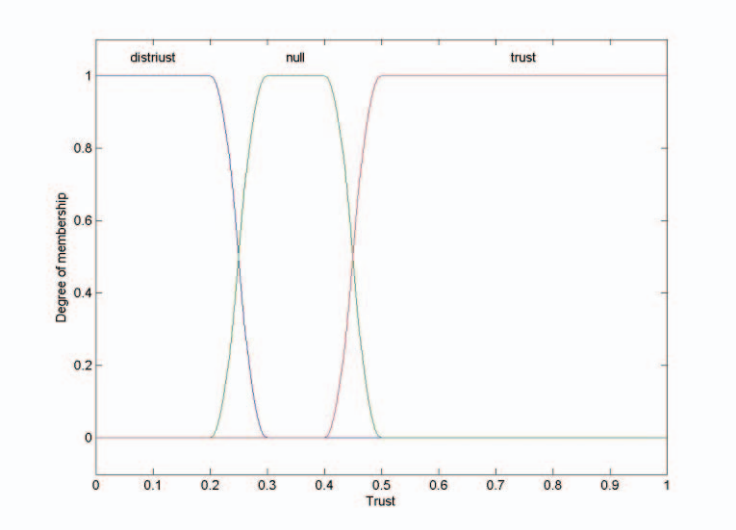

Fig. 11: The fuzzy sets of the fuzzy output variable trust.

In Fig. 12, some of the fuzzy rules in this study are illustrated. Please consider the following example to better understand the problem. Rule 1 states: assume that the value of the input variable support belongs to the no-support set, the value of the input variable session count is in the low range, and the value of the input variable average duration is in the low range as well. It means references to the desired web page are low with the weight 1 which will result in the distrust status.

1. (Support==NoSupport) & (SessionCount==low) & (AvgerageDuration==low) => (Trust=distriust) (1)
2. (Support==NoSupport) & (SessionCount==low) & (AvgerageDuration==medium) => (Trust=distriust) (0.67)
3. (Support==NoSupport) & (SessionCount==low) & (AvgerageDuration==medium) => (Trust=null) (0.33)
4. (Support==NoSupport) & (SessionCount==low) & (AvgerageDuration==high) => (Trust=distriust) (0.67)
5. (Support==NoSupport) & (SessionCount==low) & (AvgerageDuration==high) => (Trust=trust) (0.33)
6. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==low) => (Trust=distriust) (0.67)
7. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==low) => (Trust=null) (0.33)
8. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==medium) => (Trust=distriust) (0.33
9. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==medium) => (Trust=null) (0.67)
10. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==high) => (Trust=distriust) (0.33)
11. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==high) => (Trust=null) (0.34)
12. (Support==NoSupport) & (SessionCount==medum) & (AvgerageDuration==high) => (Trust=trust) (0.33)
13. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==low) => (Trust=distriust) (0.67)
14. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==low) => (Trust=trust) (0.33)
15. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==medium) => (Trust=distriust) (0.33)
16. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==medium) => (Trust=null) (0.34)
17. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==medium) => (Trust=trust) (0.33)
18. (Support==NoSupport) & (SessionCount==high) & (AvgerageDuration==high) => (Trust=distriust) (0.33)

Fig. 12: Some of fuzzy rules in this study.

Figures 13 to 15 show the effect of input parameters support and session count, input parameters support and average duration as well as the input parameters session count and average duration on the output trust. As can be seen, the trust output increases by increasing each of the input variables.

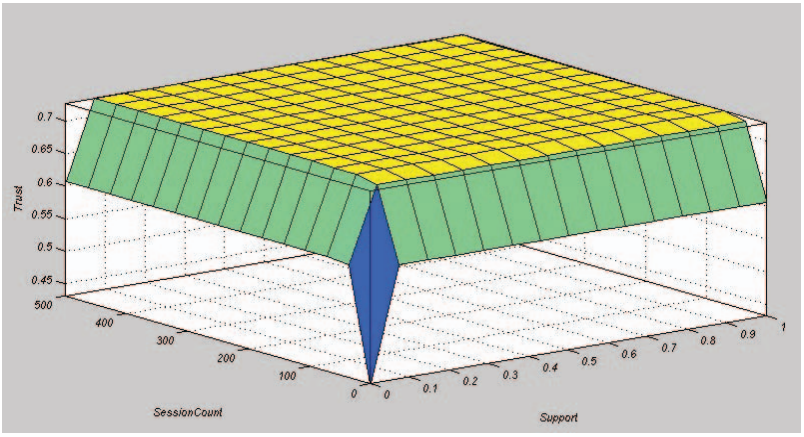

Fig. 13: The relationship between parameters support and session count with the surface view.

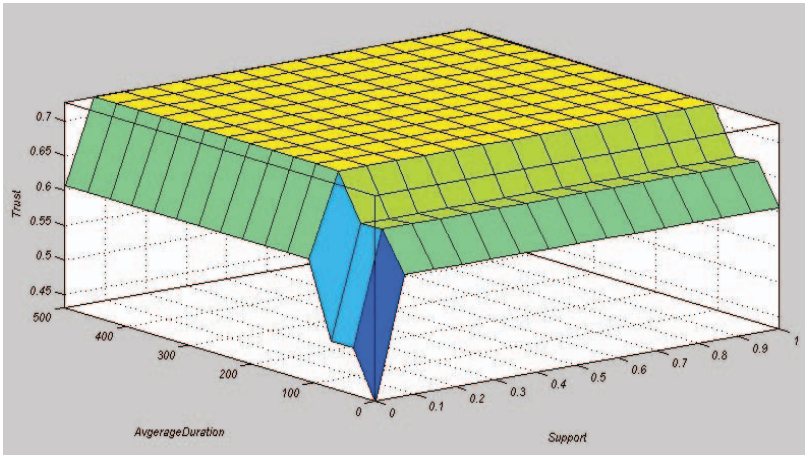

Fig. 14: The relationship between parameters support and average duration with the surface view.

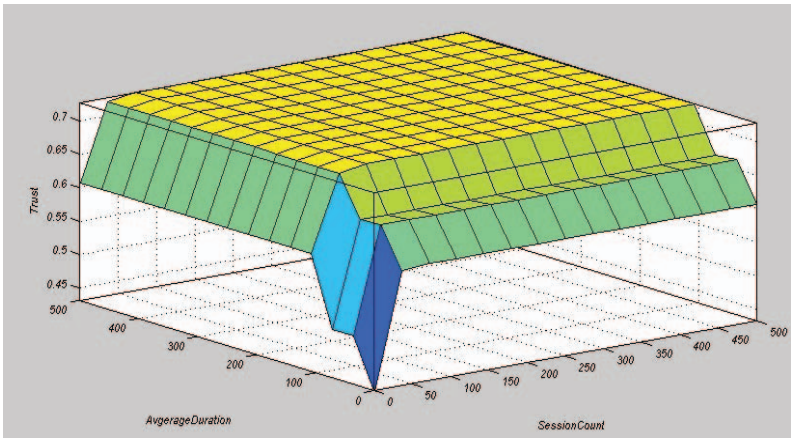

Fig. 15: The relationship between parameters session count and average duration with the surface view.

4. RESULTS AND EVALUATION

In this study, the evaluation metrics are precision, recall, and F-Score. The precision indicates the fraction of correct predictions of the proposed method among the total number of records, in other words, it determines the number of samples that are correctly predicted by the proposed method with respect to all the records. The recall determines the number of the correct predictions of the proposed method with respect to the all correct predictions.

Assume the parameter a is the value predicted by the proposed method and the parameter b shows the actual value in the dataset. The similarity between the set obtained by the proposed method and the actual set is calculated according to Eqs. 1 and 2.

$$\text{precision}\,(a,b) = \frac{a \cap b}{a} \tag{1}$$

$$\text{recall}\,(a,b) = \frac{a \cap b}{b} \tag{2}$$

In order to provide a balance, the information retrieval system uses a combination of the two above-mentioned metrics. For example, when the actual number is close to zero in the set a, the recall value is close to 1, but the precision value is close to zero. The F-score metric is a combination of recall and precision. In other words, the harmonic mean is used to determine the F-score parameter since the two metrics have the same significance. This parameter is also used for measuring the efficiency of the proposed algorithm.

$$\text{Fscore} = \frac{2 * \text{recall} * \text{precision}}{\text{arecall} + \text{precision}} \tag{3}$$

Figures 16-19 indicate the trust values obtained by the proposed fuzzy system with different number of random records 100, 500, 1000, and all records in the semantic web dataset, respectively. As can be seen, in general, the proposed fuzzy system could achieve an acceptable trust value in compared to the actual trust value.

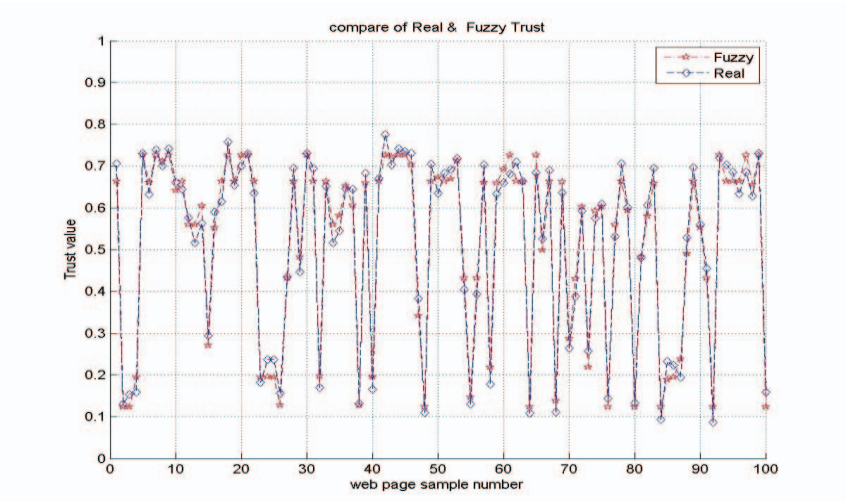

Fig. 16: Comparison between the results of the proposed method and actual data for 100 records.

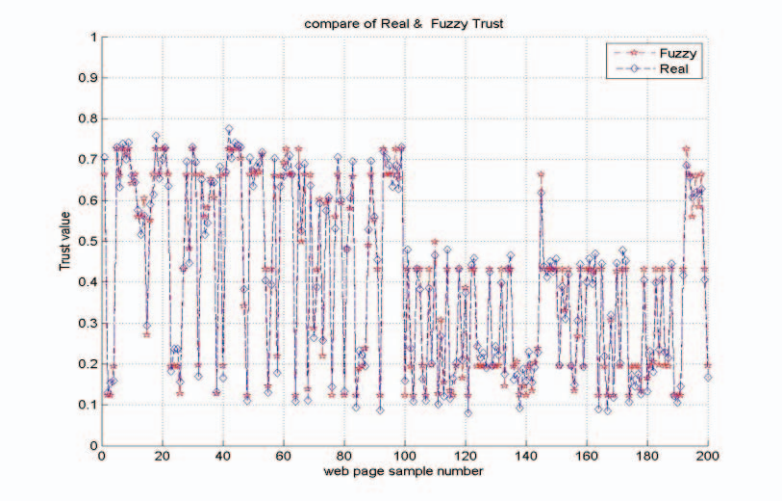

Fig. 17: Comparison between the results of the proposed method and actual data for 500 records.

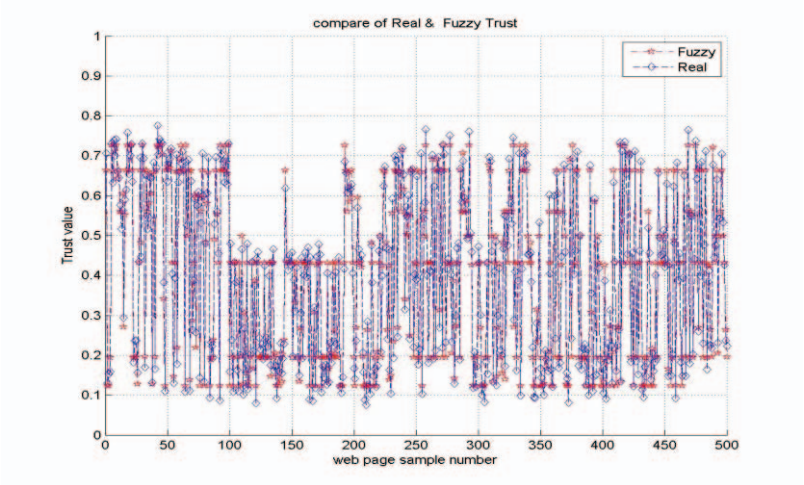

Fig. 18: Comparison between the results of the proposed method and actual data for 1000 records.

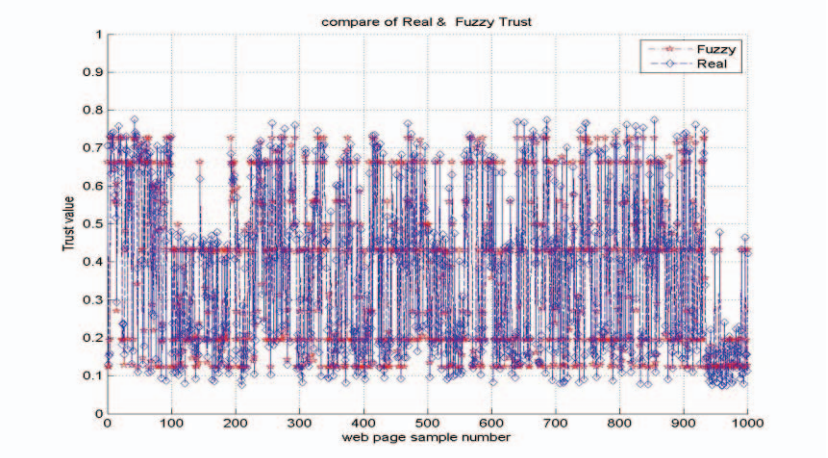

Fig. 19: Comparison between the results of the proposed method and actual data for all records.

The R is one of the accuracy evaluation metrics for the prediction methods. According to the Eq. (1), y_i represents the values predicted by the system for various input data, and x_i is the actual output values based on the dataset. $\overline{x}$ and $\overline{y}$ are the mean value of the predicted output and the mean value of the actual output for all records in the dataset, respectively. In general, the closer the value of R is to 1, the higher the precision of the predictive system.

$$R = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

The graphs in Figures 20-23 show the R value obtained by the proposed system with different number of random records 100, 500, 1000, and all records, respectively. As obvious as it seems, the closer the fuzzy output obtained by the proposed system is to the real graph, the lower the error of the proposed system.

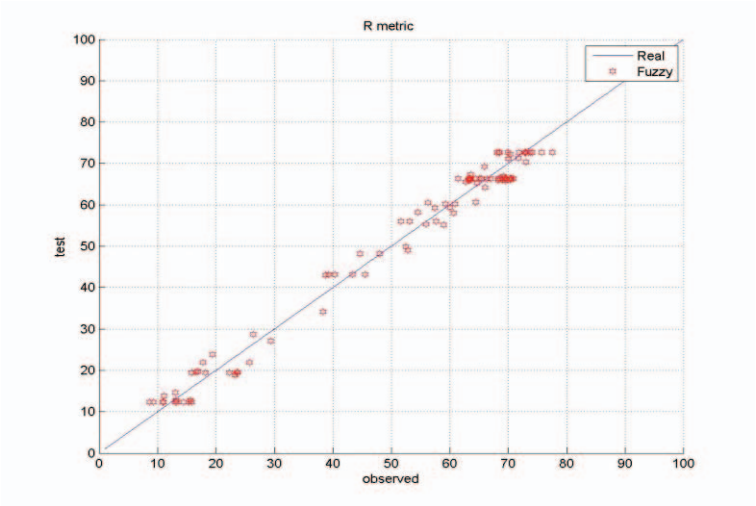

Fig. 20: Comparison between the R value obtained by the proposed fuzzy system and actual data for 100 records.

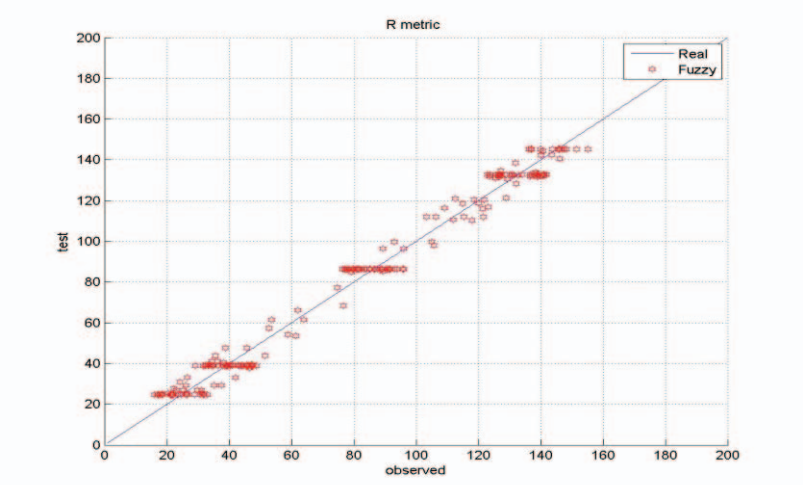

Fig. 21: Comparison between the R value obtained by the proposed fuzzy system and actual data for 500 records.

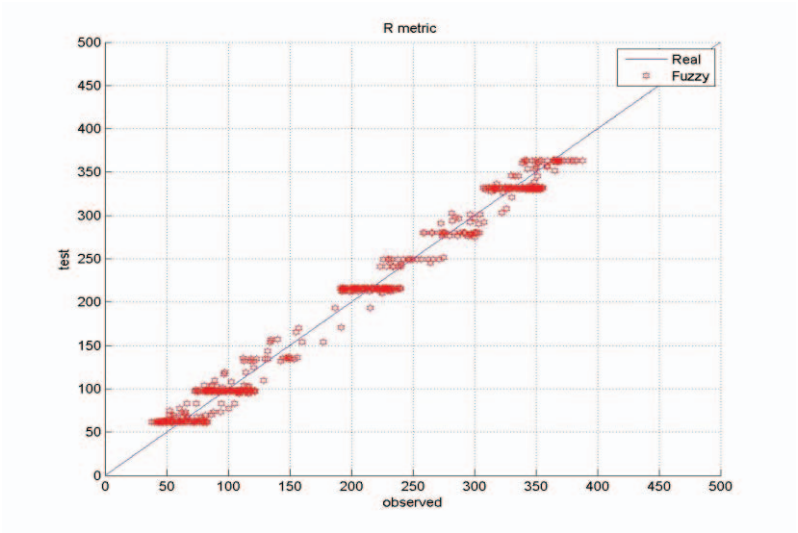

Fig. 22: Comparison between the R value obtained by the proposed fuzzy system and actual data for 1000 records.

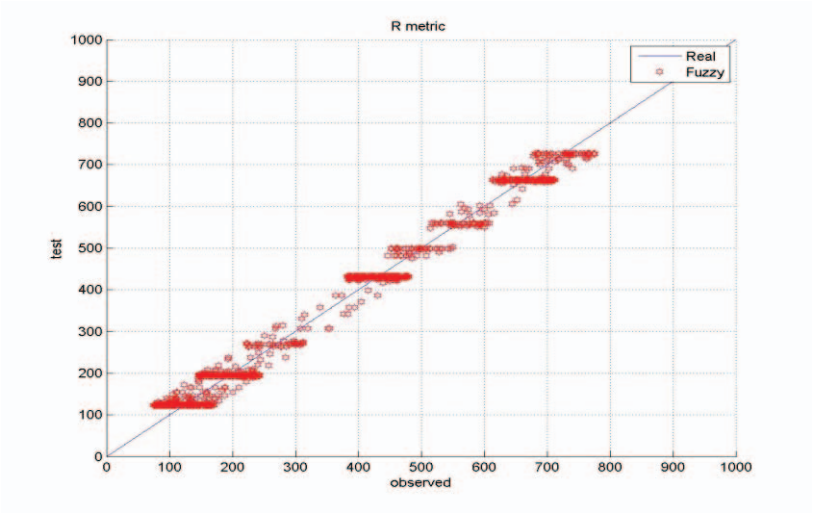

Fig. 23: Comparison between the R value obtained by the proposed fuzzy system and actual data for all records.

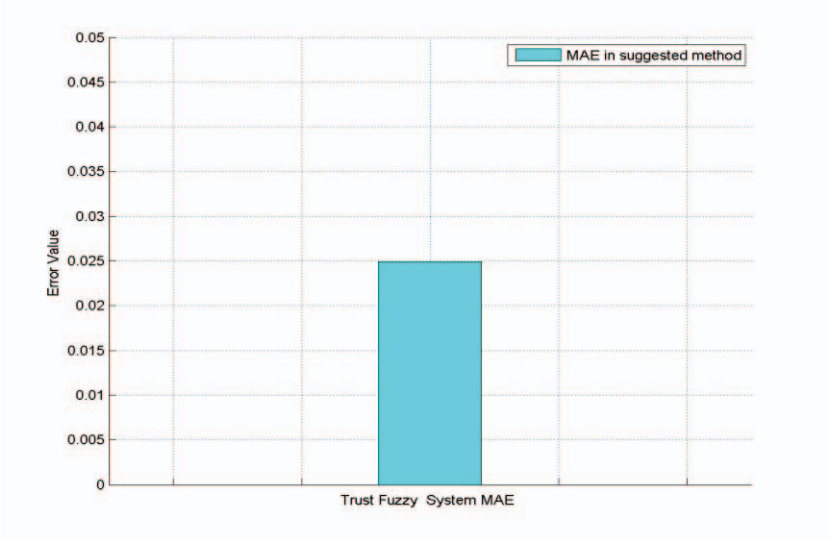

Fig. 24: The mean absolute error of the proposed fuzzy system.

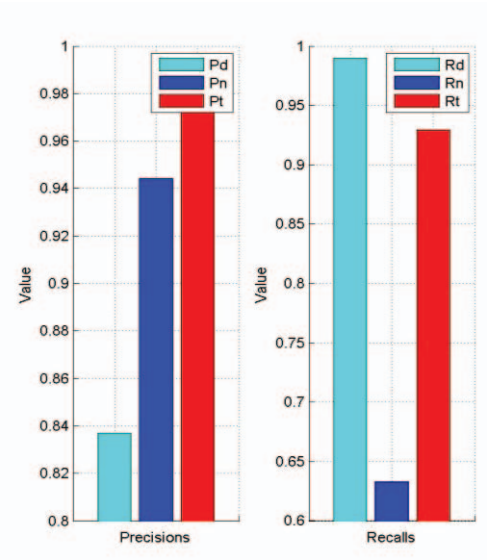

Fig. 25: Precision and recall in the proposed fuzzy system for all records.

According to Eq. (5), the mean absolute error can be used as a metric to compare the proposed fuzzy system with the actual system, where y_i is the trust value of the i^{th} sample obtained by the proposed method and x_i is the actual trust value. The comparison graph is expressed in Fig. 24.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \tag{5}$$

Precision and recall are defined based on four conditions as follows:

Trust Conditions: in which the following parameters are used to calculate the correctness:

A_t: the actual trust value in the datasets

B_t: the trust value obtained by the proposed algorithm

$$recall_t = \frac{A_t \cap B_t}{A_t} \tag{6}$$

$$precision_t = \frac{A_t \cap B_t}{B_t} \tag{7}$$

Distrust Conditions: in which the following parameters are used to calculate the correctness:

A_d: the actual distrust value in the datasets

B_d: the distrust value obtained by the proposed algorithm

$$recall_d = \frac{A_d \cap B_d}{A_d} \tag{8}$$

$$precision_d = \frac{A_d \cap B_d}{B_d} \tag{9}$$

Null Conditions: in which the following parameters are used to calculate the correctness:

A_d: the actual value of null trust in the datasets

B_d: the value of null trust obtained by the proposed algorithm

$$recall_n = \frac{A_n \cap B_n}{A_n} \tag{10}$$

$$precision_n = \frac{A_n \cap Bn}{B_n} \tag{11}$$

General Conditions: in which the following parameters are used to calculate the correctness:

$$recall = \frac{(A_t \cap B_t) + (A_d \cap B_d)}{A_t + A_d} \tag{12}$$

$$precision = \frac{(A_t \cap B_t) + (A_d \cap B_d)}{B_t + B_d}) \tag{13}$$

As shown in Fig. 25, for the precision chart, the turquoise vertical bar indicates the precision of the distrust, the dark blue one shows the null precision and the red one indicates the precision of the trust. For the recall chart, the turquoise vertical bar indicates the recall value of the distrust, the dark blue one shows the recall value of null condition and the red one indicates the recall value of the trust.

In the graph related to the total metrics, Fig. 26, the turquoise vertical bar indicates the overall precision, the dark blue one is the overall recall value and the red one is the F-Score value according to Eq. (3).

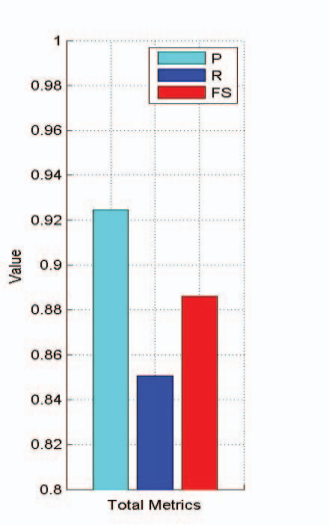

Fig. 26: Precision, recall, and F-Score in the proposed fuzzy system for all records.

5. CONCLUSION

In this study, a fuzzy system is proposed for the trust evaluation in a semantic web based on user profile data, which uses the semantic parameters such as a list of pages, user sessions, and visited pages in each session, and the time of page viewing. According to the

experiments, the proposed fuzzy method could evaluate the trust value based on user profiles with a mean absolute error of 2.5%. Accordingly, the evaluation precision of the proposed method is 97.5% in average.

In the proposed fuzzy method, the precision of distrust, null, and trust conditions are 83.5%, 94.2%, and 98.0% in average, respectively. The recall value for distrust, null, and trust conditions are 99.0%, 64.1%, and 93.0% in average, respectively. In general, therefore, the values of metrics precision, recall, and F-Score obtained by the proposed fuzzy method are 92.2%, 85.1%, and 88.3% in average, respectively.

In order to improve the precision in future work, it is recommended to use a combination of fuzzy methods, linear regression, and neural networks. Additionally, in the case of high volume of input data, it is suggested to cluster the data before applying the proposed algorithm. Also, if there are more users' features in their profiles such as interest and expertise, we suggest that users' content based categorizing according to data mining methods.

REFERENCES

[1] Artz D, Gil Y. (2007) A survey of trust in computer science and the Semantic Web. Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 5: 58–71.

[2] Arnau JP, Monedero DR, Forné J, Muñoz JL, Esparza O. (2012) Optimal tag suppression for privacy protection in the semantic Web. Data & Knowledge Engineering 81–82: 46–66.

[3] Richardson M, Argawal R, Domingos P. (2003) Trust management for the semantic web. In Proceedings of the Second International Semantic Web Conference, 351–368.

[4] Lesani M, Bagheri S. (2006) Applying and Inferring Fuzzy Trust in Semantic Web Social Networks. CSWWS, Quebec City, Canada, 23–43.

[5] Matsuo Y, Yamamoto H. (2009) Community gravity: measuring bidirectional effects by trust and rating on online social networks, in: Proceedings of the 18th International Conference on World Wide Web, ACM, New York, NY, USA Madrid, Spain, 751–760.

[6] Zolfaghar K., Aghaieb A. (2011) Evolution of trust networks in social web applications using supervised learning, Procedia Computer Science, (3) 833–839.

[7]O'Donovan J, Smyth B. (2006) Mining trust values from recommendation errors, International Journal on Artificial Intelligence Tools, (15):945–962.

[8] Victor P., Cornelis C, Cock MD, Teredesai A. (2008) Key figure impact in trust-enhanced recommender systems, AI Communications, 21(2–3):127–143.

[9] Maximilien EM, Singh MP. (2005) Multiagent system for dynamic web services selection, in: Proceedings of 1st Workshop on Service-Oriented Computing and Agent-Based Engineering (SOCABE at AAMAS), 25–29.

[10] Nguyen HT, Zhao W, Yang J. (2010) A trust and reputation model based on bayesian network for web services, in: IEEE International Conference on Web Services, IEEE, 251–258.

[11] Yahyaoui H. (2012) A trust-based game theoretical model for web services collaboration, Knowledge-Based Systems, (27):162–169.

[12] Liu FM, Wang L, Gao L, Li H, Zhao H, Men SK. (2014)A Web Service trust evaluation model based on small-world networks, Knowledge-Based Systems, (57):161-167.

[13] Sherchan W, Loke SW, Krishnaswamy S. (2006) A fuzzy model for reasoning about reputation in web services, in: Proceedings of the ACM Symposium on Applied Computing, SAC, ACM, 1886–1892.

[14] Pitsilis G, Chia PH. (2010) Does trust matter for user preferences? A study on Epinions ratings, in: The 4th IFIP International Conference on Trust Management (IFIPTM 2010) Morioka, Japan, 232–247.

[15] Shirgahi H, Mohsenzadeh M, HajSeyyedJavadi H. (2017) Trust estimation of the semantic web using semantic web clustering, Journal of Experimental & Theoretical Artificial Intelligence, 29(3): 537-556.

[16] Shirgahi H, Mohsenzadeh M, HajSeyyedJavadi H. (2018) A new method of trust mirroring estimation based on social networks parameters by fuzzy system, International Journal of Machine Learning and Cybernetics, 9(7): 1153–1168.

[17] Shekarpour S, Katebi SD. (2010) Modeling and evaluation of trust with an extension in semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 8: 26–36.

[18] Kim YA, Phalak R. (2012) A trust prediction framework in rating-based experience sharing social networks without a Web of Trust. Information Sciences, (191):128-145.