Semantic Web technologies for generating feedback in online assessment environments

María del Mar Sánchez-Vera a, Jesualdo Tomás Fernández-Breis b, Dagoberto Castellanos-Nieves c, Francisco Frutos-Morales b, María Paz Prendes-Espinosa a

a Departamento de Didáctica y Organización Escolar, Universidad de Murcia, CP30100 Murcia, Spain
b Departamento de Informática y Sistemas, Universidad de Murcia, CP30100 Murcia, Spain
c Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

ABSTRACT

Feedback is an important component of assessment in learning environments, because it constitutes a new learning opportunity which is not usually available in most eLearning systems. Feedback allows students to know their learning flaws and helps teachers to design learning contents adapted to the needs of the students. Semantic Web technologies have been applied in recent years with different purposes in Education, but their applications for generating useful feedback have not been researched enough so far. In this paper we present an approach for generating feedback to open questions in assessment tests, by making use of ontologies and semantic annotations. The feedback is generated by obtaining the semantic similarity between the annotations associated with both questions and students’ answers. The method has been implemented as an extension of our online assessment platform. Its application in a real course is also presented and discussed here.

1. Introduction

Assessment, that is, the evaluation of the knowledge or skills of the students, is a basic activity in both conventional education and eLearning. Nowadays, most learning management systems offer different possibilities for carrying out assessment tests, and this contributes to carry out online assessment (OA) processes. The analysis of relevant literature [32,12,25], reveals the following problems of OA:

- It has not changed much despite the degree and variety of changes in the teaching–learning process due to the use of Information and Communication Technologies. It is usually performed through presentational or standardized tests, because most virtual courses require an accreditation process.
- eLearning allows for focusing on the learning process of the students, but online assessment is mainly used for the final assessment of the course.
- The assessment results should be more comprehensible and useful for the students. That would contribute to make it a new learning experience.
- Feedback is important in OA, but most systems does not take it into account effectively. In this way, assessment processes do not generate new learning opportunities for the students and new teaching opportunities for the teachers.

The concept of feedback has received several definitions. In a generic sense, feedback is the control of the system through the reinsertion of the results of the process [38]. If this information is capable of producing changes in the general method, this can be considered a learning process. If this process is then applied to students, feedback would be the return of information about the learning process according to particular predefined objectives. Feedback is obviously related to assessment, because it allows for returning to the student information that has been provided by the proper student, but improved with the processing done by the teacher. As a consequence, this new information helps the student to consolidate the acquired knowledge and, consequently, improve the teaching–learning process. One of the benefits of using feedback is that it makes it possible to warn the students about the quality of their work [28]. Consequently, the need for methods for generating and including feedback in learning management systems is clear.

In literature, three types of feedback are distinguished [29]: (1) student feedback, which is given to a student during learning; (2) author feedback, which is given to an author during course authoring; and (3) group feedback, which is given to a group of learners who study a course. Our work is focused on both student and author feedback, although the methods that will be presented...
can be adapted for groups. It should be pointed out that student feedback can be classified in three main categories [10]; (1) knowledge of results, which means that the students only receives a “correct” or “wrong”-like answer as feedback; (2) knowledge of correct response, which means that the students receive as feedback which were the correct answers; and (3) elaborated feedback, which means that the students receive as feedback relevant information to inspire learners to reason or judge correct responses or results. Elaborated feedback and knowledge of correct responses are generally considered better. In this work, we will pursue methods for the generation of knowledge of correct response feedback.

On the technical side, the Semantic Web (SW) aims at adding semantic information to web contents in order to create an environment in which software agents will be capable of doing tasks efficiently [4]. The appropriateness of SW technologies for developing eLearning systems is also supported by the research efforts done in the last years from different perspectives (see, for instance, [35,11,5]).

A number of technologies are needed for the success of the SW, among which the ontology is the cornerstone one. In the literature, multiple definitions for ontology can be found (see for instance[17,19]). An ontology can be seen as a semantic model containing concepts, their properties, interconceptual relations, and axioms related to the previous elements. On the other hand, semantic annotations are metadata associated to particular information items, expressed in terms defined by an ontology. Such annotations can be used for the semantic enrichment of information [3].

The present work describes an approach based on SW technologies to produce semantically rich feedback in online assessment processes. In this context, we identify three beneficial uses of Semantic Web technologies:

1. Ontologies can provide the precise semantic specification of the domain, in this case, the knowledge the students must acquire through the course.
2. Semantic annotations can be used for getting a precise semantic specification of the questions and the answers.
3. Automatic feedback processes can be developed by combining course ontologies and semantic annotations.

Knowledge is represented in the approach using ontologies. Such ontologies will include the knowledge the students have to acquire in a particular course and will provide the semantic context for the generation of feedback. The feedback method will be developed for assessment based on open questions, and each open question and each student answer will have a set of semantic annotations associated. Such annotations will be then compared to check whether the student has acquired the expected knowledge and the corresponding feedback will be generated. This approach will reuse results from our previous works, more concretely, the OeLE platform [6], which supports semantic assessment processes.

The structure of this paper is described next. First, some related work and the description of the OeLE platform are included in Section 2. Next, the approach for generating feedback is explained in Section 3. The extended OeLE platform is described in Section 4. The results of the application in a real course are shown in Section 5. Finally, the discussion and some conclusions are included in Section 6.

2. Background

2.1. Related work

The provision of feedback has been researched in the last decades. Marshall [27] proposed an approach for generating feedback for writing reports. The study presented in [16] concluded that the consistent use of immediate feedback may optimize the effects of individual instruction in modules, and that computer-directed feedback provides a powerful means to support the self-directed learning process. According to [37,40], student performance improves significantly after receiving knowledge of correct response feedback, which is the type of feedback addressed in our approach.

Concerning the effectiveness of feedback depending on the stage it is applied, applying feedback at intermediate steps during the resolution of a problem was found more effective than doing it only at the end [31,8]. This can be interpreted in learning settings as feedback should play an important role in continuous learning and assessment rather than being only used for final assessments. The feedback method developed in our work will be applicable during the teaching–learning process and not only at the end.

The relation between SW technologies for assessment and feedback processes has also to be explored. Some standards and recommendations make the relation and need for SW-based approaches clear. The Learning Object Metadata (LOM) standard recommends the annotation and classification of learning objects using metadata to facilitate their retrieval. The IMS LD specification, which describes the learning process that takes place in learning units, is currently described by an ontology [2]. Some ontologies have been built for eLearning domains: learning contents of technical documents [22], learning objects and group work [20], semantic extensions of SCORM [13], etc. On the other hand, ontologies and semantic annotations have been mainly used for the design, preparation, and classification of course materials (see for instance [36,11,24]).

Regarding assessment and feedback, some related approaches can be described. The review of the state of the art in this field presented in [30] concluded that there is still a long way to go before reaching the ideal system, although there are already some interesting systems. In [14], each group of students creates manually its own hyperbook from a course ontology and the different hyperbooks are compared and discussed collaboratively. Thus, the students are required to have some skills in ontologies. On the other hand, the Atenea platform [1] combines natural language processing and statistical techniques to process student’s natural language answers. In [41], the authors use course ontologies to specify the structure of the course contents and for generating the board structure for the Question/Answer process and classifying the e-documents read by the students. Such course ontologies are described using three main types of relations, that is, is-a, component-of, and part-of. In [18], a method for the intelligent assessment of summaries is provided. This approach is based on latent semantics analysis for analyzing the student’s answer and the reference one. Finally, [32] presents an approach based on DL reasoning, ontologies and Model-Driven Architectures is proposed for intelligent assessment.

In summary, these research results show the interest and effectiveness of feedback for improving the teaching–learning process, and that Semantic Web technologies may play a fundamental role for the achievement of this goal. However, none of the existing approaches is capable of supporting the generation of semantic feedback using a complete semantic infrastructure, that is, combining the formalization of the semantics of the domain, building semantic annotations of the questions and answers. In this work, we will then propose such novel solution, which will provide the academic community with new methods and tools for supporting feedback based on open questions. This approach will also be useful to show the value of SW technologies for building eLearning solutions. Our approach will be based on the use of SW standards and the reuse, adaptation and extension of approaches of interest developed by the SW community.
then the system runs automatically. Once the annotations have been obtained, OeLE gets automatically the marks for each question. The algorithm for matching questions and answers is similar to the one that will be presented in Section 3.3 for generating the feedback. Both methods share the similarity functions that are described in Section 3.2. The description of the assessment technique is out of the scope of this work and this platform has demonstrated its usefulness in real settings (see for instance [6]).

However, the OeLE platform does not allow students and teacher to know the main flaws of the students from the course knowledge perspective, that is, the generation of feedback for the students and teachers, only the numerical mark obtained by the students for each question is displayed, as well as the mark for the exam. Given our current objectives and the semantic properties of the OeLE platform, we decided to reuse the semantic representation of knowledge assessment entities and the annotation extraction methods used in OeLE.

3. Feedback generation

In this section, the approach for generating feedback is presented. First, the definitions of all the elements that are needed for generating such semantic feedback will be presented. Second, we will present the similarity function that will be used by the feedback algorithm. Third, the algorithm designed for generating the feedback will be presented.

3.1. Definitions

**Definition 1. Open question**

An open question, written open_question, is a question whose answer is given in free text. It can be defined by the tuple (desc, expected_answer, [open_question_annot], value), where:

- desc is the name of the question;
- expected_answer contains the correct answer to the question in natural language;
- open_question_annot, are the semantic annotations defined for such open question;
- value is the number of marks assigned to the question.

An example, using an XML notation, is:

```xml
<open_question>
  <desc>Identify and define what audio resource appeared from a term created by Ben Hammersley. Indicate its advantages</desc>
  <expected_answer>The audio resource is Podcast. The origin of the word Podcast comes from the union of the words Ipod and broadcasting. An Ipod is a portable player developed by Apple...</expected_answer>
  <value>5</value>
</open_question>
```

**Definition 2. Open question annotation**

An annotation of an open question, written open_question_annot, defines the number of marks for a particular association.
between a question and a knowledge entity. It can be defined by the pair \( \text{entity}_\text{annot}, \text{quantitative}_\text{value} \) where

- \( \text{entity}_\text{annot} \) represents the annotation for the knowledge entity in the course ontology. There are three types of \( \text{entity}_\text{annot} \): concepts, relations and attributes. Thus, an \( \text{entity}_\text{annot} \) stands for a concept, relation or attribute of the course ontology;
- \( \text{quantitative}_\text{value} \) is the numerical score associated to the annotation, hence standing for the importance of the annotated entity.

An example is “Podcast allows for listening to the radio, whose quantitative value is 4, and its representation is:

\[
\begin{align*}
<\text{open_question_answer}> \\
<\text{entity_annotation}> \\
<\text{relation}> \\
<\text{concept1}> \text{course}\#\text{Podcast}</\text{concept1}> \\
<\text{name}> \text{course}\#\text{allows}\_\text{for}\_\text{listening}</\text{name}> \\
<\text{concept2}> \text{course}\#\text{Radio}</\text{concept2}> \\
</text_relation> \\
<\text{entity_annotation}> \\
<\text{quantitative_value}>4</\text{quantitative_value}> \\
</open_question_answer>
\]

\textbf{Definition 3. Answer annotation}

A semantic annotation of an answer, written \( \text{answer}_\text{annotation} \), is comprised of the knowledge entity and the linguistic expression from the text of the answer associated to that knowledge entity. Therefore, it can be defined by the pair \( \{ \text{entity}_\text{annot}, \text{ling}\_\text{exp} \} \), where

- \( \text{entity}_\text{annot} \) is defined as \( \text{open_question}_\text{annotation} \);
- \( \text{ling}\_\text{exp} \) represents the text of the answer associated to the knowledge entity.

For example, the answer annotation for “Press is periodically published” is:

\[
\begin{align*}
<\text{answer_annotation}> \\
<\text{entity_annotation}> \\
<\text{relation}> \\
<\text{concept1}> \text{course}\#\text{Press}</\text{concept1}> \\
<\text{name}> \text{course}\#\text{is}\_\text{a}</\text{name}> \\
<\text{concept2}> \text{course}\#\text{Periodical}\_\text{Publications}</\text{concept2}> \\
</text_relation> \\
<\text{entity_annotation}> \\
<\text{ling}\_\text{exp}> \text{Press}\_\text{is}\_\text{periodically}\_\text{published}</\text{ling}\_\text{exp}> \\
</\text{answer_annotation>}
\]

\textbf{Definition 4. Open question answer}

The answer to an open question, written \( \text{open_question_answer} \), is comprised of the answered text plus its semantic annotations associated. Consequently, it can be defined by the pair \( \{ \text{text}\_\text{answer}, \{ \text{answer}\_\text{annotation}\} \} \), where

- \( \text{text}\_\text{answer} \) is the answer of the student in natural language;
- \( \text{answer}\_\text{annotation} \), are the semantic annotations obtained from the textual answer, which are defined next.

An example is:

\[
\begin{align*}
<\text{open_question_answer}> \\
<\text{text_answer}> \ldots</\text{text_answer}> \\
<\text{annotations}> \\
<\text{answer_annotation}> \ldots </\text{answer_annotation}> \\
\ldots \\
<\text{answer_annotation}> \ldots </\text{answer_annotation}> \\
</\text{annotations}> \\
<\text{quantitative_value}>4</\text{quantitative_value}> \\
</open_question_answer>
\]

\textbf{Definition 5. Feedback annotation}

A feedback annotation, written \( \text{feedback}_\text{annotation} \), permits to capture the quality of the answers given by the students. Each semantic annotation of an answer has a feedback annotation associated. In case that answer annotation is considered correct by the marking algorithm, then the value for \( \text{feedback}_\text{annotation} \) will be \( \text{correct} \). Otherwise, its value will be \( \text{wrong} \).

\textbf{Definition 6. Positive feedback annotation for an answer annotation}

A positive feedback annotation for an answer, written \( \text{positive_feedback}_\text{annotation} \), is comprised of an annotation of the answer and its corresponding feedback annotation. Therefore, it can be defined by the triple \( \langle \text{answer}_\text{annotation}, \text{open_question}_\text{annotation}, \text{feedback}_\text{annotation} \rangle \), where the value of \( \text{feedback}_\text{annotation} \) is \( \text{correct} \).

The positive feedback annotation for the example shown in \textbf{Definition 3} would be:

\[
\begin{align*}
<\text{positive_feedback_annotation}> \\
<\text{answer_annotation}> \\
<\text{entity_annotation}> \\
<\text{relation}> \\
<\text{concept1}> \text{course}\#\text{Press}</\text{concept1}> \\
<\text{name}> \text{course}\#\text{is}\_\text{a}</\text{name}> \\
<\text{concept2}> \text{course}\#\text{Periodical}\_\text{Publications}</\text{concept2}> \\
</text_relation> \\
<\text{entity_annotation}> \\
<\text{ling}\_\text{exp}> \text{Press}\_\text{is}\_\text{periodically}\_\text{published}</\text{ling}\_\text{exp}> \\
</\text{answer_annotation}> \\
<\text{open_question_annotation}> \\
<\text{entity_annotation}> \\
<\text{relation}> \\
<\text{concept1}> \text{course}\#\text{Press}</\text{concept1}> \\
<\text{name}> \text{course}\#\text{is}\_\text{a}</\text{name}> \\
<\text{concept2}> \text{course}\#\text{Periodical}\_\text{Publications}</\text{concept2}> \\
</text_relation> \\
<\text{entity_annotation}> \\
<\text{quantitative_value}>4</\text{quantitative_value}> \\
</open_question_annotation> \\
<\text{feedback_annotation}> \\
<\text{correct}> </\text{correct}> \\
</\text{feedback_annotation}> \\
</positive_feedback_annotation>
\]
Definition 7. Negative feedback annotation for an answer annotation

A negative feedback annotation for an answer, written negative_feedback_annotation, is comprised of an annotation of the answer and its corresponding feedback annotation. Therefore, it can be defined by the triple (answer_annotation, open_question_annotation, feedback_annotation), where the value of feedback_annotation is wrong.

Definition 8. Feedback for an open question answer

The feedback for a particular answer A, written student_feedback (A), is the answer provided by the student plus the feedback generated for the answer annotations associated with A. Consequently, it can be defined by the triple (text_answer, feedback (A)), where

- text_answer is the answer given by the student to the question.
- feedback (A) is positive_feedback (A) U negative_feedback (A), where
  - positive_feedback (A) is the set of positive_feedback_annotations for the answer_annotations associated with A.
  - negative_feedback (B) is the set of negative_feedback_annotations for the answer_annotations associated with A.

Definition 9. Teacher feedback for an open question

The feedback generated for a teacher for a particular question Q, written teacher_feedback (Q), is the union of the feedback generated for the students for that question. Therefore, it can be defined as U(student_feedback(A)_i), where A_i is an answer given by a student to the question Q.

An example of teacher feedback would be: (student_feedback_{a_1, …, student_feedback_{a_n}}, where q is the question and n is the number of students that answered that question.

Finally, it should be noted that Definitions 1–4 are reused from OeLE, whereas Definitions 5–9 are needed for the generation of the semantic feedback.

3.2. The semantic similarity functions

The feedback algorithm requires to know how similar two annotations are. For this purpose, the following functions for measuring similarity between concepts, attributes and relations are used. Each individual similarity function returns a value in [0, 1].

3.2.1. Concepts

The similarity between concepts is evaluated through the function (1). It is calculated as the weighted average of three factors: the proximity of the concepts in the taxonomic structure of the ontology (concProx), the linguistic similarity of the terms associated with the concepts (eqName) and the similarity of the set of properties associated with the concepts (propSim). The weights cP1, cP2, and cP3 define the importance of each factor. In this way, cP1 is the importance of the ontological distance between the concepts, cP2 is the importance of the similarity of the sets of properties, and cP3 refers to the importance of the linguistic similarity.

\[
concSim(c_i, c_j) = cP1 \times concProx(c_i, c_j) + cP2 \times propSim(c_i, c_j) + cP3 \times eqName(term(c_i), term(c_j))
\]

where \(\sum cP_i = 1\) and \(0 < cP_i < 1\).

The conceptual proximity is the distance of the concepts in the taxonomic hierarchy of the ontology. The function is defined by

\[
concProx(c_i, c_j) = 1 - \frac{\text{dist}(c_i, c_j)}{\text{Nodes}}
\]

where dist stands for the amount of concepts between \(c_i\) and \(c_j\) through the closest common taxonomic parent; nodes stands for the total amount of concepts in the ontology; and \(\text{anc}(c_i, c_j)\) returns the common taxonomic parents of the concept \(c_i\). In case there is no common parent, the conceptual proximity function returns 0.

The linguistic distance (see Eq. (3)) is obtained by getting the linguistic similarity between the terms associated with the concepts. This approach uses the Levenshtein distance [26], written \(L(x, y)\).

\[
eq Name(s, s_j) = \frac{1}{1 + L(s, s_j)}
\]

Finally, propSim accounts for the similarity between the sets of properties associated with the respective concept. It is calculated by the following functions [33]:

\[
propSim(c_i, c_j) = \frac{|C(c_i, c_j)|}{G(c_i, c_j)}
\]

\[
C(c_i, c_j) = \text{commonAttributes}(c_i, c_j) \cup \text{commonRelations}(c_i, c_j)
\]

\[
G(c_i, c_j) = |C(c_i, c_j)| + \beta_1 \times |\text{nC}(c_i, c_j)| + \beta_2 \times |\text{nC}(c_i, c_j)|
\]

The factor \(C(c_i, c_j)\), refers to the amount of properties both concepts share, and it is calculated as follows: (1) \(\text{commonAttributes}\) is the set of attributes having the same name, type and value restrictions in \(c_i\) and \(c_j\); (2) \(\text{commonRelations}\) is the set of relations having the same name, with \(c_i\) and \(c_j\) playing the same role in the relation and the other participant being the same in both relations.

On the other hand, \(\text{nC}(c_i, c_j)\) is calculated analogously but considering the set of attributes and relations which do not appear in both concepts. Finally, \(\beta_i\) is calculated as defined in Eq. (7). In this function, the value of \(\beta_i\) depends on the depth of the concepts \(c_i\) and \(c_j\) in the taxonomy, according to the definition provided in [34].

\[
\beta_1 = \frac{\text{depth}(c_i)}{\text{depth}(c_i) + \text{depth}(c_j)}
\]

\[
\beta_2 = 1 - \beta_1
\]

3.2.2. Attributes

The similarity between two attributes, written attSim, is calculated by using three factors: (1) the linguistic similarity; (2) the similarity of their value sets; and (3) the similarity of the concepts they refer to. These elements are combined in

\[
attSim(a_i, a_j) = at1 \times eqName(term(a_i), term(a_j)) + at2 \times valSim(a_i, a_j) + at3 \times concSim(concept(a_i), concept(a_j))
\]

where \(\sum at_i = 1\) and \(0 \leq at_i \leq 1\).

The weights at1, at2, and at3 are similar to the cp_i for the concepts: at1 weights the importance of the linguistic distance, at2 indicates the importance of the similarity of the values of the attributes, and at3 refers to the similarity of the concepts associated with the attributes. The first and the third factors have already been described for concepts. The second factor, written valSim, calculates the similarity among value sets as shown in

\[
valSim(a_i, a_j) = \frac{|values(a_i) \cap values(a_j)|}{\min_k[|values(a_k)|]}
\]
3.2.3. Relations

The similarity between two relations depends on the relationships themselves and on the similarity of the concepts that participate in such relationships. Hence, the similarity between two relations, written \( \text{relSim}_i, r_j \), is calculated by using two factors: (1) the linguistic similarity between the relations; and (2) the similarity between their participants. Both factors are combined in

\[
\text{relSim}(r_i, r_j) = r_{l1} \cdot \text{eqName}(\text{term}(r_i), \text{term}(r_j)) + r_{l2} \\
\quad \cdot \text{concSim}(r_i, \text{concept}_1, r_j, \text{concept}_1) \\
\quad \cdot \text{concSim}(r_i, \text{concept}_2, r_j, \text{concept}_2)
\]

(10)

where \( \sum r_{li} = 1 \) and \( 0 \leq r_{li} \leq 1 \) and \( r_i, \text{concept}_j \) stands for the \( j \)th concept associated to the \( i \)th relation.

3.3. The feedback algorithm

This algorithm has three main input parameters: the open question, a particular answer for that question, and the set of assessment parameters. We describe now how this algorithm works. First, the semantic similarity between the annotations associated with the question and the answer are calculated by applying the functions described in the previous section. This generates a similarity matrix, whose rows correspond to the answer, and whose columns correspond to the question. The algorithm matches answers annotations with questions ones, with the following constraints: (1) a question annotation can only be associated with one answer annotation; (2) an answer annotation can only be associated with one question annotation; and (3) the associated annotations must be of the same knowledge category, that is, concept, relations, attributes. An eligibility matrix supports the process. The cells in the similarity matrix can be in two states: eligible and uneligible. A cell is eligible if its associated annotations satisfy the three constraints. Otherwise, it is ineligible.

The algorithm executes now iterations on the following actions:

1. Get the next eligible cell with the highest similarity score. If none, go to 5.
2. Update the eligibility of the cells of the matrix.
3. If the similarity score is lower than the threshold, then generate a negative feedback annotation for the answer annotation. Otherwise, generate a positive feedback annotation for the answer annotation.
4. Add this new feedback annotation to the feedback annotations of the answer, and go to 1.
5. If no feedback annotation has been produced for an answer annotation, then generate a negative feedback annotation for it.

The algorithm consists then in initialization and matching. The initialization requires to calculate \( n \times m \) similarity functions, where \( n \) is the number of annotations of the questions and \( m \) is the number of annotations of the answer. To improve the efficiency of the algorithm, this process generates the matrix of eligible cells, which is implemented as a sorted list for efficiency purpose. Each element of this list includes information about the question and answer annotations and the similarity score. Once this list has been initial-
ized, it is sorted by decreasing similarity score. The current matching algorithm is greedy and provides the matching in n decisions. Each decision takes into account all the eligible cells, but from a practical perspective, the next decision is located at the first position of the eligibility list. Updating the eligibility matrix has linear complexity, since it requires to remove from the list all the elements whose annotations refer to the corresponding question annotation. Thus, in terms of complexity, the algorithm provides good results.

Next, an example is shown. Let us suppose that we have the following elements: annotations of the question (see Table 1), annotations of the answer (see Table 2), and the threshold value is 0.8, meaning that a student gets marks for an answer annotation if the similarity with its associated question annotation is not lower than 0.8.

The similarity matrix is displayed in Table 3, whereas the initial eligibility of cells is shown in Table 4. There, x means that the cell is not eligible. The final result is shown in Table 5. The cells that became ineligible during this process are marked as Sy, where y is the step in which they become ineligible.

Step 1: The next eligible cell would be either (A1, Q1) or (A3, Q2). In this case, the algorithm would take the first one. Since 1 is greater than 0.8, we add (A1, correct) to the feedback.

Step 2: The next eligible cell is (A3, Q2). Since 1 is greater than 0.8, we add (A3, correct) to the feedback.

Step 3: The next eligible cell is (A4, Q6). Since 0.18 is lower than 0.8, we add (A4, wrong) to the feedback.

Step 4: The next eligible cell is (A5, Q3). Since 0.15 is lower than 0.8, we add (A5, wrong) to the feedback.

Step 5: There is not any eligible cell, so exit.

Given that no feedback annotation has been provided for A2, then (A2, wrong) is added to the feedback. The resulting feedback annotation structure is {(A1, correct), (A2, wrong), (A3, correct), (A4, wrong), (A5, wrong)}.

Thus, the generation of the feedback for students is basically the application of the previous algorithm to all the questions of their exam. On the other hand, the feedback for the teacher of a subject is created by analyzing all the answers of the students to the same exam. In this way, the previous algorithm is applied to all the an-

![Fig. 3. Providing feedback to teachers in OeLE.](image-url)
answers of the students for each question, and the feedback for each question is added to the feedback for the exam.

4. The extended OeLE platform

Feedback has been incorporated in the OeLE platform for both teachers and students. The new OeLE platform is depicted in Fig. 2. The main difference with Fig. 1 is the inclusion of the feedback subsystem, capable of generating useful information for teachers and students. The OeLE platform offers a desktop application for teachers and a web-based access for students, so the corresponding software artifacts had to be modified appropriately. This section has then two streams, one per type of agent involved in the teaching–learning process: teacher and student.

4.1. Feedback for teachers

The feedback received by a teacher for a particular exam is shown in Fig. 3. The upper part of the dialog contains general information about the exam, showing some statistics such as mean, standard deviation, highest and lowest scores, and the description of the marking criterion used (“Standard”). The lower part of the dialog provides the teacher feedback. This analysis calculates how many students have answered correctly each ontological entity associated to the questions, and how many have done it wrong.

To this end, such entities are classified into two sets: (a) best acquired aspects; and (b) worst acquired aspects. Both sets are shown in the lower part of the dialog. In Fig. 3, the concept “Object of Study” has been correctly answered by all the students, and the concept “Sensation of reality” has not been correctly answered by all the students. If the teacher considers this concept important, she would know that it has not been acquired by her students and could take corrective actions.

Graphical feedback is generated for the teacher (see Fig. 4) when the teacher selects a set of entities and the desired type of graph. The OeLE platform generates bar and pie graphs for teachers. Bar graphs (see the upper part of the figure) allow for viewing the selected course knowledge items ordered by decreasing percentage. On the other hand, pie ones (see the lower part of the figure) allow for representing and analyzing course knowledge items in relative terms. In this case, we can see that “Navigation” has been, for instance, acquired better than “Didactic Resource” since its portion of the graph is larger.

4.2. Feedback for students

The semantic feedback is generated for each open question of the exam. The feedback shown for each question is: the description of the question, the numerical score obtained by the student in that question, the aspects that have not been acquired and the explanation of the marking of the question in terms of the semantic annotations of the student’s answer. An example of such feedback is shown in Figs. 5 and 6. The score, the expected answer, and aspects to improve are shown in Fig. 5. The list of aspects to improve contains the knowledge items that were expected to be answered in this question, but the student did not answer correctly. In this example, the student did not answer correctly concepts such as “freedom of expression”. On the other side, Fig. 6 analyses the annotations of the answers of the student. Correct answers are marked in green and wrong ones in red. In this example, concepts such as “Simplicity” or “Knowledge transfer” were right answers, whereas the concept “Subscription” was not in the expected answer. The knowledge entities are underlined since they contain links to their definition in the course ontology.

5. Application in a real course

5.1. General description and research hypotheses

The course “Design and Production of Educational Materials” is an elective module of the Bsc on Pedagogy in the University of Murcia. This course took place in the second semester of 2008/
2009, and had 25 students. This subject has a practical orientation, whose main goal is training the students in the aspects of design, production and evaluation of content in didactic processes. The academic performance of the students is evaluated with an e-portfolio and other different activities throughout the 9 themes of the program and also the participation of the students in several communication situations (videoconferences, forum and collaborative works). The students had also to take a written exam in a concrete assessment environment. OeLE was the assessment environment used in this course, and our research is focused on this part of the course.

The experiment was executed as follows. The students took an exam using the OeLE platform. Those exams were assessed by OeLE, which generated the feedback information for each student and for the teacher, and were also marked by one of the teachers of the course. At that stage, the students were given access for one week to reinforce contents. Therefore, the students had the chance to revise the contents associated with the course knowledge items that they did not have answered correctly in the exam. Then, the students took a second exam, which was marked as the first one. Finally, the students were asked to answer a questionnaire about the experience, and the experience was analyzed. More details about the process will be given in Section 5.2. The execution of the process can be accessed at our website.²

The students of the course are the population of this experiment. The selection of the sample was not probabilistic, because intentional sampling was used. The selection criterion was to be registered in the course. The size of the sample was 22, 95.2% of the sample are females and 4.8% are males, which is usual in the degree of Pedagogy. Next, our main research hypotheses are listed, although additional aspects were also investigated and will be reported later in this paper:

- **Hypothesis 1:** The automatic marks obtained by OeLE are similar to the manual ones obtained by a human teacher.
- **Hypothesis 2:** The students are satisfied with the technical details of the feedback learning contents.

² http://miuras.inf.um.es/~oele/feedback/.
Hypothesis 3: The opinion of the students about feedback in assessment processes and their opinion about the usefulness of the feedback generated by the OeLE platform in this experience are significantly related.

Hypothesis 4: The marks obtained by the students in the first exam are not significantly different from the ones obtained in the second one.

5.2. The methodology applied in the study

In this section, the activities of this experience are described sorted by their chronological order of execution:

1. Development of the course ontology.
2. Design of the exams.
3. Preparation of reinforcement contents.
4. Design of the questionnaire.
5. Assessment and feedback.

5.2.1. The course ontology

The teachers of the course were responsible for building the ontology, and they were supervised by experts in ontology engineering, who had previously trained the teachers in ontology engineering and the use of tools for building ontologies. This was the first experience of the teachers with the development of ontologies, so they found this process hard for an average user. The OWL ontology was implemented using Protégé, which is the most popular tool for creating ontologies. The ontology (see an excerpt of the ontology as viewed in the OeLE platform in Fig. 7) has 111 classes, 71 object properties, 51 data type properties, including also disjoint and cardinality constraints. Its consistency was checked using Pellet, and the ontology has ALC(H)N (D) DL expressivity. This ontology, which can be found at our website, was then imported into OeLE and used in this experience.

5.2.2. The assessment tests

The teachers designed in OeLE the two exams (see 6,7) that were applied in this experience. First, the teachers created a series of possible questions. Second, the teachers selected five open questions for each exam. In order to facilitate the automatic assessment of the answers of the students, one of the teachers was responsible for supervising the semantic annotation of the expected answers using OeLE.

5.2.3. The reinforcement contents

The selection of the questions of the exam, and the corresponding semantic annotations allowed the teachers to design specific reinforcement contents for the exams. Given that the current version of the OeLE platform does not support content management, an external website was developed. The website contains a series of HTML learning objects that were designed and associated to the concepts of the course ontology. This website was developed using ExeLearning, and it was structured according to the questions of the exam.

---

3 http://protege.stanford.edu/.
5 http://miuras.inf.um.es/~oele/diseno.owl.
6 http://miuras.inf.um.es/~oele/feedback/exam.html.
7 http://miuras.inf.um.es/~oele/feedback/exam2.html.
8 http://miuras.inf.um.es/oele/objetos/.
9 http://exelearning.org/.
the first exam, for which the feedback was generated. An example of reinforcement content is shown in Fig. 8. This website was designed having in mind that the students would receive a semantic feedback as shown in Section 4.2, so the most semantic annotations of each question were also shown in the menus for speeding up the access to the reinforcement content.

5.2.4. The questionnaire

The questionnaire was designed by the teachers. It included questions referred to both technical and educational issues. The questionnaire contained 28 questions. Three questions had yes/no answers, six were multiple choice, two were open questions, and seventeen were Likert. It should be noted that Likert questions have usually 5 or 7 possible answers. However, in this questionnaire, we removed the intermediate category and used four possible values: very disagree, disagree, agree, very agree. The questionnaire was created using Formlogix. The reliability of the questionnaire was checked using Cronbach’s alpha (see Cronbach [9]), obtaining a value 0.75, which means the questionnaire is highly reliable.

5.2.5. Assessment and feedback

Both exams were executed similarly. The students were given one hour for completing each exam in OeLE, and they were allowed to use any online or offline material for answering the questions. 22 students took the first exam and 21 the second one. One teacher supervised the annotation extraction of the answers of the students for each exam and then, the exams were marked by OeLE. Simultaneously, other teacher manually marked the exams. It should be noted that the teacher that performed the manual assessment was different from the one who supervised the annotation of the answers of the students. The marks of both exams are available at our website. The feedback was automatically generated for the first exam by the OeLE platform and shown to the students and teachers.

5.2.6. Evaluation of the experience

After the execution of the previous phases, the students were given two weeks for answering the questionnaire. Some answers (in Spanish) can be found at our website. The interpretation of the answers and the results will be provided in Section 5.3.

5.3. Results

5.3.1. Data analysis process

The research hypotheses have been analyzed using univariate descriptive methods and, for those requiring checking the relation between different variables, a bivariate descriptive analysis was performed. Due to the sample size, the analysis of the questionnaire was done using non-parametric statistical techniques. For hypotheses with qualitative variables, the values were captured and they were analyzed by frequency and percentage. For those with quantitative variables, basic statistics, such as mode or dispersion, were used.

The data matrix of the marks of the exams has three scale-type variables, which are the three marking methods: (1) manual marking of the first exam; (2) automatic marking of the first exam; and (3) automatic marking of the second exam. Such data were analyzed using ANOVA with repeated measures, in which the average mark for each exam was also included in the analysis. This parametric technique can be used because, even though there are 22 subjects, the number of observations is 66. However, we also executed Friedman and Wilconox (see [15,39]) tests to check whether the results are similar to the ones obtained by ANOVA with repeated measures.

The process can be described as follows. First, a direct reading of data was performed through frequency calculus, which allowed for knowing the distribution of data. Second, a cross-reading of the data was executed, so contingency tables and their corresponding association test were obtained. For the qualitative variables, contingency tables and Pearson’s chi-square tests (see Chernoff and Lehmann [7]) were applied for checking the existence of significant association between the variables. The correlations were calculated using Pearson product-moment correlation coefficient as describe in Katz [23].

5.3.2. Hypothesis 1: The automatic marks obtained by OeLE are similar to the manual ones obtained by a human teacher

Our previous research [6] concluded that the similarity between the marks provided by the tool and the human marking could not be rejected, but we wanted to perform this study in this experience. The initial analysis of the average marks of two methods for the first exam reveals that the average mark is 6.030 for the manual process and 6.569 for the automatic one. We applied an analysis of repeated measures using the marking method as factor. The analysis of the combination 1–2 allows for testing the existence of a significant difference between the results of both types of marking. The result of this analysis is \( d = -0.099; \ p = 0.135 \) as
shown in Table 6. The p-value means that we can reject the hypothesis of the existence of such differences, so the hypothesis 1 cannot be rejected.

5.3.3. Hypothesis 2: The students are satisfied with the technical details of the feedback learning contents website

The frequency analysis reveals a positive opinion of the students. 100% accessed the website easily, 90.5% found easily the information they searched, 68.2% found it friendly and usable, and 95.2% found positive that OeLE indicated them the errors they had made in the exam. Most students expressed a positive opinion for all the items related to this website.

The average and typical deviation allowed us to compare the results for the website and OeLE. We recorded the variables referred to the technical aspects of both tools in two variables, one per tool, and this has allowed us to find a positive, significant relation between them, as it can be seen in Table 7, where webLO refers to the learning object website. There, the correlation table reinforces this result. Consequently, the hypothesis 2 cannot be rejected.

5.3.4. Hypothesis 3: The opinion of the students about feedback in assessment processes and their opinion about the usefulness of the feedback generated by the OeLE platform in this experience are significantly related

The results obtained through frequency analysis indicate that 86.4% of the students found that the feedback process was not a waste of time. 77.3% said that the feedback allowed them to understand better some course contents. 86.4% checked the learning contents between 1 and 10 times, and 36.4% did not only checked the items suggested by OeLE, but also those in which they had special interest. All the students had checked at least one reinforcement content. 71.4% said that the reinforcement process helped them to get a better result in the second exam, and 76.2% disagreed with “the reinforcement contents are not necessary because the contents of the course are available online”.

A negative, significant correlation between “I could have obtained the same results in the second exam without the feedback” and “the feedback contents have helped me to do better the second exam” has been found (see Fig. 9). It can then be said that there is a significant relation between the general opinion of the students about the feedback as an element of the assessment process and the concrete usefulness of the feedback in this experience. The relations established are coherent, because the students that could have obtained the same result without feedback also say that the feedback has helped them to do a better second exam. There is a relation between the positiveness of receiving the feedback and the results obtained in the second exam. Overall, the hypothesis 3 cannot be rejected.

5.3.5. Hypothesis 4: The marks obtained by the students in the first exam are not significantly different from the ones obtained in the second one

The marks compared in this case were the ones provided by OeLE. The average mark for the first exam is 6.129, where the average for the second is 6.569. An analysis of repeated measures was done, using the marking method as factor. The results shown in Table 8 (\(d = -0.440; p = 0.069\)) indicate that there is a marginally significant difference between the results of both exams. Given that the p-value is close to 0.05, we calculated the determination coefficient. In this case, \(R^2_a\) is 0.53, which could be due to the sample size. Therefore, it can be said that there are significant differences between the results of both exams, so the hypothesis 4 can be rejected.

---

### Table 6

<table>
<thead>
<tr>
<th>(I) factor 1 (J) factor 1</th>
<th>I – J</th>
<th>Std error</th>
<th>Sig(^a)</th>
<th>Conf. int. 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-0.099</td>
<td>0.064</td>
<td>0.133 ( -0.232, 0.033)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-0.539</td>
<td>0.247</td>
<td>0.041 ( -1.054, 0.024)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-0.099</td>
<td>0.064</td>
<td>0.135 ( -0.232, 0.033)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-0.440</td>
<td>0.229</td>
<td>0.069 ( -0.917, 0.037)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.440</td>
<td>0.228</td>
<td>0.069 ( 0.033, 0.917)</td>
</tr>
</tbody>
</table>

Based on estimated marginal measures.

\(^a\) Adjustment for multiple comparisons. Least significant difference (equivalent to the absence of adjustment).

### Table 7

<table>
<thead>
<tr>
<th>Tool/tool</th>
<th>webLO</th>
<th>OeLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation of Pearson</td>
<td>1</td>
<td>0.589</td>
</tr>
<tr>
<td>Sig (bilateral)</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tool/tool</th>
<th>webLO</th>
<th>OeLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation of Pearson</td>
<td>0.589</td>
<td></td>
</tr>
<tr>
<td>Sig (bilateral)</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Model summary</th>
<th>R square</th>
<th>Adjusted R</th>
<th>Standard error of estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.317a</td>
<td>0.100</td>
<td>0.053</td>
</tr>
</tbody>
</table>

---

Fig. 9. The contingency table for H3.
It must be noted that we cannot justify that the students have learned by having obtained a higher mark in the second exam. This result was expected since both exams covered the same knowledge area of the subject and there were not many days between both exams. This is why the usefulness of the feedback was also studied analyzing the questionnaire.

5.3.6. Other results
In this section we present other results, which are not directly related to feedback, obtained in this experience by analyzing the questionnaire filled by the students.

- The students are satisfied with this assessment experience. 72.8% of the students enjoyed doing open questions-based online exams. 81.8% of the students think that this modality of assessment should be used more frequently. The average degree of satisfaction is 3.41 (3 means some satisfaction), with a typical deviation of 0.796. For all the items related to the students’ experience with this assessment and feedback experience, the average result was above 3. It should be taken into account that the rating was 1 to 4.

- The students are satisfied with the technical performance of OeLE. Our frequency analysis reveals that 72.7% of the student did not feel uncomfortable using OeLE. 77.3% of them found friendly how the information is displayed, and 72.7% said that they could execute the exams without usability and technical problems. 95.5% of the students said that the access of the assessment exams in the platform was simple. Consequently, it might be said that the students are satisfied with the technical performance of OeLE.

6. Discussion and conclusions
OeLE is an innovative online assessment system capable of marking exams and identifying the main learning flaws and strengths of the students. This allows teachers knowing the individual and collective performance and learning of their students. Our results have evaluated feedback positively, which is in line with the conclusions of other studies [16,37,40]. With the current extension for providing automatic feedback to both students and teachers, OeLE goes beyond other currently available intelligent assessment tools, like the ones described in Section 2.1. In this paper, we have described how the basic Semantic Web Technologies can be used for generating useful feedback in online assessment environments.

Our approach addresses both student and author feedback. Concerning student feedback, most students considered it positive, and we found a positive relation between the idea that feedback is useful in online assessment environments and the usefulness the students perceived of the feedback received in the experience. Thus, from a pedagogical point of view, the information offered by OeLE is a fundamental, positive aspect that can be used in educational processes. The students also liked the reinforcement contents prepared for the experience and found them very useful. The whole experience provided the students with concrete suggestions for revision, as suggested in [31]. Unfortunately, the current version of the OeLE platform is not capable of managing learning objects, so the contents had to be manually found by the students. To overcome such limitations, the SICARA platform [13] which allows for the semantic annotation of SCORM learning objects, will be integrated into OeLE.

On the author side, the feedback generated by the platform allowed the teachers to know which aspects had been acquired best and worst by the students. This was found interesting by the teachers because each knowledge item had a relative importance associated to each annotation in the questions. In this experience, the best acquired concepts were those with higher relative value.

Unfortunately, the current version of the OeLE platform does not exploit this relative importance in its feedback analysis, so it will be part of our future work.

The teachers needed to make an initial effort to create the ontology, and supervise the annotation of the questions, the last one being a semi-automatic process in OeLE. Given that this was the first time OeLE was used in that subject, the knowledge base was empty and the teachers were asked to supervise the annotations suggested by the tool. In this experience, the time spent in annotating the exams was higher than the time spent in the manual correction for the first exam, and almost the same for the second exam. The possibility of reusing the ontologies, questions, annotations, etc. for different exams, different years and the possibilities offered by the feedback generation makes it worthy, because the teachers would have needed more time to generate manually the personalized feedback. Consequently, in a global perspective, the teachers found that OeLE saved them time and found it useful. Besides, the development of an ontology for this course and a base of annotated questions, which can be reused in further editions of the course, were also valuable results of this work for the teachers. The use of feedback should be carefully planned by the teachers. In this sense the recommendations of [8] should be taken into account, since the method might be applied at different stages of the course. Analysing the actions taken by the teachers to improve the course contents after receiving the feedback was out of the scope of the present work. Additional research should also be done in this area, following the indications of [21] for using the feedback as a way for increasing the quality of the course.

The effective sharing possibility offered by the Semantic Web would also open a series of interesting possibilities, such as the creation of communities of teaching professionals, who could share not only learning contents, but also assessment questions and the analysis of flaws and strengths of students could be analyzed not only for a particular group of a school but also at wider scale. This would also contribute to make the teaching–learning process homogeneous in different schools. To this end, we plan to facilitate the data sharing, publishing and consumption by using the Linked Open Data approach and that would transform OeLE into Semantic Web platform.

In summary, we have presented in this paper a system which generates semantic feedback for both teachers and students by applying Semantic Web technologies. According to our results, such feedback has been considered useful by both students and teachers. Consequently, the results are promising, but they correspond to a single experience. More experiences in courses from different knowledge areas should be made to analyze the quality, effects and usefulness of the semantic feedback generated by the OeLE platform.

Acknowledgements
This work has been possible thanks to the Seneca Foundation, through Grants 08756/PI/08 and 15295/PI/10.

References