A Multi-granular Linguistic Hierarchical Model To Evaluate The Quality Of Web site Services

F. Herrera[1], E. Herrera-Viedma[1], L. Martínez[2], L.G. Pérez[2], A.G.López[1], and S. Alonso[1]

[1] Dept. of Computer Science and Artificial Intelligence.
University of Granada, 18071 - Granada, Spain
herrera,viedma@decsai.ugr.es
[2] Dept. of Computer Science.
University of Jaén, 23071 - Jaén, Spain
martin,lgonzaga@ujaen.es

Summary. The explosion in the use of Internet has contributed to arise a lot of web sites that offer many kind of services (products, information, etc). At the beginning the quality of these web sites was not too important because the most important fact was that people knew that there was a web site because there was not a big competence. But recently, there are many web sites related to the same topics in Internet and the quality of their services has become a critical factor. Different evaluation approaches for different types of web sites have been developing [2, 22, 35] in which the users provide their opinions in a predefined numerical scale to evaluate their services. Nevertheless, the information provided by users is related to their own perceptions. Usually, human perceptions are subjective and not objective, therefore to assess perceptions with precise information is not very suitable and the results are not accurate. Therefore, in this chapter we propose a linguistic quality evaluation model to evaluate the services offered by the web sites. The use of the fuzzy linguistic approach has provided good results managing human perceptions. Our proposal will consist of a hierarchical model to evaluate the services offered by general purpose web sites, such that, it will choose a few quality dimensions to be evaluated, where each one has different criteria. The users will provide their knowledge about these criteria by means of linguistic assessments. But different users can have different knowledge about the web site's criteria, so the evaluation model should take into account this point. Therefore, our model will be defined in a multi-granular linguistic information context, such that, different users can express their opinions in different linguistic term sets according to their knowledge. In order to develop this evaluation model we shall use different tools and resolution schemes based on decision techniques that are able to deal with multi-granular linguistic information.

Keywords: web quality, evaluation, linguistic variables, aggregation, decision analysis, multi-granular linguistic information, web services.

1 Introduction

Nowadays, we can assert that the World Wide Web is the largest available repository of data and services with the largest number of visitors. The World Wide Web is a distributed, dynamic, and rapidly growing information resource that has stimulated new and useful developments in areas such as digital libraries, information retrieval, education, commerce, entertainment, government, and health care.

At the beginning, the quality of the web sites did not play a key role for the companies because the most important point was that people knew that the company was in Internet by means of its web site. But recently, due to the fact, there exist a lot of web sites competing in the same area in Internet, the quality of their services has become a critical factor for the competitiveness of the companies. In such a context, Web quality evaluation tools are necessary to filter web resources in order to avoid the bad information and services that users could receive from the web.

When we talk about the quality of a web site services, we want to show how well it meets the consumers necessities and so, it is associated with consumer satisfaction [27]. Companies have noticed that offering quality services has become an essential ingredient for successful competition and they need tools that allow them to evaluate the quality of their services.

One of the first points we must fix is the meaning of quality and satisfaction. Quality can be described as conformance to requirements, while satisfaction has been defined as conformance to expectation. The ideal situation would be that there were no difference between consumer judgement of quality and experienced satisfaction. But, in fact, it is very difficult to meet all the consumers' requirements.

Due to this increasing interest in the evaluation of the services offered by the web sites we can find in the literature different models applied to specific types of web sites [1, 2, 11, 22, 28, 29, 30, 35]. However, the evaluation of the quality offered by a web site is not an easy task because the aspects of the evaluated services are related to the users own perceptions, and usually, most of these perceptions are about subjective aspects. In spite of this fact, most of the evaluation models use precise numerical assessments that it is not very suitable so, sometimes try to hide them labelling the numbers with words or symbols without semantics [22]. Nevertheless, the final results are expressed by means of numbers, so this causes a lack of precision and effectiveness in the evaluation. To overcome these problems, we propose the use of the use of the fuzzy linguistic approach [39] that has provided successful results to manage human perceptions in different topics, such as for example, "information retrieval" [4, 10, 19], "clinical diagnosis" [7], "marketing" [37], "risk in software development" [24], "technology transfer strategy selection" [5], "education" [23], "decision making" [8, 13, 34], etc.

In this chapter, we shall propose a linguistic hierarchical quality evaluation model for general purpose web sites based on decision analysis techniques

that could be specialized for specific types of web sites (e-commerce, e-bank, etc, ...). This evaluation model is user centered because it characterizes the quality of the web sites services using judgements provided by different users that surf in those web sites. It takes into account that users are assessing subjective aspects that are related to the quality of the web site services and allow the people to use words (linguistic labels) instead of numbers in their assessments. In addition, we have realized that it should offer to each user the possibility to use a linguistic term set that allow him to express his evaluation values according to their knowledge about the problem or the evaluated aspects (multi-granular linguistic context). Finally, the results generated by our evaluation model will use the same expression domain used by its users and therefore, they will be easier understood by the users and/or companies. To accomplish our aims, the evaluation model will be based on decision analysis techniques and on fuzzy tools that have been used to deal with multi-granular linguistic information [12, 14, 16, 18].

Our proposal for the linguistic multi-granular hierarchical evaluation scheme has the following steps (graphically, Fig 1):

1. *Evaluation framework:* this model will define an evaluation framework which will be composed by a few number of quality dimensions and their respective criteria that will be evaluated by the users to obtain an evaluation measure of the web site. These dimensions and criteria will be chosen according to their importance and will be chosen taking into account different points of view that we can find in the literature [2, 22, 35]. Once the dimensions and their criteria have been selected our model will gather the information provided by the users offering a multi-granular linguistic context.
2. *Evaluation process:* Once the evaluation framework has been defined, we propose a hierarchical evaluation process based on two steps:
   a) *Quality of each dimension:* the input assessments provided by the users are aggregated to obtain an evaluation assessment for each dimension. The input assessments could be expressed in different linguistic term sets with different granularity or semantics (multi-granular linguistic information). So we shall use fuzzy tools that have been used in the literature to manage and aggregate this type of information [12, 14, 16]
   b) *Global Quality Evaluation:* to compute a global quality measurement of the evaluated web site we shall calculate a satisfaction degree from the quality dimension values obtained in the before phase. This satisfaction degree will be computed by means of a weighting function that will depend on the web site we are evaluating.

This chapter is structured as follows: in the section 2 we revise some linguistic foundations we shall use in our evaluation process. In the section 3, we show in short several evaluation models for different types of web sites, in the section 4 we shall present our proposal for a multi-granular linguistic

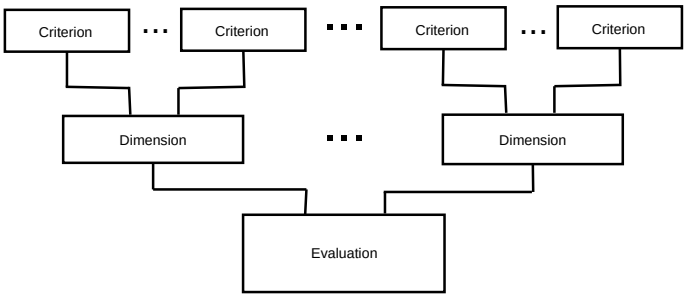

Fig. 1. Hierarchical evaluation model

hierarchical evaluation model for web site services and in the section 5 we shall show an example of this evaluation model. Eventually, some concluding remarks are pointed out.

2 Linguistic Background

In this section, we shall make a brief review of the fuzzy linguistic approach and of the fuzzy linguistic 2-tuple model. Because, we shall use them in the development of our evaluation process to manage human perceptions and multi-granular linguistic information.

2.1 Fuzzy Linguistic Approach

Usually, we work in a quantitative setting, where the information is expressed by means of numerical values. However, many aspects of different activities in the real world cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case a better approach may be to use linguistic assessments instead of numerical values. The variables which participate in these problems are assessed by means of linguistic terms [39]. This approach is adequate in some situations, for example, when attempting to qualify phenomena related to human perception, we are often led to use words in natural language. This may arise for different reasons. There are some situations where the information may be unquantifiable due to its nature, and thus, it may be stated only in linguistic terms (e.g., when evaluating the "comfort" or "design" of a car, terms like "bad", "poor", "tolerable", "average", "good" can be used [25]. In other cases, precise quantitative information may not be stated because either it is not available or the cost of its computation is too high, then an "approximate value" may be tolerated (e.g., when evaluating the speed of a car, linguistic terms like "fast",

"very fast", "slow" are used instead of numerical values). The linguistic approach is less precise than the numerical one, however some advantages may be found using it:

1. The linguistic description is easily understood by human beings even when the concepts are abstract or the context is changing.
2. Furthermore, it diminished the effects of noise since, as it is known the more refined assessment scale is, then more sensitive to noise and consequently the more error facedown it becomes.

In short, the linguistic approach is appropriated for many problems, since it allows a more direct and adequate representation when we are unable to express it with precision. Hence, the burden of qualifying a qualitative concept is eliminated.

The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables:

Definition 1 [39].- *A linguistic variable is characterized by a quintuple (H,T(H),U,G,M) in which H is the name of the variable; T(H) (or simply T) denotes the term set of H, i.e., the set of names of linguistic values of H, with each value being a fuzzy variable denoted generically by X and ranging across a universe of discourse U which is associated with the base variable u; G is a syntactic rule (which usually takes the form of a grammar) for generating the names of values of H; and M is a semantic rule for associating its meaning with each H, M(X), which is a fuzzy subset of U.*

Usually, depending on the problem domain, an appropriate linguistic term set is chosen and used to describe the vague or imprecise knowledge. The number of elements in the term set will determine the granularity of the uncertainty, that is, the level of distinction among different counting of uncertainty. In [3] the use of term sets with an odd cardinal was studied, representing the mid term by an assessment of "approximately 0.5", with the rest of the terms being placed symmetrically around it and the limit of granularity being 11 or no more than 13.

One possibility of generating the linguistic term set consists of directly supplying the term set by considering all terms distributed on scale on which total order is defined [36]. For example, a set of seven terms S, could be given as follows:

$$S = \{s_0 : none, s_1 : verylow, s_2 : low, s_3 : medium, s_4 : high, s_5 : veryhigh, s_6 : perfect\}$$

Usually, in these cases, it is required that in the linguistic term set there exist:

1. A negation operator $\mathrm{Neg}(s_i) = s_j$ such that $j = g$-i (g+1 is the cardinality).
2. A max operator: $\max(s_i, s_j) = s_i \; if \; s_i \geq s_j$.
3. A min operator: $\min(s_i, s_j) = s_i \; if \; s_i \leq s_j$

The semantics of the terms is given by fuzzy numbers. A computationally efficient way to characterize a fuzzy number is to use a representation based on parameters of its membership function [3]. The linguistic assessments given by the users are just approximate ones, some authors consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments. The parametric representation is achieved by the 4-tuple (a, b, d, c), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function [3]. A particular case of this type of representation are the linguistic assessments whose membership functions are triangular, i.e., $b = d$, then we represent this type of membership functions by a 3-tuple (a, b, c). And example may be the following:

$$P = Perfect = (.83, 1, 1)\ VH = Very\_High = (.67, .83, 1)$$
$$H = High = (.5, .67, .83)\ M = Medium = (.33, .5, .67)$$
$$L = Low = (.17, .33, .5)\quad VL = Very\_Low = (0, .17, .33)$$
$$N = None = (0, 0, .17),$$

which is graphically shown in Figure 2.

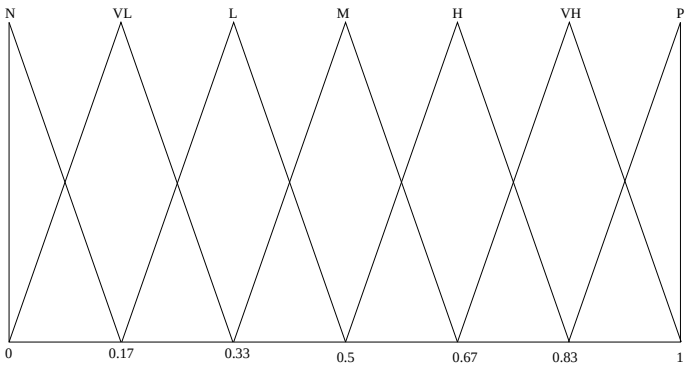

Fig. 2. A Set of Seven Terms with its Semantic

Other authors use a non-parametric representation, e.g., Gaussian functions [4].

The use of linguistic variables implies processes of computing with words such as their fusion, aggregation, comparison, etc. To perform these computations there are different models in the literature:

- *The linguistic computational model based on the Extension Principle*, which allow us to aggregate and compare linguistic terms through computations on the associated membership functions [7].
- *The symbolic method* [9]. This symbolic model makes direct computations on labels, using the ordinal structure of the linguistic term sets.

- *The 2-tuple fuzzy linguistic computational model* [14]. It uses the 2-tuple fuzzy linguistic representation model and its characteristics to make linguistic computations, obtaining as results linguistic 2-tuples. A linguistic 2-tuple is defined by a pair of values, where the first one is a linguistic label and the second one is a real number that represents the value of the symbolic translation.

In the following subsection we shall review the 2-tuple model due to the fact, that it will be the computational model we shall use in our model to deal with multi-granular linguistic information.

2.2 The 2-tuple Fuzzy Linguistic Model

This model has been presented in [14] and has shown itself as useful to deal with heterogeneous information [17, 18], such as the multi-granular linguistic information that we shall use in this paper.

This linguistic model takes as a basis the symbolic aggregation model [9] and in addition defines the concept of Symbolic Translation and uses it to represent the linguistic information by means of a pair of values called linguistic 2-tuple, (s, α), where s is a linguistic term and α is a numeric value representing the symbolic translation.

Definition 2. *Let β be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set $S = \{s_0, ..., s_g\}$, i.e., the result of a symbolic aggregation operation. $\beta \in [0, g]$, being $g+1$ the cardinality of S. Let $i = round(\beta)$ and $\alpha = \beta - i$ be two values, such that, $i \in [0, g]$ and $\alpha \in [-.5, .5)$ then α is called a Symbolic Translation.*

Graphically, it is represented in Figure 3.

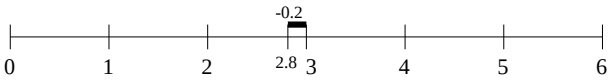

Fig. 3. Example of a Symbolic Translation

From this concept in [14] was developed a linguistic representation model which represents the linguistic information by means of 2-tuples (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-.5, .5)$.

This model defines a set of transformation functions between linguistic terms and 2-tuples, and between numeric values and 2-tuples.

Definition 3.[14] *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value supporting the result of a symbolic aggregation operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:*

$$\Delta : [0, g] \longrightarrow S \times [-0.5, 0.5)$$
$$\Delta(\beta) = \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-.5, .5) \end{cases}$$

where round is the usual round *operation, s_i has the closest index label to "β" and "α" is the value of the symbolic translation.*

Proposition 1.[14] *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and (s_i, α) be a 2-tuple. There is a Δ^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g] \subset \mathcal{R}$.*

Proof.

It is trivial, we consider the following function:

$$\Delta^{-1} : S \times [-.5, .5) \longrightarrow [0, g]$$
$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

Remark 1: From definitions 1 and 2 and from proposition 1, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consist of adding a value 0 as symbolic translation:

$$s_i \in S \Longrightarrow (s_i, 0)$$

This representation model has associated a computational model that was presented in [14]:

1. **Aggregation of 2-tuples:** The aggregation of linguistic 2-tuples consist of obtaining a value that summarizes a set of values, therefore, the result of the aggregation of a set of 2-tuples must be a linguistic 2-tuple. In [14] we can find several 2-tuple aggregation operators based on classical ones. Here we review the 2-tuple arithmetic mean and the 2-tuple weighted average operators, because we shall use them in our evaluation model:
   Definition 4:*Let $x = \{(r_1, \alpha_1), \cdots, (r_n, \alpha_n)\}$ be a set of 2-tuples, the extended Arithmetic Mean AM^* using the linguistic 2-tuples is computed as,*

$$AM^*\left((r_1, \alpha_1), \ldots, (r_n, \alpha_n)\right) = \Delta\left(\sum_{i=1}^{n} \frac{1}{n} \Delta^{-1}\left(r_i, \alpha_i\right)\right) = \Delta\left(\frac{1}{n}\sum_{i=1}^{n} \beta_i\right)$$

   Definition 5:*Let $x = \{(r_1, \alpha_1), \cdots, (r_n, \alpha_n)\}$ be a set of 2-tuples and $W = \{w_1, \cdots, w_n\}$ his associated weights. The 2-tuples weighted mean, W_AM^*, is computed as:*

$$W\_AM^*\left((r_1, \alpha_1), \ldots, (r_n, \alpha_n)\right) = \Delta\left(\frac{\sum_{i=1}^{n} \Delta^{-1}\left(r_i, \alpha_i\right) \cdot w_i}{\sum_{i=1}^{n} w_i}\right) =$$
$$= \Delta\left(\frac{\sum_{i=1}^{n} \beta_i \cdot w_i}{\sum_{i=1}^{n} w_i}\right)$$

   More linguistic 2-tuple aggregation operators were defined in [14].

2. **Comparison of 2-tuples:** The comparison of information represented by 2-tuples is carried out according to an ordinary lexico-graphic order.
   - if $k < l$ then (s_k, α_1) is smaller than (s_l, α_2)
   - if $k = l$ then
     a) if $\alpha_1 = \alpha_2$ then (s_k, α_1), (s_l, α_2) represents the same information
     b) if $\alpha_1 < \alpha_2$ then (s_k, α_1) is smaller than (s_l, α_2)
     c) if $\alpha_1 > \alpha_2$ then (s_k, α_1) is bigger than (s_l, α_2)
3. **Negation Operator of a 2-tuple:** The negation operator over 2-tuples is defined as:
$$Neg\left(s_i, \alpha\right) = \Delta\left(g - \Delta^{-1}\left(s_i, \alpha\right)\right)$$
where $g + 1$ is the cardinality of S, $s_i \in S = \{s_0, \ldots, s_g\}$.

3 An overview of Quality Evaluation Models for Web Services

In this section we shall make a short review about different quality evaluation models applied to evaluate web site services, that we can find in the literature [2, 33, 35]. Our aim is to show their working and their problems to deal with human perceptions, in order to use several concepts of these models and to overcome their limitations in our proposal.

3.1 Evaluation of web-based decision support systems

Web-based decision support systems are being employed by organizations as decision aids for employees as well as customers. A common usage of web-based DSS (decision support systems) has been to assist customers configure product and service according to their needs. These systems allow each customer to design their own product by choosing from a group of alternatives. Some examples of web-based DSS can be found in www.dell.com, www.ibm.com, www.landsend.com and www.vermontteddybears.com.

In [2] we can see a study about the quality of the services (user satisfaction) of web-based DSS. In this model, the quality of the service is determined by three dimensions where each **dimension** has several *criteria* to evaluate the web-based DSS:

- **System quality**, whose criteria are:
  - *System reliability, Convenient to access, System easy of use, System flexibility.*
- **Information quality**:
  - *Information accuracy, Information completeness, Information relevance, Information content needs, Information timeless.*
- **Information presentation**:
  - *Presentation graphics, Presentation colour, Presentation style, Navigationally efficient.*

The process for reaching a solution in their evaluation problem is as follows:

1. *Data collection:* users provide their opinions about the criteria using numerical scale, in spite of most of them are qualitative aspects.
2. *Evaluation process:* The opinions provided by the users are combined in order to evaluate each dimension and finally a global quality measurement of the web-based DSS is obtained. All the results are expressed by means of numerical values

The results are analyzed using the structural equation model (SEM) approach to study the correlation among the dimensions and the quality, and among the criteria and the quality.

3.2 The Extended Web Evaluation Model (EWAM)

The EWAM (Extended Web Evaluation Model) [33] defines an evaluation grid with a set of criteria for appraising the quality and success of existing e-commerce applications (e.g. assessed some Australian grocery web sites in [22]). The EWAM examines the three classic transaction phases of electronic markets, which include information, agreement, and settlement phases. A fourth element, the community component, is integrated as a link between the actual purchase transaction and the necessary trust relationship in the virtual realm.

In [33] was presented a study based on the EWAM and established the questions about quality, satisfaction or success of a e-commerce application must be allotted to one of the four transaction phases of electronic markets (information, agreement, settlement, and after-sale), to the community component, or to the category "Final Section" which concerns all phases. For instance, in [22] the *criteria* that they used for each **dimension** were:

- **Information phase:** whose criteria are:
  - *Accessibility of the web site, Structure of the contents, Quantity of information, Quality of the content, Passing on price benefits.*
- **Agreement phase:**
  - *Design of the ordering procedure, Models and methods of pricing.*
- **Settlement phase:**
  - *Integration of generic services, Tracking and tracing.*
- **After-sale phase:**
  - *Access to customer support, Performance of customer support.*
- **Community component:**
  - *Sharing opinions.*
- **Final section:**
  - *Availability of the system, The design of the user interface, Increasing productivity by gaining time, The trustworthiness of the web site.*

The dimensions in [33] are formulated in general terms and are valid in every sector but are differentiated by their importance ratings. In order to take into account the differences between the individual sectors, assessor(s) provides weights corresponding to the different sector profiles and their relevance in the sector.

The evaluation of an e-commerce web site with EWAM begins by assigning the concerned web site to a sector. Then, there are two steps involved in the evaluation:

1. *Subjective importance of every dimension:* The assessor(s) declares the subjective importance of a dimension. This importance is recorded on a scale of "unimportant" (-2), "less important" (-1), "important" (+1) and "very important" (+2).
2. *Evaluation of all web site in the concerned sector:* In [22] they used a rating based on a five-point scale: from -2 (very bad) to +2 (very good) to evaluate every question of each criterion. The evaluation process has the following steps:
   a) *Data collection:* EWAM gathers the input information from each user about the satisfaction on each criterion. In EWAM we use a numerical scale although the method offers labels to assess the criteria.
   b) *Evaluation process:* The input information is combined by an aggregation operator and we obtain a global assessment for each dimension. And after according to the importance provided by the assessor(s) to each dimension a global evaluation value is obtained. All the results are expressed by means of numerical values, although seemingly the users provides linguistic information but in fact they are providing numerical values.

3.3 The Servqual scale adapted to electronic services

In [35] is proposed an "e-satisfaction" model that evaluate the users satisfaction on quality services of information search and purchase web sites based on the Servqual scale [31]. This scale identified six dimensions representing Internet information search satisfaction (reliability, convenience, entertainment, assurance, site design, virtual environment) and three dimensions representing satisfaction with Internet purchase experience (security,product offer,convenience). Each **dimension** is defined by a group of *criteria* that describe the e-satisfaction of the users with that dimension. In [35] are used the following dimensions and criteria:

- For Internet information search satisfaction:
  - **Information reliability,** whose criteria are:
    - *Up-to-date information, Information depth, Search result, Uncluttered web pages, Easy search paths, Easiness in comparing information.*
  - **Convenience:**

  - · *Economy of time spent, Effort spent, Easy access, Fast information transmission, Interaction capacity.*
  - **Entertainment:**
    - · *Interesting places to visit, Pleasant browsing, Entertainment and leisure, Easy browsing, Information diversity.*
  - **Assurance:**
    - · *Data transmission assurance and Privacy.*
  - **Site design:**
    - · *Advertising contents and Attractive presentation.*
  - **Virtual Environment:**
    - · *Capacity of simulating reality and Personal contact absence.*
- For Internet purchase experience satisfaction:
  - **Security:**
    - · *Payment security, Trust in supplier, Privacy of purchase, Personal-sales absence, Pleasant way of buying.*
  - **Product Offer:**
    - · *Easy to compare products' characteristics, Diversity of products' brands, Product guarantee, Price reduced products, Possibility to return.*
  - **Convenience:**
    - · *Fast delivery and Easy way of buying.*

Values on satisfaction are interpreted directly as performance measures (un-weighted) and they evaluate each criterion using a five-point scale (5=High satisfaction,. . . ,1=Low Satisfaction).

The evaluation process is composed by the following phases:

1. *Data collection:* the users provide their opinions about the criteria using numerical values. Each criterion using a five-point scale (5=High satisfaction,. . . ,1=Low Satisfaction).
2. *Evaluation process:* the opinions provide by the users are combined to obtain an evaluation assessment for each dimension. Finally, in this case a global evaluation assessment is obtained using an un-weighted aggregation operator. These results are expressed by means of numerical values.

3.4 Current Web Evaluation Methods: Problems and Working

Reviewing the before web evaluation models we have realized that these models are user centered. Due to the fact that the evaluation of the web site depends on the opinions provided by the users that use the web site.

To evaluate the web site services all the models choose a set of dimensions with several outstanding criteria that have to be assessed by the users according to their perception.

These models present different problems [20], such as, they try to avoid explicit numerical values, because this type of information is not suitable to assess human perceptions. Hence, they try to hide the numbers behind a scale

of labels or symbols, but in fact they are using numbers and in addition all the users are forced to use the same scale for all the criteria, despite they can have different knowledge about the criteria or about an specific criterion this causes a lack of expressiveness that means a loss of information and accuracy in the results of the evaluation process. Besides, the final results are expressed by means of numerical values that are far from the user expression domain so sometimes the results are difficult to understand by the users and then a feedback cycle to improve the web site is almost impossible.

We shall propose an linguistic evaluation model to solve the problem of lack of expressiveness using multi-granular linguistic context, in which each user can provide their opinions in a linguistic term set according to his knowledge. In addition, this model will deal with linguistic information and the results will be expressed by means of linguistic values.

This model will have an hierarchical structure to evaluate separately each dimension and from these evaluations to obtain a global evaluation assessment for the web site.

4 A Linguistic Multi-criteria Hierarchical Evaluation Model For Web sites Services

The aim of this paper is to develop an user centered hierarchical evaluation model for web sites services, in which, the users can express their opinions about the web sites by means of linguistic terms that can be assessed in different linguistic term sets.

In short, we can define mathematically our problem as an evaluation process in which a set of users (experts) $E = \{e_1, \cdots, e_n\}$ will evaluate a web site, W, providing their opinions about a set of quality dimensions, $D = \{d_1, \cdots, d_q\}$, such that each dimension, d_i, has a set of criteria, $C_i = \{c_{1i}, \cdots, c_{ti}\}$, to be evaluated . Therefore, each time a web site is evaluated every expert, e_k, provides his assessments about the different criteria by means of an utility vector:

$$e_k \rightarrow \{u_{11}^k, \ldots, u_{t1}^k \ldots u_{1q}^k, \ldots, u_{tq}^k\},\ u_{tq}^k \in S^k$$

Where u_{tq}^k is the linguistic assessment provided by the expert e_k for the criterion c_{tq} that belongs to the dimension d_q . Due to the fact that our model is defined in a multi-granularity linguistic context, we assume that each user may use a different linguistic term set for each criterion to evaluate the web site services according to their knowledge about the problem. Therefore, each user, e_k, can express his opinions for criterion, c_{tq}, in a linguistic term set $S_{tq}^k = \left\{s_0^{k_{tq}}, \ldots, s_l^{k_{tq}}\right\}$ where $l+1$ is the cardinality of S_{tq}^k.

In the following subsections, we shall present the evaluation framework we shall use to evaluate the web site services, after we shall present our multi-granular linguistic hierarchical evaluation model.

4.1 Evaluation Framework

To develop an evaluation model, first of all we have fix which is the evaluation framework we shall use to evaluate the web sites services.

In the section 3, we have reviewed different evaluation models for different web-based services. All of them chose a few dimensions to evaluate the quality of the web site. Each dimension was evaluated using different criteria, so the dimensions as the criteria depended on the topic related to the evaluated web site. Nevertheless our aim is to define an evaluation model for general purpose web sites, so taking into account the models presented in [1, 2, 6, 21, 22, 26, 33, 35, 38] we have to choose the dimensions and their criteria that our evaluation model will use to define the evaluation framework of our problem.

Our general purpose evaluation model shall use the following **dimensions** and *criteria*:

- **Entertainment**: this criterion is related to amusement and pleasure contents. The criteria we use to assess this dimension are:
  - *Interesting place to visit.*
  - *Pleasant browsing.*
  - *Entertainment and leisure*
  - *Easy browsing.*
  - *Information diversity.*
- **Convenience**: it is related to speed, easy access to information at low cost. The criteria are:
  - *Economy of time spent.*
  - *Effort spent.*
  - *Easy access.*
  - *Fast information transmission.*
  - *Interaction capacity.*
  - *Fast delivery*
  - *Easy way of buying.*
- **Information reliability**: it is about reliability of the information. Information reliability is related with diversity, depth and actuality of information contents. And its criteria are:
  - *Up-to-date information.*
  - *Information depth.*
  - *Search result.*
  - *Uncluttered web pages.*
  - *Easy search paths.*
  - *Easiness in comparing information.*
- **Security and assurance**: security perception and privacy "assurance" are known to have a big impact in user satisfaction of quality web services. Its criteria are:
  - *Payment security.*
  - *Trust in supplier.*

  - *Privacy of purchase.*
  - *Data transmission assurance.*
  - *Privacy.*
- **Site Design**: it depends on functional and attractive elements: ease of browsing, a standard language use, interface design. Its criteria are:
  - *Advertising contents.*
  - *Attractive presentation.*
- **Virtual Environment**: it is used to minimize the absence of human contact and amusement associated to shopping. Its criteria are:
  - *Capacity of simulating reality.*
  - *Personal contact absence.*
  - *Personal-sales absence.*
- **Product Offer**: it is concerning product diversity and available brands. And its criteria are:
  - *Easy to compare products' characteristics*
  - *Diversity of product's brands*
  - *Product guarantee.*
  - *Price reduced products.*
  - *Possibility to return.*

Once the dimensions and criteria to evaluate the web site services have been chosen, the assessments provided by the experts can be expressed in different linguistic term sets S_{tq}^k according to their knowledge and the characteristics of the criterion.

This evaluation framework offers several advantages with regards to the evaluation models reviewed in the Section 3:

- The dimensions and criteria used in this framework are general enough to evaluate any kind of web site. Due to the fact we want to develop a general purpose evaluation model for web site services.
- This framework models the information provided by the users by means of linguistic information according to the fuzzy linguistic approach that it is more suitable than numerical values to model human perceptions. In addition, it offers to the users the possibility of using different linguistic term sets to provide their information and in this way avoid lack of expressiveness, lose of information, and probably bad results.

4.2 Evaluation process: Evaluating the quality of the web services

Once we know the evaluation framework we are going to use in our problem, we are going to present the hierarchical process we propose to evaluate the web site services.

Our proposal to evaluate the web sites services consist of a hierarchical process with the following phases (graphically, Fig 4) :

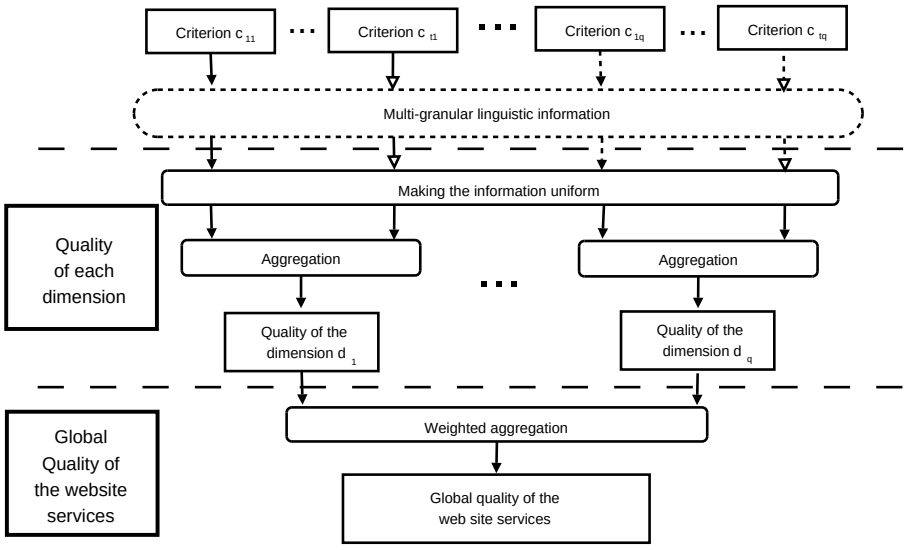

Fig. 4. Evaluation process

1. **Quality of each dimension**: in this phase we want to obtain an evaluation for each dimension of our evaluation framework. To do so, we have to aggregate the criteria belong to the dimension, but the difficulty comes from the multi-granular linguistic context in which it has been defined our framework because there are not standard aggregation operators for this type of information. Therefore this aggregation process will consist of the following steps:
   a) *Making the information uniform:* The input information provided by the users could be expressed in different linguistic term sets with different granularity or semantics (multi-granular linguistic information).Therefore to combine the input assessments, we need to unify the input information, that it is multi-granular, into an unique expression domain. We shall unify this information in a Basic Linguistic Term Set (BLTS) by means of fuzzy sets and afterwards, we shall transform these unified input information into linguistic 2-tuples expressed in the BLTS.
   b) *Aggregation phase:* it combines the unified input assessments provided by the users to obtain a collective value for each dimension.
2. **Global Quality of the web site services**: we want to obtain a global evaluation of the web site services. To do so, we shall aggregate the evaluation assessments obtained for each dimension of quality. In this phase, the aggregation will be carried out by means of a weighting aggregation operator, where the weights assigned to each dimension will depend on the

evaluated web site. And with these weights we can annul one or several dimensions in certain type of web sites. So we can use the same framework for general purpose web sites.

In the next subsections we present each phase of the evaluation model in further detail.

Quality of each Dimension

We want to obtain a collective assessment on a dimension according to the individual opinions provided by the users regarding the different criteria (assessed in multi-granularity linguistic term sets). We shall aggregate the information according to the following steps:

1. Making the information uniform by means of fuzzy sets.
2. Transforming into 2-tuple.
3. Calculating an evaluation assessment for the dimension.

Now we present the working of each step in detail.

1. Making the Information Uniform

With a view to manage the information we must make it uniform, i.e., the multi-granular linguistic information provided by the users must be transformed into a unified linguistic term set, called BLTS and denoted as S_T.

Before defining a transformation function to unify the multi-granular linguistic information into this BLTS, S_T, we have to decide how to choose S_T. We consider that S_T must be a linguistic term set which allows to express a quality scale easy to understand and maintain the uncertainty degree associated to each expert and the ability of discrimination to express the performance values. So in our case, we propose the following linguistic term set as, BLTS:

$$S_T = \{N, VL, L, M, H, VH, P\},$$

whose semantics has been shown in the Figure 2.

We shall unify the multi-granular linguistic information by means of fuzzy sets in the BLTS. The process of unifying the information involves the comparison between fuzzy sets. These comparisons are usually carried out by means of a measure of comparison. We focus in measures of comparison which evaluate the resemblance or likeness of two objects (fuzzy sets in our case) [32]. For simplicity, in this paper we shall choose a measure based on a possibility function $S(A, B) = \max_x \min(\mu_a(x), \mu_B(x))$, where μ_A and μ_B are the membership function of the fuzzy sets A and B respectively.

The next step in this process of unifying the information is to define a transformation function that we allow us to express the input information in the BLTS. We shall define a transformation function that will unify the input linguistic multi-granular information by means of fuzzy sets in the BLTS:

Definition 6. *Let $S = \{l_0, \ldots, l_p\}$ and $S_T = \{s_0, \ldots, s_g\}$ be two linguistic term sets. Then, a linguistic transformation function, τ_{SS_T}, is defined as:*

$$\tau_{SS_T} : S \rightarrow F(S_T)$$
$$\tau_{SS_T}(l_i) = \{(s_k, \gamma_k^i) \,/\, k \in \{0, ..., g\}\}, \forall l_i \in S$$
$$\gamma_k^i = \max_y \min\{\mu_{l_i}(y), \mu_{s_k}(y)\}$$

where $F(S_T)$ is the set of fuzzy sets defined in S_T, and $\mu_{l_i}(\cdot)$ and $\mu_{s_k}(\cdot)$ are the membership functions of the fuzzy sets associated with the terms l_i and s_k, respectively.

The result of τ_{SS_T} for any linguistic value of S is a fuzzy set defined in the BLTS, S_T. Therefore, after unifying the input information with this transformation function the opinions provided by the experts are expressed by means of fuzzy sets in the BLTS.

Remark 2: In the case that the linguistic term set, S, of the non-homogeneous contexts let be chosen as BLTS, then the fuzzy set that represents a linguistic term will be all **0** except the value correspondent to the ordinal of the linguistic label that will be **1**.

Example. Let $S = \{l_0, l_1, \ldots, l_4\}$ and $S_T = \{s_0, s_1, \ldots, s_6\}$ be two term set, with 5 and 7 labels, respectively, and with the following semantics associated:

$$\begin{array}{ll} l_0 = (0, 0, 0.25) & s_0 = (0, 0, 0.16) \\ l_1 = (0, 0.25, 0.5) & s_1 = (0, 0.16, 0.34) \\ l_2 = (0.25, 0.5, 0.75) & s_2 = (0.16, 0.34, 0.5) \\ l_3 = (0.5, 0.75, 1) & s_3 = (0.34, 0.5, 0.66) \\ l_4 = (0.75, 1, 1) & s_4 = (0.5, 0.66, 0.84) \\ & s_5 = (0.66, 0.84, 1) \\ & s_6 = (0.84, 1, 1) \end{array}$$

The fuzzy set obtained after applying τ_{SS_T} for l_1 is (see Fig. 5):

$$\tau_{SS_T}(l_1) = \{(s_0, 0.39), (s_1, , 0.85), (s_2, 0.85), (s_3, 0.39) \\ (s_4, 0), (s_5, 0), (s_6, 0)\}$$

2. Transforming into 2-tuple

In this phase, we transform the unified information expressed by means of fuzzy sets on the BLTS into linguistic 2-tuples in the BLTS, to facilitate the computation of the satisfaction assessment. In [15] was presented a function χ that transformed a fuzzy set into a numerical value and using Δ we transformed this numerical value in a 2-tuple. In this contribution we have redefined χ in a way that transforms directly a fuzzy set in $F(S_T)$ into a 2-tuple:

$$\chi : F(S_T) \rightarrow S_T x\, [-0.5, 0.5)$$

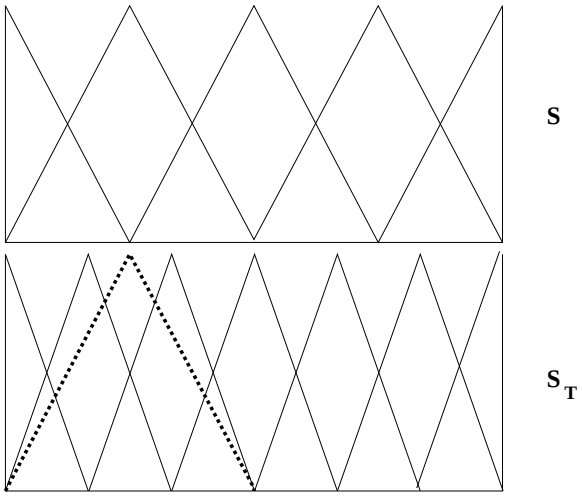

Fig. 5. Transforming $l_1 \in S$ into a Fuzzy Set in S_T

$$\chi\left(F\left(S_T\right)\right) = \chi\left(\left\{\left(s_j, \gamma_j\right), j = 0, \ldots, g\right\}\right) = \Delta\left(\frac{\sum_{j=0}^{g} j\gamma_j}{\Sigma_{j=0}^{g}\gamma_j}\right) =$$
$$= \Delta\left(\beta\right) = \left(s, \alpha\right)$$

After applying χ to the fuzzy sets in the BLTS obtained in the before step, we shall obtain the linguistic 2-tuples in the BLTS that express the opinions provided by the users.

Example.

We want to transform the fuzzy set $(0, 0, 0, .41, 1, .19, 0)$ in the BLTS to a 2-tuple (figure 2):

$$\chi\left(\left(0, 0, 0, .41, 1, .19, 0\right)\right) = \Delta\left(\frac{\sum_{j=0}^{6} j\gamma_j}{\Sigma_{j=0}^{6}\gamma_j}\right) =$$

$$\Delta\left(4.33\right) = \left(H, .33\right)$$

and the result is $(H, .33)$.

3. Calculating an evaluation assessment for each dimension

Our objective is to obtain an evaluation value for each dimension according to the opinions provided by all the users for its criteria. At this moment, these values are expressed by means of linguistic 2-tuples in the BLTS.

Therefore, to reach our objective we follow the next steps:

1. Computing collective values for each criterion: Each dimension, d_i, has a set of criteria, $C_i = \{c_{1i}, \cdots, c_{ti}\}$, so first of all we shall compute a collective value for each criterion according to all the users. In this proposal

we shall use a non-weighted aggregation operator as the arithmetic mean for 2-tuples (Definition 4), but weighted operator could considered in the future.
Therefore, the collective value for the criterion (CVC), $c_{ti} \in C_i$, will be computed as:

$$CVC_{ti} = AM^*(\left(u_{ti}^k, \alpha\right),\ k = 1...n) = (u_{ti}, \alpha)$$

2. Computing an evaluation assessment for each dimension: So far, we have computed a collective value for every criterion that belongs to d_i. Now, we want to obtain an evaluation assessment for each dimension. To do so, we shall aggregate the collective values of its criteria by means of an aggregation operator. As well as before, we shall use a non-weighted aggregation operator as the arithmetic mean for 2-tuples, although could be considered a weighted operator in the future.
Therefore, to obtain an evaluation assessment for a dimension (ED), d_i, will be computed as:

$$ED_i = AM^*((u_{ji}, \alpha),\ j = 1...t) = (u_i, \alpha)$$

So now, we have an evaluation assessment for each dimension of quality, d_i, of our evaluation model. And we can evaluate separately each dimension to improve just certain drawbacks of our services. In order to improve the global evaluation of our web site services.

Global Quality of the web site services

Our final aim it is to obtain a global evaluation assessment, EAW, for the web sites services we are evaluating. To do so, we shall aggregate the quality assessments obtained for each dimension. In this case we shall use a weighted aggregation operator, because although our model is to evaluate general purpose web site, we think that depends on the specific web site different dimensions could have different importance, even some of them their value can be null in some occasions. So, to obtain the evaluation of the web site services some expert/s provides a weighting vector that indicates the importance of each dimension, $W = \{w_1, ..., w_q\}$. Afterwards we shall apply the 2-tuple linguistic weighting average operator (Definition 5):

$$EAW = W\_AM^*((u_i, \alpha), i = 1, ..., q) = (u, \alpha)$$

We have obtained a global linguistic evaluation for the quality of the web site services that is expressed in the BLTS (linguistic evaluation scale).

In the next section we shall apply this evaluation model to evaluate a lecturer's web site, in order to know the satisfaction of their students regarding the services offered by the web site.

5 Application: Evaluating a Lecturer's Web site

Let us suppose that we want to evaluate Lecturer's web site. We have four students that will provide us their opinions about the web site services. Every student could use a different linguistic term set for each criterion, but, to simplify the problem and show the resolution process easily we shall assume that every student choose a linguistic term set to evaluate all the criteria according to his/her knowledge:

- The first user has chosen the linguistic term set A of 3 labels (figure 6).
- The second user has chosen the linguistic term set B of 5 labels (figure 7).
- The third user has chosen the linguistic term set C of 7 labels (figure 8).
- The fourth user has chosen the linguistic term set D of 9 labels (figure 9).

In the table 1 we can see the semantics of the linguistic term sets.

Lingustic term set A	Linguistic term set B	Linguistic term set C	Linguistic term set D
$a_0 = (0, 0, 0.5)$	$b_0 = (0, 0, 0.25)$	$c_0 = (0, 0, 0.16)$	$d_0 = (0, 0, 0.12)$
$a_1 = (0, 0.5, 1)$	$b_1 = (0, 0.25, 0.5)$	$c_1 = (0, 0.16, 0.34)$	$d_1 = (0, 0.12, 0.25)$
$a_2 = (0.5, 1, 1)$	$b_2 = (0.25, 0.5, 0.75)$	$c_2 = (0.16, 0.34, 0.5)$	$d_2 = (0.12, 0.25, 0.37)$
	$b_3 = (0.5, 0.75, 1)$	$c_3 = (0.34, 0.5, 0.66)$	$d_3 = (0.25, 0.37, 0.5)$
	$b_4 = (0.75, 1, 1)$	$c_4 = (0.5, 0.66, 0.84)$	$d_4 = (0.37, 0.5, 0.62)$
		$c_5 = (0.66, 0.84, 1)$	$d_5 = (0.5, 0.62, 0.75)$
		$c_6 = (0.84, 1, 1)$	$d_6 = (0.62, 0.75, 0.87)$
			$d_7 = (0.75, 0.87, 1)$
			$d_8 = (0.87, 1, 1)$

Table 1. Semantic of the linguistic term sets A, B, C and D

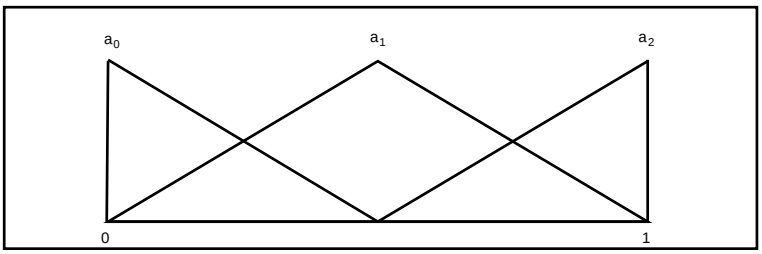

Fig. 6. Semantic of the lingustic term set A

The opinions provided by the students for the dimensions and criteria proposed in the section 4.1 are:

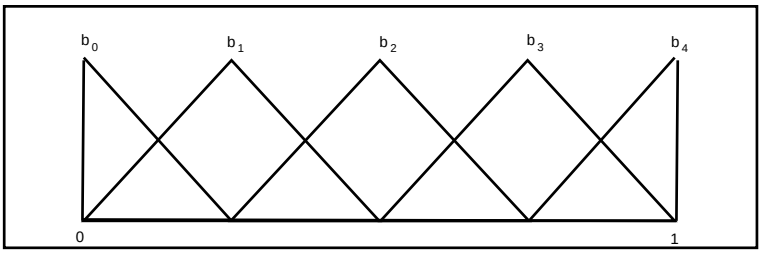

Fig. 7. Semantic of the linguistic term set B

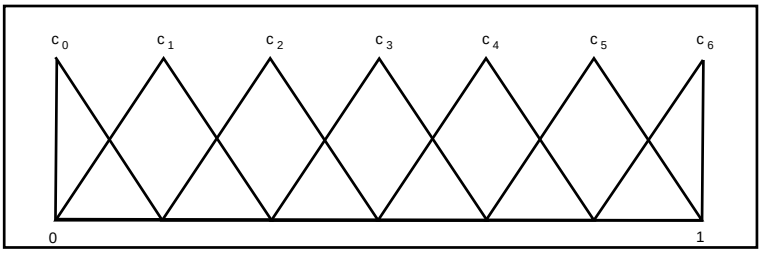

Fig. 8. Semantic of the linguistic term set C

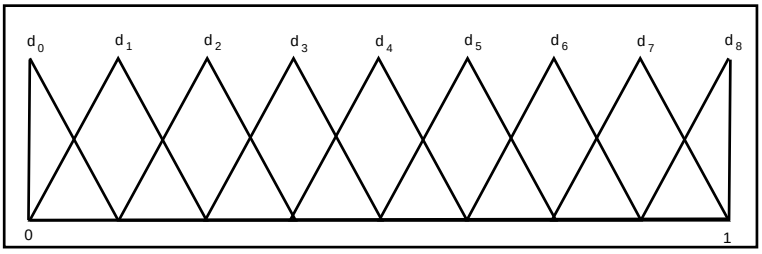

Fig. 9. Semantic of the linguistic term set D

$$e_1 = \{\mathbf{a_2}, \mathbf{a_1}, \mathbf{a_0}, \mathbf{a_2}, \mathbf{a_2}, a_0, a_1, a_1, a_1, a_2, a_0, a_1, a_0, a_0, a_2, a_2, a_2, a_2, a_0, a_0, a_0, a_0, a_0, a_1, a_2, a_2, a_1, a_2, a_0, a_0, a_0, a_0, a_0\}$$

$$e_2 = \{\mathbf{b_1}, \mathbf{b_4}, \mathbf{b_0}, \mathbf{b_0}, \mathbf{b_0}, b_2, b_3, b_4, b_2, b_2, b_3, b_0, b_4, b_0, b_4, b_1, b_4, b_1, b_0, b_0, b_0, b_0, b_0, b_1, b_3, b_3, b_1, b_2, b_0, b_0, b_0, b_0, b_0\}$$

$$e_3 = \{\mathbf{c_4}, \mathbf{c_0}, \mathbf{c_5}, \mathbf{c_1}, \mathbf{c_6}, c_0, c_6, c_2, c_1, c_3, c_2, c_2, c_1, c_1, c_3, c_5, c_1, c_4, c_0, c_0, c_0, c_0, c_0, c_0, c_3, c_3, c_1, c_6, c_0, c_0, c_0, c_0, c_0\}$$

$$e_4 = \{\mathbf{d_7}, \mathbf{d_7}, \mathbf{d_1}, \mathbf{d_2}, \mathbf{d_8}, d_4, d_3, d_7, d_2, d_2, d_0, d_3, d_7, d_4, d_6, d_5,$$
$$d_0, d_1, d_0, d_0, d_0, d_0, d_0, d_4, d_2, d_3, d_0, d_3, d_0, d_0, d_0, d_0, d_0\}$$

We shall resolve this evaluation problem but for a better comprehension of the evaluation process we just show the main operations over the dimension Entertainment (the first five assessments of every student that are in bold).

Once we have obtained the opinions from the students we have to apply the evaluation model to obtain: firstly, the quality of each dimension and afterwards we shall compute the global quality of the web site.

Quality of each dimension

We want to remark that in this case the BLTS we have chosen for our model is the linguistic tern set C of the multi-granular linguistic context used in this example.

To obtain the quality value for each dimension we shall apply the process presented in the section 4.2:

1. Make the information uniform: We unify the input information by means of fuzzy sets in the BLTS and afterwards, we shall transform these unified information into linguistic 2-tuples expressed in the BLTS.
   a) Transforming into the BLTS: We use the functions τ_{AC}, τ_{BC}, τ_{CC} and τ_{DC} (definition 6). The results obtained for the criteria belonging to the dimension entertainment are:

$$e_1 = \{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.25}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0}), (\mathbf{0.25}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0.25}), (\mathbf{0}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0.25}, \mathbf{0}, \mathbf{0}, \mathbf{0}),$$
$$(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.25}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.25}, \mathbf{0.5}, \mathbf{0.5}, \mathbf{0}), \ldots\}$$

$$e_2 = \{(\mathbf{0.4}, \mathbf{0.8}, \mathbf{0.8}, \mathbf{0.4}, \mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.2}, \mathbf{0.6}, \mathbf{0}), (\mathbf{0}, \mathbf{0.6}, \mathbf{0.2}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}),$$
$$(\mathbf{0}, \mathbf{0.6}, \mathbf{0.2}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0.6}, \mathbf{0.2}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}), \ldots\}$$

$$e_3 = \{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{0}), (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$$
$$, (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{1}), \ldots\}$$

$$e_4 = \{(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.28}, \mathbf{0.85}, \mathbf{0.57}), (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.28}, \mathbf{0.85}, \mathbf{0.57}), (\mathbf{0.57}, \mathbf{0.85}, \mathbf{0.28}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}),$$
$$(\mathbf{0.14}, \mathbf{0.71}, \mathbf{0.71}, \mathbf{0.14}, \mathbf{0}, \mathbf{0}, \mathbf{0}), (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0.42}, \mathbf{0}), \ldots\}$$

   Where, for example, the corresponding fuzzy set for c_{11} of the e_1 is obtaining as:

$$\tau_{AC}(a_2) = (0, 0, 0, 0.25, 0.5, 0.5, 0)$$

   b) Transforming into 2-tuples. The fuzzy sets are transformed into linguistic 2-tuples by means of the function χ. The results for the dimension Entertainment are:

$$e_1 = \{(\mathbf{c_4}, \mathbf{0.2}), (\mathbf{c_3}, \mathbf{0}), (\mathbf{c_2}, \mathbf{-0.2}), (\mathbf{c_4}, \mathbf{0.2}), (\mathbf{c_4}, \mathbf{0.2}), \cdots\}$$

$$e_2 = \{(\mathbf{c_2}, \mathbf{-0.5}), (\mathbf{c_5}, \mathbf{-0.25}), (\mathbf{c_1}, \mathbf{0.25}), (\mathbf{c_1}, \mathbf{0.25}), (\mathbf{c_1}, \mathbf{0.25}), \cdots\}$$

$$e_3 = \{(\mathbf{c_4}, \mathbf{0}), (\mathbf{c_0}, \mathbf{0}), (\mathbf{c_5}, \mathbf{0}), (\mathbf{c_1}, \mathbf{0}), (\mathbf{c_6}, \mathbf{0}), \cdots\}$$

$$e_4 = \{(\mathbf{c_5}, \mathbf{0.16}), (\mathbf{c_5}, \mathbf{0.16}), (\mathbf{c_1}, \mathbf{-0.16}), (\mathbf{c_2}, \mathbf{-0.5}), (\mathbf{c_5}, \mathbf{0}), \cdots\}$$

For example, the transformation into a 2-tuple of c_{11}^1 is computed as:

$$\chi((0, 0, 0, 0.25, 0.5, 0.5, 0)) = \Delta(4.02) = (c_4, 0.2)$$

2. Aggregation phase: we shall combine the unified input assessments.
   a) Computing collective values for each criterion. We shall compute a collective value for each criterion according to all the users. We shall use the arithmetic mean for 2-tuples(Definition 4):

$$CVC_{1i} \Rightarrow \{(\mathbf{c_4}, \mathbf{-0.28}), (\mathbf{c_3}, \mathbf{0.22}), (\mathbf{c_2}, \mathbf{0.22}), (\mathbf{c_2}, \mathbf{-0.01}), (\mathbf{c_4}, \mathbf{0.11}), \cdots\}$$

Where, for example the first assessment is computed as:

$$CVC_{11} = AM^*((c_4, 0.2), (c_2, -0.5), (c_4, 0), (c_5, 0.16)) = (c_4, -0.28)$$

   b) Computing an evaluation assessment for each dimension. Now, we want to obtain an evaluation assessment for each dimension. We shall aggregate the collective values of its criteria by means of the arithmetic mean for 2-tuple (Definition 4):

$$ED_i \Rightarrow \{(\mathbf{c_3}, \mathbf{0.05}), (c_3, -0.27), (c_3, 0), (c_1, 0.01), (c_3, -0.41), (c_3, -0.02), (c_1, 0.01)\}$$

where, the first assessment is obtained according to this expression:

$$ED_1 = AM^*((c_4, -0.28), (c_3, 0.22), (c_2, 0.22), (c_2, -0.01), (c_4, 0.11)) = (c_3, 0.05)$$

Global Quality of the web site services.

Depending on the kind of web site services we are assessing, the importance of every dimension is different. In this case we have established that Entertainment, Convenience and Information reliability are more importance. The weighted vector we shall use is $= W\{0.2, 0.2, 0.2, 0, 0.2, 0.2, 0\}$ because, we are evaluating an educational web site, so the dimensions *security and assurance* and *product offer* are not crucial at all, hence the users don't provide

proper assessments about them. We shall apply the 2-tuple linguistic weighting average operator (Definition 5) to compute the EAW and finally we shall obtain:

$$EAW = W\_AM^{*}\left(\left(c_3, 0.05\right), \left(c_3, -0.27\right), \left(c_1, 0.01\right), \left(c_3, 0.12\right), \left(c_3, -0.41\right), \left(c_3, -0.02\right), \left(c_1, 0.01\right)\right) =$$

$$= \left(c_3, -0.06\right)$$

where $(c_3, -0.06)$ is the global linguistic evaluation for the quality of the lecturer's web site.

6 Concluding remarks

The evaluation of web site services have become a critical factor for users and companies in order to improve their commercial exchanges. So recently, different evaluation methods for this topic have arisen. The evaluation of these services is user centered because it depends on the opinions provided by the users according to their perceptions.

Current evaluation methods use numerical information to model users opinions. This modelling is not suitable for human perceptions. Therefore, we have proposed a hierarchical evaluation model that models the users opinions by means of linguistic information and these opinions could be assessed in different linguistic term sets to offer a greater flexibility to the users that take part in the evaluation process.

To manage the multi-granular linguistic information of the evaluation framework we have used fuzzy tools and the linguistic 2-tuples representation model.

Acknowledgments

This work has been partially supported by the Research Projects TIC 2003-03348 and TIC 2002-11942-E

References

1. M. Banârte, V. Issarny, F. Leleu, and B. Charpiot. Providing quality of service over the web: A newspaper-based approach. *Computer Networks and ISDN Systems*, (29):1457–1465, 1997.
2. P. Bharati and A. Chaudhury. An empirical investigation of decision-making satisfaction in web-based decision support systems. *Decision Support Systems*, (37):187–197, 2004.
3. P.P. Bonissone and K.S. Decker. *Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-Off Precision and Complexity.* In L.H. Kanal and J.F. Lemmer, Editors., Uncertainty in Artificial Intelligence. North-Holland, 1986.

4. G. Bordoga and G. Pasi. A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science*, (44):70–82, 1993.
5. P. Chang and Y. Chen. A fuzzy multicriteria decision making method for technology tranfer strategy selection in biotechnology. *Fuzzy Sets and Systems*, 63:131–139, 1994.
6. L.-D. Chen and J. Tan. Technology adaption in e-commerce: Key determinants of virtual stores acceptance. *European Management Journal*, 22(1):74–86, 2004.
7. R. Degani and G. Bortolan. The problem of linguistic approximation in clinical decision making. *International Journal of Approximate Reasoning*, 2:143–162, 1988.
8. M. Delgado, J.L. Verdegay, and M.A. Vila. Linguistic decision making models. *International Journal of Intelligent Systems*, 7:479–492, 1993.
9. M. Delgado, J.L. Verdegay, and M.A Vila. On aggregation operations of linguistic labels. *International Journal of Intelligent Systems*, 8:351–370, 1993.
10. E.Herrera-Viedma, O.Cordon, M.Luque, A.G.Lopez, and A.N. Muñoz. A model of fuzzy linguistic irs based on multi-granular linguistic information. *International Journal of Approximate Reasoning*, 34(3):221–239, 2003.
11. M. Gertz, M.T. Ozsu, G. Saake, and K.U. Sattler. Report on the dagstuhl seminar "data quality on the web". *Sigmod Record*, 33(1):127–132, 2004.
12. F. Herrera, E. Herrera-Viedma, and L. Martinez. A fusion approach for managing multigranularity linguistic term sets in decision making. *Fuzzy Sets and Systems*, (114):43–58, 2000.
13. F. Herrera, E. Herrera-Viedma, and J.L. Verdegay. A sequential selection process in group decision making with linguistic assessment. *Information Sciences*, 85:223–239, 1995.
14. F. Herrera and L. Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems*, 8(6):746–752, 2000.
15. F. Herrera and L. Martínez. An approach for combining linguistic and numerical information based on 2-tuple fuzzy representation model in decision-making. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8(5):539–562, 2000.
16. F. Herrera and L. Martínez. The 2-tuple linguistic computational model. Advantages of its linguistic description, accuracy and consistency. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(Suppl.):33–49, 2001.
17. F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.
18. F. Herrera, L. Martínez, and P.J. Sánchez. Managing non-homogeneus information in group decision making. *European Journal of Operational Research*, (To appear), 2004.
19. E. Herrera-Viedma. Modeling the retrieval process of an information retrieval system using an ordinal fuzzy linguistic approach. *Journal of the American Society for Information Science and Technology (JASIST)*, 52(6):460–475, 2001.
20. E. Herrera-Viedma and G. Pasi. Fuzzy approaches to access information on the web: Recent developments and research trends. In *Proceedings Interna-*

tional Conference on Fuzzy Logic and Technology (EUSFLAT03), pages 25–31, Zittau(Germany), 2003.
21. J. Van Iwaarden, T. Van der Wiele, L. Ball, and R. Millen. Perceptions about the quality of web sites: A survey amongst students at northeastern university and erausmus university. *Information and Management*, (41):947–959, 2004.
22. S. Kurnia and P. Schubert. An assessment of australian web sites in the grocery sector. pages 229–236. IADIS International Conference e-Commerce 2004.
23. C.K. Law. Using fuzzy numbers in educational grading system. *Fuzzy Sets and Systems*, (83):311–323, 1996.
24. H.M. Lee. Group decison making using fuzzy sets theory for evaluating the rate of aggregative risk in software development. *Fuzzy Sets and Systems*, 80:261–271, 1996.
25. E. Levrat, A. Voisin, S. Bombardier, and J. Bremont. Subjective evaluation of car seat comfort with fuzzy techniques. *International Journal of Intelligent Systems*, (12):891–913, 1997.
26. C. Liu and K.P. Arnett. Exploring the factors associated with web site success in the context of electronic commerce. *Information and Management*, (38):23–33, 2000.
27. P.J.A. Nagel and W.W. Cilliers. Customer satisfaction: A comprehensive approach. *International Journal of Physical Distribution and Logistic Management*, 20(6):2–46, 1990.
28. F. Naumann. Quality-driven query answering for integrated information systems. *LNCS*, 2261, 2002.
29. S. Negash, T. Ryan, and M. Igbaria. Quality and effectiveness in web-based customer support systems. *Information and Management*, (40):757–768, 2003.
30. L. Olsina and G. Rossi. Measuring web application quality with WebQEM. *IEEE Mul-timedia*, October-December:20–29, 2002.
31. A. Parasuraman, L.L. Berry, and V.A. Zeithaml. Servqual:a multiple-item scale for measuring customer perceptions of service quality. *In Journal of Retailing*, 1988.
32. M. Rifqi, V. Berger, and B. Bouchon-Meunier. Discrimination power of measures of comparison. *Fuzzy Sets and Systems*, (110):189–196, 2000.
33. P. Schubert. Extended web assessment method (EWAM) - evaluation of electronic commerce applications from the customer's viewpoint. *International Journal of Electronic Commerce*, 7(2):51–81, 2003.
34. M. Tong and P.P. Bonissone. A linguistic approach to decision making with fuzzy sets. *IEEE Transactions on Systems, Man and Cybernetics*, 10:716–723, 1980.
35. A. I. Torres and F. Vitorino. An e-satisfaction model - application to internet information search and purchase. pages 313–319. AIDIS International Conference e-Commerce, 2004.
36. R.R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, (12):237–261, 1995.
37. R.R. Yager, L.S. Goldstein, and E. Mendels. Fuzmar: an approach to aggregating market research data based on fuzzy reasoning. *Fuzzy Sets and Systems*, (68):1–11, 1994.
38. Z. Yang, S. Cai, Z. Zhou, and N. Zhou. Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Information and management*, (Article in press), 2004.

39. L.A. Zadeh. The concept of a linguistic variable and its applications to approximate reasoning. *Information Sciences, Part I, II, III*, 8,8,9:199–249,301–357,43–80, 1975.