Problem-based learning supported by semantic techniques

Esther Lozano [a] , Jorge Gracia [a] , Oscar Corcho [a] , Richard A.
Noble [b] & Asunción Gómez-Pérez [a]

[a] Ontology Engineering Group, Universidad Politécnica de Madrid,
Madrid, Spain

[b] Hull International Fisheries Institute, University of Hull, Hull,
HU6 7RX, UK
Version of record first published: 27 Nov 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

Problem-based learning supported by semantic techniques

Esther Lozano[a]*, Jorge Gracia[a], Oscar Corcho[a], Richard A. Noble[b] and Asunción Gómez-Pérez[a]

[a]*Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain;* [b]*Hull International Fisheries Institute, University of Hull, Hull HU6 7RX, UK*

Problem-based learning has been applied over the last three decades to a diverse range of learning environments. In this educational approach, different problems are posed to the learners so that they can develop different solutions while learning about the problem domain. When applied to conceptual modelling, and particularly to Qualitative Reasoning, the solutions to problems are models that represent the behaviour of a dynamic system. The learner's task then is to bridge the gap between their initial model, as their first attempt to represent the system, and the target models that provide solutions to that problem. We propose the use of semantic technologies and resources to help in bridging that gap by providing links to terminology and formal definitions, and matching techniques to allow learners to benefit from existing models.

Keywords: problem-based learning; qualitative reasoning; semantic grounding

1. Introduction

Over the last three decades, problem-based learning (PBL) has been successfully applied in both authentic and situated learning environments, and its use has been widely reported in a number of domains, e.g. medicine (Barrows, 2000), architecture (Maitland, 1998), or chemical engineering (Woods, 1996). This approach consists of posing problems to learners, so that they can learn about a particular domain by developing solutions to them. Problem-based learning provides learners an opportunity to develop skills in problem definition and problem solving, to reflect on their own learning, and develop a deep understanding of the domain content (Spiro, Coulson, Feltovich, & Anderson, 1988). When compared to approaches driven by textbooks or other prepared instructional materials (Pierce & Jones, 2000), PBL students remember the information for longer period (Norman & Schmidt, 1992), and are more likely to apply scientific knowledge appropriately in the future (Allen, Duch, & Groh, 1996).

In problem-based learning environments, some target solution(s) usually exist to the proposed problem. The learner's task consists of developing their own solutions to the problem, and testing/evaluating them against some existing target solutions(s). In doing so, they move from their initial solution to the target solution. However,

---

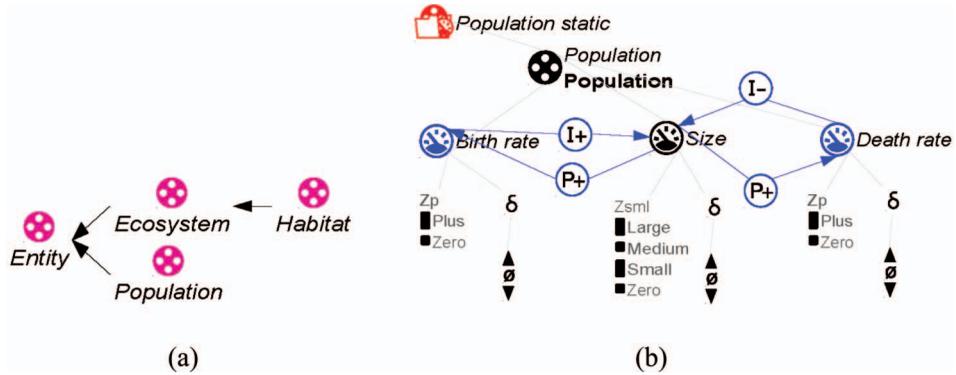*Corresponding author. Email: elozano@fi.upm.es

Figure 1.   Model of population growth based on competing processes. (a) Hierarchy of entities. (b) Relations between ingredients.

bridging the gap between these two solutions is not a trivial task, especially for complex domains or for those concepts that are very new to learners.

In this work, we focus on one of the strategies that have been commonly applied to problem-based learning for complex domains: conceptual modelling (Bredeweg & Forbus, 2004). In particular, we focus on Qualitative Reasoning (QR) modelling, which enables modellers to specify conceptual representations of complex, dynamic systems, and to predict their behaviour through reasoning (Forbus, 2008). In fact, QR has been already successfully applied for education across multiple disciplines, such as physics, economics, or environmental science.

In QR, two relevant limitations hamper the scaffolding of model building and the autonomy of the learner towards learning targets: (1) tools do not allow the alignment of terminology in the models to common shared vocabularies, thus leading to heterogeneous models that are more difficult to compare, and (2) tools do not enable the automated reuse of knowledge coming from other models, thus losing learning opportunities by discovering relevant knowledge from experts or peers that could enrich the learner model.

To illustrate this, let us consider a modelling task undertaken by a group of learners using one of the existing QR modelling tools. The students are given the task of representing in a QR model the behaviour of a population growth based on competing processes (e.g. Nuttle, Salles, Bredeweg, & Neumann, 2009) as shown in Figure 1. The learners use one of the existing modelling tools to build the model and start to define some concepts and their relations. At some point in time, the learners consider their models finished and require feedback on their models to check whether they have an appropriate solution to the task. In a typical problem-based learning task, the teacher will have also specified their own reference model representing a target solution to the problem. It can be anticipated that, with learners working individually they will have developed a diverse range of models, representing their individual understanding of the behaviour of the population. As the learners and the teacher have created their models independently, potentially using different terminologies and thus these different models are not easily aligned (for instance, a student may call "Death" the same concept that other student refers as "Mortality rate"). When the learners want to compare their models to detect whether there are any differences, they can only do it by manually inspecting the models. First, they need to agree on the terminology and align their models before being able to directly

compare them. Second, they also lack the autonomy to validate on their own solution. In fact, the student models have to be validated by the teacher, either by directly correcting them by hand or by providing clear rules to allow self-correction by the student. In any case, this is a costly process and typically does not benefit of the knowledge contained in other expert or peer models maybe not known by the teacher in advance.

1.1.   A proposal for semantically-enhanced QR-based PBL

In this paper, the use of some of the resources available in the Web of Data (Bizer, Heath, & Berners-Lee, 2009), more specifically DBpedia[1], is proposed to ease the modelling task and avoid some of the aforementioned limitations. Since QR models can be represented as ontological models (Liem & Bredeweg, 2007), the application of semantic-based techniques is possible. The approach consists of two main components: (1) grounding of terminologies and vocabulary in different models to a common vocabulary to align different models considering similar concepts and (2) the generation of semantic-based feedback to the modellers. Both components aim to ensure lexical correctness of their terms, as well as facilitate the interoperability among models at a terminological level. The goal, from an educational perspective, is twofold:

(1) Allow learners to link their particular terminology to well established vocabularies, thus having the opportunity of learning the correct terminology and modifying their model accordingly.
(2) Enable automatic model comparisons to discover pieces of common knowledge among the models as well as to detect possible discrepancies or misconceptions. Suggestions have to be created from these discrepancies and communicated as feedback to the learner. These suggestions are aimed to stimulate learners to evaluate their models and to scaffold their progress, modifying their models and understanding, directing their representations towards the representations found in reference models.

The rest of this manuscript is organised as follows. Section 2 provides background on Qualitative Reasoning and the zone of proximal development as scaffolding strategy for learning. Section 3 describes the system architecture and its main components. Sections 4 and 5 provide a detailed description of the semantic grounding and the feedback generation processes. Section 6 details the evaluations run to validate this approach. The related work is discussed in Section 7. Finally, conclusions and future work are presented in Section 8.

2.   Context

In this section, some preliminary concepts of QR modelling and development levels are introduced to allow a better understanding of the remainder of the paper.

2.1.   Qualitative Reasoning as a learning environment

Qualitative Reasoning modelling allows modellers to create non-numerical descriptions of system structure and behaviour concerning domain concepts and

phenomena. Qualitative Reasoning can support models that preserve important properties and qualitative behavioural distinctions of systems behaviour and can automate reasoning and problem solving about the physical world (Forbus, 2008). Qualitative Reasoning modelling requires that modellers define the model in terms of ingredients that can define both the system structure and notions of causality from a systems viewpoint. These ingredients include: entities, configurations, quantities and quantity spaces. Entities define the concepts with which the physical components/ structure of the system is described, typically organised in taxonomies (e.g. "Ecosystem" and "Population" in Figure 1(a) are two disjoint entities and "Habitat" is a subtype of "Ecosystem"). Configurations define structural relations between entities (e.g. part of, contains, and lives in), and quantities represent the features of entities that may change during simulation of dynamic behaviour. Quantities are defined by a set of possible qualitative values (quantity spaces). Quantity spaces represent the possible values a magnitude (or derivative) of a quantity can have, and contain an ordered set of possible values. All of these ingredients are defined by their type and by the name given to them by the modeller. Some of the ingredients described above can be put in relation by means of causal relationships, correspondences, inequalities, and other generic building blocks representing mechanisms of change within the model. For instance, in Figure 1(b), e.g. the positive influence (I+) from "Birth rate" of "Population" makes the size of that population increase. On the other hand, the "Size" of the population is propagated to "Birth rate" via a positive proportionality (P+). Similarly, the negative influence (I−) from "Death rate" makes the size of population decrease, while the "Size" is positively proportional to "Death rate".

Therefore, given that all ingredients in the QR approach are defined by their name and the terminology used by the modeller, it is this process of defining ingredients that is fundamental to the creation of models. It is also during modelling that learners will require feedback and support on how to construct an appropriate explanation for the behaviour of a certain system, and how to use and define domain-appropriate vocabulary.

2.2. *Development levels as a framework for scaffolding*

During the modelling process, the learners' models are constructed based on the modellers' current understanding of the system or phenomena they are representing. Therefore, this model can be considered to be an external representation of the modellers' understanding (mental model) of the system. The amount of knowledge a user has about a particular domain determines, together with their modelling skills, the quality of this representation and how much this model may differ from a reference or target model for the subject/problem. The assumption is that the closer the learner model is to the target one, the better the user model represents the reality. In this process, four classes of models can be identified, following a similar approach to (Tran, Cimiano, Rudolph, & Studer, 2008):

(1) **Mental model (M_m).** This corresponds to the learner's internal mental representation/understanding of the system that they intend to represent by means of a conceptual model. The mental model is composed of mental structures (from distinct and well structured concepts and relations to vaguely formulated ideas) that are relevant to the topic to represent.

The mental model is understood as the user's initial knowledge about the system.

(2) **Learner model (M_l)**. This is the result of the learner's intent to formalise and externalise their mental model in terms of a formal modelling vocabulary. In most cases, not all the knowledge contained in the mental model is successfully represented in the learner model, depending to a great extent on the learner's modelling skills and the expressiveness of the knowledge representation language.

(3) **Target model (M_t)**. This is an ideal (and typically unknown) formal representation of the studied system, assumed to be complete and to cover all aspects of system. There is no unique target model (as there is no unique view of the world). It depends on several factors such as scope, granularity and perspective.

(4) **Reference model (M_r)**. As the target model (ideal representation of the system) is typically unknown, an expert user's model is taken as reference model, and is considered an approximation of the target model.

Vygotsky (1978) proposed the concept of a *zone of proximal development* (ZPD), defined as "the distance between the actual development level as determined by independent problem solving and the level of potential development [...] in collaboration of more capable peers". In this context, M_m represents the initial point in the actual development level space, and M_l constitutes the first effort to formalise the learner's knowledge, while M_r is an intermediate point before M_t, which is typically unreachable from the current level of the learner (see Figure 2).

In our case, both learner and reference models are represented as QR models. Given this, the modelling process can be seen as a progression from an initial representation of the learners' understanding of the phenomena and the QR approach towards a better system understanding and greater ability with the QR tool. The systematic evolution of the learner model towards the reference model(s) should act to refine the learners' mental model of the system and to therefore enhance their understanding of the concept. To achieve these, learners will require scaffolding during their modelling task to support their appropriate use of domain knowledge and also their use of the QR vocabulary. Automation of this support and scaffolding requires that the QR tool can provide situated feedback and suggestions to the student on how their model could be improved. This support should provide suggestions that are within the zone of proximal development for the student.

Figure 2. Development levels.

3.   System overview

In traditional approaches driven by textbooks or other prepared instructional materials, the progression between M_l to M_t requires "manual" comparisons among models, which has inherent limitations in time (expert/teacher validation is time consuming) and scope (reference models has to be known beforehand). In this work, the use of two semantic techniques, semantic grounding and semantic-based feedback, is proposed to overcome these limitations. This approach aims to support a learner's transition from M_l to M_t by providing automated feedback based on "intermediate" reference models extracted from a pool of models created by the community of users.

To support such pool of models, an online semantic repository has been conceived where the models created by all the users are stored and remain accessible for later reuse. Then, during the construction of the learner model, the system can perform the following actions (see Figure 3):

- **Semantic grounding**. Determining links between the unrestricted terminology used by learners and resources of the Web of Data (particularly DBpedia), making easier the interoperability and later alignment among models.
- **Semantic-based feedback**. Analysis of the differences between the learner model and the reference model and generation of suggestions provided as feedback for the learner.

The DynaLearn[2] workbench, a system intended to allow learners to acquire conceptual knowledge, is used for creating the model. In this workbench, the conceptual knowledge contained in QR models is extracted (Liem & Bredeweg, 2007) and expressed in OWL to ease their ontology-based description. During the modelling process, the learner grounds the terms in their models to online semantic
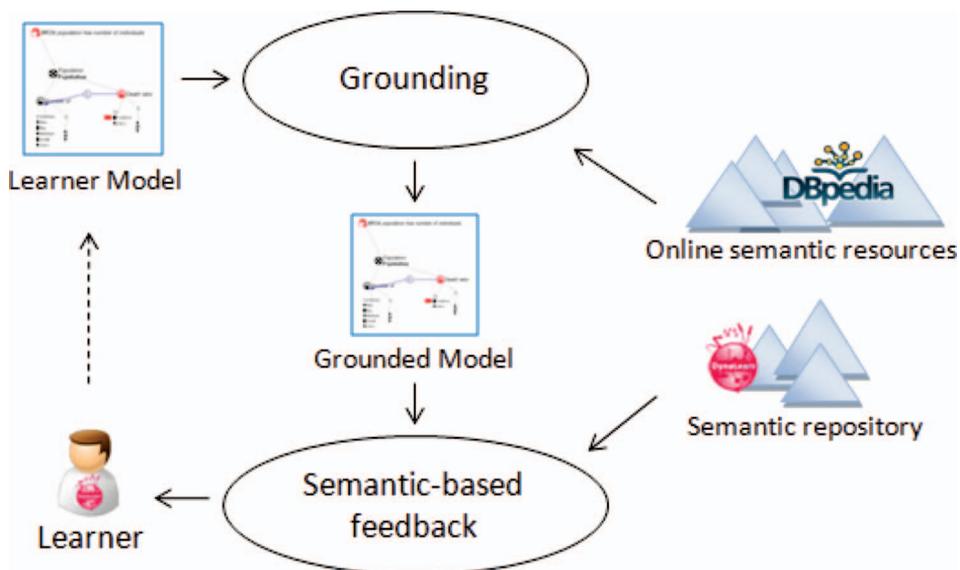

Figure 3.   General overview of the semantic technologies.

resources. The choice of DBpedia for that was motivated by its good coverage in comparison to other semantic resources (Gracia et al., 2010). At a certain point, the learner asks for semantic feedback on their model. Then, relevant models contained in the semantic repository are searched. The most relevant model is selected as reference model and compared with the learner's model for the generation of feedback. As a result, a list of discrepancies is obtained which are communicated back to the learner as suggestions to improve their model. The learner can then decide whether to follow such suggestions or not, thus changing or maintaining the model accordingly. The whole process should scaffold a learner to evaluate their own model and evolve their model towards capturing a more insightful understanding of the concepts they are modelling and the terminology they are using.

4.   Grounding of terms

The semantic grounding process allows conceptual models to be linked to other existing ones by means of common terms that come from external resources. Figure 4 illustrates the main steps of the semantic grounding process.

The first step in the grounding process is to normalise the term (e.g. replacing "_" symbols with blank spaces). Also, if the term is composed of multiple words, it is syntactically analysed[3] and split in the parts delimited by separators (e.g. "the change of energy flow" becomes "change" and "energy flow"). The original term and its parts constitute the list of words to be grounded. If one of these parts is still a multiword, their constituent words and their possible combinations are added to the list.

The next step is to look up in DBpedia the given list of words to be grounded. The obtained DBpedia terms form the list of candidate groundings. A ranking process is applied on this list of candidates. This performs comparisons, based on vector space modelling (Raghavan & Wong, 1986) and string-based distances (Levenshtein, 1966), between each candidate grounding and the original term and its context (related terms in the QR model). As a result, the candidate groundings are ranked according to their relevance to the QR model.

Additionally, a stemming process (Miller, 1995) is applied to get a canonical form from plurals, inflected forms, etc., and a spelling suggestion service[4] is accessed to get alternative forms in case of misspelling errors. These suggestions are communicated to the learner as alternatives to the original term. In case the candidate groundings were not satisfactory to the learner, they could repeat the search using one of the alternatives instead of their original term.

Finally, in order to cover the cases in which the groundings provided by DBpedia are not satisfactory, the system allows the users to create the so called *anchor terms*.
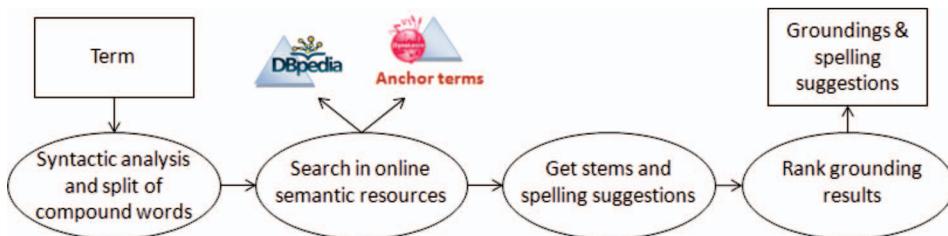

Figure 4.   Main steps of semantic grounding.

An anchor term contains the label and description of a user term that could not be grounded in DBpedia, and is stored apart in an online ontology that remains available for later reuse in other models. Therefore, this solution also enables the emergence of much specialised common vocabularies shared by the users of the system. This common vocabulary is also used as source during the search of groundings, and the results are added to the list of candidate groundings.

5. Semantic-based feedback

As a consequence of the preliminary grounding process, the comparison of models becomes a feasible task, and thus the possibility of generating valuable feedback for the learner from existing models. This feedback informs the learner about the discrepancies between their model and the model used as reference in the comparison. In particular, feedback is given on:

- *Terminological discrepancies*. During the modelling task, the selection of the appropriate vocabulary to be used in the model is first, and therefore very important, step. An incorrect choice of the terminology in the learner model could indicate a misunderstanding of the problem to be solved or a lack of knowledge in the particular domain. The comparison of the terminologies of the learner model and a reference model allows the detection of some of these problems.
- *Taxonomic discrepancies*. Two equivalent entities should share the same terminology but also have the same equivalent position in their respective hierarchies. The system informs of this inconsistent situation, so the learner could revise the hierarchy and change it accordingly.
- *Structural discrepancies*. Correctly representing the causal dependencies (influences and proportionalities) between quantities is a difficult task in QR modelling. By comparing the causal dependencies between equivalent quantities in the reference and the learner model, the relations that are missing, extra or different in the learner model can be detected.

5.1. Generation of feedback

The generation of feedback is based on the comparison between the learner model and a reference model. The selection of this reference model can be done in two manners: the teacher can select the model that learners have to pursue, or it can be left to the system to automatically select an appropriate reference from the models currently available based on the common groundings between the learner model and the models in the repository. The assumption here is that the existence of common groundings indicates that the models address similar concepts and share the same domain.

Figure 5 shows the different processes that compose the generation of semantic feedback.

5.1.1. Grounding-based alignment

First, the learner and reference models are compared using a grounding-based alignment, which identifies the common groundings between the two models. These common groundings are used to generate a preliminary set of mappings, which are added to the learner model as owl:EquivalentClass statements. As a result, pairs of

terms considered semantically equivalent are obtained, and these constitute the preliminary set of mappings. To illustrate this, we use Figure 1 and Figure 6 as examples of reference model and learner model, respectively. In this example, the learner's model has a term labelled *Birth* that is grounded to the DBpedia term *Birth rate*[5], and the reference model has a term labelled *Birth rate* that is also grounded to the same DBpedia resource *Birth rate*. To allow later inference, *Birth* and *Birth rate* are identified as equivalent terms (expressed using *owl:EquivalentClass*).

5.1.2. *Ontology matching*

Then, ontology matching techniques (Gracia, Bernad, & Mena, 2011) are applied between the enriched version of the learner model and the reference model to discover new equivalences beyond the common groundings, based on the semantic similarity of the terms. These new equivalences, together with the grounding-based ones form the final set of mappings. These techniques have been adapted to the QR
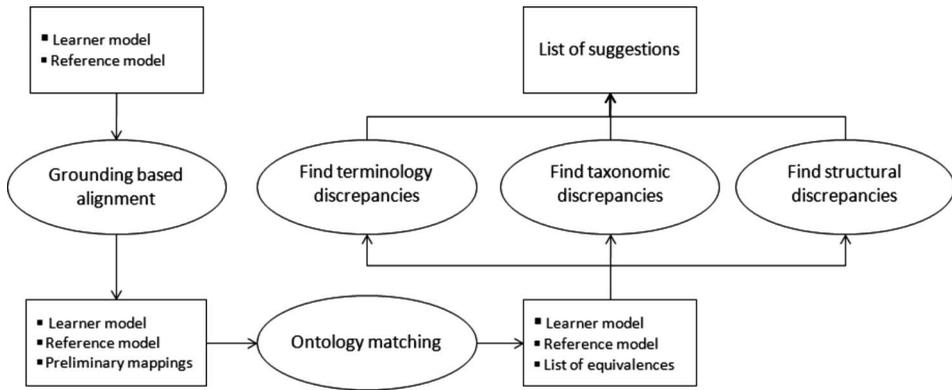

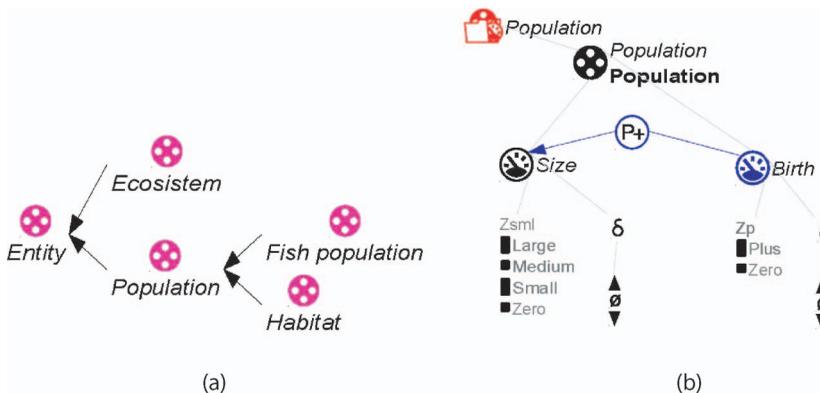Figure 5. Generation of semantic feedback.

Figure 6. Learner model used in the examples. (a) *Ecosystem* and *Population* are two disjoint entities, and *Habitat* and *Fish population* are subtype of *Population* and disjoint between them too. (b) The *Birth* of *Population* is propagated to *Size* via a positive proporcionality (P+).

scenario and applied to the models, which are treated as ontologies (Liem & Bredeweg, 2007). The obtained mappings are added to the learner model as *owl:EquivalentClass* statements.

Finally, the system compares the pairs of equivalent terms, obtained from the ontology matching, to detect the possible differences between them. This detection is carried out through three different processes, according to the type of difference we want to find:

5.1.3. Detection of terminological discrepancies

Each pair of equivalent terms is analysed to find the possible terminological discrepancies between the models. Depending on the type of differences, the system finds:

- *Discrepancies between labels or groundings:* The system compares the label and grounding of each pair of equivalent terms and detects the differences. In our example, the system would suggest that the quantity *Birth* in the learner's model could be renamed as *Birth_rate*.
- *Missing and extra ontological elements:* Those terms from the reference model that are not present in the learner model are seen as missing elements. The system would suggest to add them to the learner model, indicating their relation with the existing elements in the model. In the same way, the terms in the learner model that are not present in the reference model could be seen as extra elements to be removed from the model. In our example, the entity *Fish population* would be suggested as extra since it does not appear in the reference model and could be removed from learner's model.

5.1.4. Detection of taxonomic discrepancies

After the ontology matching process, the tool uses semantic reasoning techniques (Sirin, Parsia, Cuenca-Grau, Kalyanpur, & Katz, 2007) to detect inconsistencies between the taxonomy of the two models. For instance, the entity *Habitat* cannot be subtype of *Population* in learner's model and subtype of *Ecosystem* in the reference model given that *Population* and *Ecosystem* are disjoint classes.

5.1.5. Detection of structural discrepancies

Once the equivalences at the schema level have been identified, the system compares the models at the instance level. The discrepancies between the instances of equivalent concepts point out differences in how these concepts are used in the models. The particular semantics of QR models is used, as well as some QR modelling rules, to identify common patterns in the models and detect the possible discrepancies:

- *Missing and extra instances of existing terms in the learner model:* A missing instance in the learner model could indicate some missing behaviour in the model. Similarly, an extra instance could represent a superfluous or incorrect behaviour in the model. Let us assume that the learner has already defined the quantity *Death* in their model, though it is not being used. Then, the system

suggests that, according to the reference model, an instance of *Death* should be added to the entity *Population* in their model.

- *Discrepancies in the causal dependencies between quantities:* By comparing the causal dependencies between equivalent quantities in the reference and the learner model, the relations that are missing, extra or different in the learner model can be detected. In the example, the quantities *Size* and *Birth* in learner's model are causally related by a positive proportionality. However, according to the reference model they should be related through a positive influence, and a positive proportionality only applies in the opposite direction.

Finally, to scaffold the learners though their zones of proximal development, the detected discrepancies are expressed in the form of suggestions, as the actions that learner should do to modify their model towards the target model. These suggestions are communicated to the learner through a particular interface within the modelling tool, indicating the involved terms in the learner model and a little description of the detected problem.

6.   Evaluation

There has been previous work showing the adequacy of DBpedia as a source of knowledge for grounding QR models (Gracia et al., 2010), with up to 78% of coverage in the domain of environmental science. We focus here on studying the effect of the grounding and the other proposed semantic techniques in a real learning scenario[6]. In particular, the research hypotheses to be tested are the following:

- H1   The system is capable of generating suitable feedback from the comparison of a learner model and a reference model.
- H2   The generated feedback supports model development and learning through the zone of proximal development.

6.1.   Evaluation 1: adequacy of generated feedback

To evaluate the adequacy of the generation of feedback, 10 pairs of expert and learner models[7] were taken in the domain of Natural Sciences, all of them previously grounded. The learner models were obtained from two different groups of learners, all beginners in QR modelling. Two domain experts participated in the evaluation, both experts in QR modelling.

Each pair of models was used as input for the semantic feedback. A total of 156 test cases (suggestions) were obtained. For each pair of models, the images of the learner and reference model were shown to the evaluators, together with the list of suggestions generated as feedback. Then, the evaluators assessed whether they agreed or not with the proposed suggestions. Some examples of the generated suggestions are:

- Terminology: The entity "Shallow water" in your model should be renamed as "Surface water".
- Inconsistency: "Osmosis" appears as quantity instead of entity in the reference model.

- Different dependency: The quantity "Population size" of the entity "Fish" should have a positive proportionality instead of a positive influence to the quantity "Death and sinking rate" of the entity "Fish".
- Missing term: The entity "Cytoplasm" grounded to http://dbpedia.org/resource/Cytoplasm should be added to your model.

As a result, 87% of the generated suggestions were considered correct by the experts. The Kappa coefficient, that measures the level of agreement between evaluators, was 0.68, which indicates an adequate level of agreement.

6.2. *Evaluation 2: effect of the feedback in the learners' progress*

To test the second hypotheses, an experiment was set up using the tool DynaLearn in a real classroom scenario. The experiment was run simultaneously in two European universities: University of Hull (Group 1, seven undergraduate students from the department of Biological Science) and Universidad Politécnica de Madrid (Group 2, six students from the department of Artificial Intelligence). In both cases, students were unfamiliar to QR and received a couple of training lessons in QR modelling before the evaluation activity. Then, they were asked to elaborate a model about a certain topic which was described in natural language. The topic proposed for Group 1 was to study the "interaction of the brine shrimp Artemia with a hypersaline environment" (extracted from an undergraduate physiology book), while for Group 2 was the "physics of the flow of a liquid between two communicating vessels". Once the students considered their model complete, they grounded the terms in DBpedia and asked for feedback. Then, they modified their models by accepting or not each of the given suggestions and iterated again until they considered the model finished.

In a first attempt to evaluate the automatic selection of reference models for the generation of feedback, Group 1 was divided into two subgroups. Four of the students received feedback from reference models automatically selected among the whole pool of models (that contained 190 models at the time of the evaluation). The other three students received feedback from a single reference model selected beforehand by the teacher, as all the learners from Group 2.

To assess whether the final models of the students (after feedback) were closer than the initial model (before feedback) to the reference models, the opinion of two experts in QR were asked. In order to avoid bias, each pair of models to be assessed were randomised (the expert did not know which model was generated after/before feedback). The experts scored how close the inspected model was to the reference one, in terms of the adequate conceptualisation, terminology and representation of causal dependencies for a pre-defined list of key concepts related to the system being modelled. Tables 1 and 2 show the results. "Improvement/worsening" indicates that, according to the evaluator, the model produced after feedback was closer/farther than the reference one. "Neutral" means that the changes did indicate neither a significant improvement nor worsening.

In order to evaluate the inter-rate agreement between the evaluators, we considered the amount of test cases in which both evaluators detected a correct progression or at least stability, thus rating the case as "improvement" or "neutral". In that case, the direct agreement between the evaluators was 77% for the cases that showed some improvement or at least remained as far from the reference model than

Table 1. Test results for Group 1.

Evaluator	Improvement	Neutral	Worsening
Evaluator 1	4	3	0
Evaluator 2	0	5	2
Total	4	8	2

Table 2. Test results for Group 2.

Evaluator	Improvement	Neutral	Worsening
Evaluator 1	3	2	1
Evaluator 2	4	2	0
Total	7	4	1

before the feedback. The corresponding Kappa was 0.54, what indicates moderate agreement. In fact, Evaluator 2 was much more conservative when considering potential improvements than Evaluator 1. When inspecting the cases that really presented some improvement (i.e. the cases rated as "improvement" by at least one evaluator), the direct agreement was 62%.

6.3. Discussion

The first experiment was aimed at confirming the hypothesis H1. As a result of this experiment, 87% of the evaluated suggestions were considered adequate by the experts, which can be considered as a promising result. The main problems identified in the learner models were:

- Representation of multiple concepts in the same term. For instance, "Water concentration in mucus" was represented as one quantity. However, this could be split into the individual concepts and represent "Water" and "Mucus" as entities and "Concentration" as a quantity of "Water".
- Representation of concepts using different types of ingredients. For example, Osmosis was a quantity in the reference model, though it appeared as entity in some learner models.
- Learners did not fully ground their models. By grounding the terms, some of the above problems can be detected, since not satisfactory groundings can be found for the term and learners can then reformulate the labels. In addition, the grounding facilitates the later alignment with the reference model.

The strong differences between the learner models and the reference model indicate a big gap between the learner's background in QR and very complex models. This makes difficult the task of ontology matching, thus obtaining few mappings between the models, and hence limits the potential of the feedback. In fact, the problem here was that the reference model was too advanced for those learners and with different granularity. Some terminology issues can be still detected and suggest some changes to the learners, but most of structural problems in the model

would be hard to detect when the reference model is so distant and with such low overlap with the learner model. There are two ways of alleviating this issue: either using the feedback mechanism when learners have a better insight of QR or introducing simpler reference models in the repository.

To confirm the hypothesis H2, the second experiment was run. Although improvement was detected in 62% of cases after feedback, the results are still modest. In particular, the progression of learners was particularly slow for participants of Group 1. Analysing the data for this group, we found that the models had a very low quality and revealed an important lack of knowledge about the domain to be modelled. Although learners from both groups were beginners in QR modelling, participants from Group 2 did know the problem better (communicating vessels), while Group 1 were new in the domain, since the objective of that activity was to learn about it. As a consequence, initial models in Group 1 were long way away from the reference model. In these cases, the reference model could be considered out of their zone of proximal development. Therefore, although they made changes it may not have been sufficient to improve. The solution to this problem could be either to improve the initial knowledge of the learner (thus approaching their user models to the reference one) or generate an easier version of the reference model (thus approaching the reference model to the learner models).

About the comparison between the two different ways to obtain reference models, automatically selected by the system or pre-selected by the teacher, there was no observable difference in the quality of the feedback generated. However, further experimentation will continue in this regard.

7.   Related work

Currently, there are a range of different modelling tools based on QR that have been developed for learning purposes (Bredeweg, Linnebank, Bouwer, & Liem, 2009; Dehghani & Forbus, 2009; Forbus, Carney, Sherin, & Ureel II, 2004; Soloway et al., 1997), though to our best knowledge the approach described in this paper has no counterpart in the field of QR modelling. Other modelling and simulation tools, such as Betty's brain (Leelawong & Biswas, 2008) or Stella (Costanza & Voinov, 2001) neither ground terms to a common vocabulary nor get quality feedback from existing models. There exist other modelling tools as well based on conceptual modelling in general, beyond QR (Cañas et al., 2004; Novak & Gowin, 1984), though these do not rely on Semantic Web standards to maximise their interoperability or not perform an effective grounding with external concepts.

In the field of ontology engineering, there are some works aimed to improve an ontology through techniques of ontology debugging (Stuckenschmidt, 2008) and ontology evolution (Stojanovic, 2004). These systems analyse the formalisation and implementation of the ontology to detect the possible inconsistencies. An approach to the sequential diagnosis of ontologies is presented in Shchekotykhin and Friedrich (2010). Lehmann and Bümann (2010) describe an ontology debugging tool to repair inconsistent ontologies and enrich them. However, these approaches are mostly based on diagnosis methods (Friedrich & Shchekotykhin, 2005; Kalyanpur, Parsia, Horridge, & Sirin, 2007) and they do not provide recommendations based on run-time discovered shared domain knowledge, as our system does.

8.   Conclusions and future work

In this paper, a novel approach is presented to support learning in problem solving environments by the use of semantic techniques, and in the particular scenario of using QR models. This approach links the learners' models to common vocabularies to facilitate their later alignment, thus easing the comparison of models to detect the possible differences between them. These differences are expressed as suggestions and communicated back to the learner as feedback, aim to scaffold the learner's progress through their zone of proximal development.

The human-based evaluations run with an early implementation of the feedback utilities showed promising results although it is suggested that a more detailed scaffolding of naive modellers is needed to ensure first proper terminology, and then proper usage of those terms in the model.

Regarding the grounding of terms to common vocabularies, additional knowledge and resources can be incorporated to the system. For example, DBpedia contains rich multilingual textual descriptions, links to pictures and web pages, etc. as part of a term description. This information can be imported if the term is grounded on that knowledge source, and shown to the user in the modelling tool.

Finally, we will run more detailed evaluations focused on particular aspects of the modelling process and their effect during the generation of feedback. In this context, we will study the influence of anchor terms created by learners, as well as the common modelling problems that could be taken into for a more specialised feedback.

Notes

1. http://dbpedia.org
2. www.dynalearn.eu
3. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/
4. http://code.google.com/p/google-api-spelling-java/
5. http://dbpedia.org/resource/Birth/_rate
6. All the experimental data and more details about the evaluation can be found in www.oeg-upm.net/files/dynalearn/experiments/2012/Evaluations.rar
7. Obtained as result of real modelling exercises conducted at University of Hull and in Tel Aviv University.

Notes on contributors

Esther Lozano is a PhD student within the Ontology Engineering Group (OEG) at Departamento de Inteligencia Artificial) Facultad de Informática, Universidad Politécnica de Madrid (Artificial Intelligence Department, School of Computer Science, UPM), where she attained a PhD research grant of the Spanish Ministry of Education. She also belongs to the Ontology Engineering Group (OEG) of that faculty. Previously she worked as AI research assistant in Fundació Clínic per a la Recerca Biomèdica. She holds a degree in Computer Science (Autonomous University of Barcelona, 2008), where she was awarded with the Academic merit Recognition of that year, and a MSc in Artificial Intelligence (Technical University of Catalonia, 2009). Her research activities are focused in Semantic Web and Ontological Engineering. Esther participated in the European project DynaLearn for the creation of engaging and informed tools for learning conceptual system knowledge.

Jorge Gracia is a researcher in the Ontology Engineering Group (OEG) at Departamento de Inteligencia, Artificial Facultad de Informática, Universidad Politécnica de Madrid (Artificial Intelligence Department, School of Computer Science, UPM) where he attained a Juan de la Cierva postdoctoral position. From 2004 to 2009 he worked at the University of Zaragoza, teaching and researching as Part-time Lecturer and as *Ayudante* (Assistant). He also stayed for six months at Knowledge Media Institute (Open University, United Kingdom) as Visiting Researcher. From 1998 to 2003, he worked at companies belonging to the **Accenture** group and the **Telecom Italia** group, taking part in software engineering projects for important telecommunications and service firms (Retevisión, Auna, Danone).

Oscar Corcho is an Associate Professor at Departamento de Inteligencia, Artificial Facultad de Informática, Universidad Politécnica de Madrid (Artificial Intelligence Department, School of Computer Science, UPM), and he belongs to the Ontology Engineering Group. His research activities are focused on Semantic e-Science and Real World Internet, although he also makes research in the more general areas of Semantic Web and Ontological Engineering. In these areas, he has participated in a number of EU projects (SemsorGrid4Env, ADMIRE, OntoGrid, Esperonto, Knowledge Web and OntoWeb), and Spanish R&D projects (CENITS mIO! and Buscamedia, GeoBuddies), and has also participated in privately-funded projects like ICPS (International Classification of Patient Safety), funded by the World Health Organisation, and HALO, funded by Vulcan Inc. Previously, he worked as a Marie Curie research fellow at the University of Manchester, and was a research manager at iSOCO. He holds a degree in Computer Science, an MSc in Software Engineering and a PhD in Computational Science and Artificial Intelligence from UPM. He was awarded the Third National Award by the Spanish Ministry of Education in 2001. He has published two books, ''Ontological Engineering'' and "A layered declarative approach to ontology translation with knowledge preservation", and more than 50 papers in journals, conferences and workshops. He usually participates in the organisation or in the programme committees of relevant international conferences and workshops. In 2011, he will be the programme chair of the KCAP conference.

Richard A. Noble has a diverse research interest in the field of fish ecology and environmental management. His research includes: fisheries science; the role of fish communities in wetland conservation; ecological assessment of rivers and lakes using fish assemblages; the role of hydro-morphology in structuring fish communities and population dynamics; and the management and rehabilitation of aquatic environments for fish. In addition to this, he has developed an interest in qualitative reasoning modelling in ecology and the role conceptual modelling can play in the enhancement of environmental science education.

Prof. Dr. Asunción Gómez-Pérez is Full Professor at the Universidad Politécnica de Madrid (UPM). She is the Director of the Artificial Intelligence Department (2008) and Director or the OEG at UPM (1995). She has a BA in Computer Science (1990), MSc on Knowledge Engineering (1991), PhD in Computer Sciences (1993) and MSc on Business Administration (1994). She was visiting (1994–1995) the Knowledge Systems Laboratory at Stanford University. She also was the Executive Director (1995–1998) of the Artificial Intelligent Laboratory at the School. She is a member of the Technical Committee of Standardization of the Spanish Association for Standardisation and Certification (AENOR) number 148 (AEN/CTN148) Geographic Information and participates in the Working Group developing the ISO 19150 standard. Her main research areas are: Ontological Engineering, Semantic Web and Knowledge Management. She led at UPM the following EU projects: Ontoweb, Esperonto, Knowledge Web, NeOn, SEEMP, OntoGrid, ADMIRE, DynaLearn, SemSorGrid4Env, SEALS and MONNET. She coordinated OntoGrid and now she is coordinating SemSorGrid4Env and SEALS. She is also leading at UPM projects funded by Spanish agencies. The most relevant are: España Virtual, WEBn+1, GeoBuddies and the Spanish network on Semantic Web. She participates in Linked Data Initiatives as Linked Data at the Spanish National Library. She has published more than 150 papers and she is author of one book on Ontological Engineering and co-author of a book on Knowledge Engineering. She has been co-director of the summer school on Ontological Engineering and the Semantic Web since 2003 up to now. She was program chair of ASWC'09, ESWC'05 and EKAW'02 and co-organiser of many workshops on ontologies. She reviews papers in many conferences and journals.

References

Allen, D.E., Duch, B.J., & Groh, S.E. (1996). The power of problem-based learning in teaching introductory science courses. *New Directions for Teaching and Learning*, 68, 43–52.

Barrows, H.S. (2000). *Problem-based learning applied to medical education*. Springfield, IL: Southern Illinois University School of Medicine.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. In T. Heath, M. Hepp, & C. Bizer (Ed.), *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5, 1–22.

Bredeweg, B. & Forbus, K.D. (2004). Qualitative modelling in education. *AI Magazine*, 24, 35–46.

Bredeweg, B., Linnebank, F., Bouwer, A., & Liem, L. (2009). Garp3 – workbench for qualitative modelling and simulation. *Ecological Informatics*, 4, 263–281.

Cañas, A.J., Hill, G., Carff, R., Suri, N., Lott, J., Gómez, G., ... Carvajal, R. (2004). *Cmaptools: A knowledge modeling and sharing environment*. Paper presented at the first international conference on concept mapping, Pamplona, Spain.

Costanza, R., & Voinov, A. (2001). Modeling ecological and economic systems with Stella: Part III. *Ecological Modelling*, 143 (1–2), 1–7.

Dehghani, M., & Forbus, K.D. (2009). QCM: A QP-based concept map system. In J. Zabkar & I. Bratko (Eds.), *Proceedings of the 23rd international workshop on qualitative reasoning (QR09)*, Ljubljana, Slovenia (pp. 16–21). Ljubljana: University of Ljubljana.

Forbus, K.D. (2008). Qualitative modelling. In F. van Harmelen, V. Lifschitz, & B. Porter (Eds.), *Handbook of knowledge representation* (pp. 361–393). Amsterdam, The Netherlands: Elsevier.

Forbus, K.D., Carney, K., Sherin, B.L., & Ureel II, L.C. (2004). *Vmodel: A visual qualitative modeling environment for middle-school students*. Paper presented at the 16th conference on innovative applications of artificial intelligence, San Jose, California.

Friedrich, G., & Shchekotykhin, K. (2005). A general diagnosis method for ontologies. In Y. Gil, E. Motta, V.R. Benjamins, & M. Musen (Eds.), *Proceedings of the 4th international semantic web conference (ISWC-05)* (pp. 232–246). Berlin Heidelberg: Springer-Verlag.

Gracia, J., Bernad, J., & Mena, E. (2011). *Ontology matching with CIDER: Evaluation report for OAEI 2011*. Paper presented at the 6th ontology matching workshop (OM'11) at ISWC'11, Bonn, Germany, CEUR-WS.

Gracia, J., Liem, J., Lozano, E., Corcho, O., Trna, M., Gómez-Pérez, A., & Bredeweg, B. (2010). Semantic techniques for enabling knowledge reuse in conceptual modelling. In P.F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J.Z. Pan, ... B. Glimm (Eds.), *Proceedings of the 9th international semantic web conference (ISWC2010)*, Shanghai, China (pp. 82–97). Berlin Heidelberg: Springer-Verlag.

Kalyanpur, A., Parsia, B., Horridge, M., & Sirin, E. (2007). Finding all justifications of OWL DL entailments. In K. Aberer, K.S. Choi, N. Noy, D. Allemang, K.I. Lee, L. Nixon, ... P. Cudré-Mauroux (Eds.), *Proceedings of the 6th international semantic web conference and 2nd Asian semantic web conference (ISWC/ASWC–07)* (pp. 267–280). Berlin Heidelberg: Springer-Verlag.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's brain system. *International Journal of Artificial Intelligence*, 18, 181–208.

Lehmann, J., & Bümann, L. (2010). Ore – A tool for repairing and enriching knowledge bases. In P.F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J.Z. Pan, ... B. Glimm (Eds.), *Proceedings of the 9th international semantic web conference (ISWC2010)*, volume 6497 of Lecture Notes in Computer Science (pp. 177–193). Berlin Heidelberg: Springer-Verlag.

Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10, 707–710.

Liem, J., & Bredeweg, B. (2007). OWL and qualitative reasoning models. In C. Freksa, M. Kohlhase, & K. Schill (Ed.), *KI 2006: Advances in artificial intelligence. Proceedings of the 29th annual German conference on AI*, Bremen, Germany, number 4314 in Lecture Notes in artificial intelligence (pp. 33–48), Berlin Heidelberg: Springer-Verlag.

Maitland, B. (1998). Problem-based learning for an architecture degree. In D. Boud & G. Feletti (Ed.), *The challenge of problem-based learning,* 2nd ed. (pp. 211–217). London: Kogan Page.

Miller, G.A. (1995). A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Norman, G.R., & Schmidt, H.G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine*, *67*, 557–565.

Novak, J., & Gowin, D. (1984). *Learning how to learn*. Cambridge, UK: Cambridge University Press.

Nuttle, T., Salles, P., Bredeweg, B., & Neumann, M. (2009). Representing and managing uncertainty in qualitative ecological models. *Ecological Informatics*, *4*, 358–366.

Pierce, J.W., & Jones, B.F. (2000). Problem-based learning: Learning and teaching in the context of problems. In *Contextual teaching and learning: Preparing teachers to enhance student success in the workplace and beyond* (pp. 69–95). Columbus, OH: ERIC Clearinghouse on Adult, Career and Vocational Education.

Raghavan, V., & Wong. S. (1986). A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Science*, *37*, 279–287.

Shchekotykhin, K., & Friedrich, G. (2010). Query strategy for sequential ontology debugging. In P.F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J.Z. Pan, . . . B. Glimm (Eds.), *Proceedings of the 9th international semantic web conference (ISWC2010),* volume 6497 of Lecture Notes in Computer Science (pp. 696–712). Berlin Heidelberg: Springer-Verlag.

Sirin, E., Parsia, B., Cuenca-Grau, B., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, *5*, 51–53.

Soloway, E., Pryor, A.Z., Krajcik, J.S., Jackson, S., Stratford, S.J., Wisnudel, M., & Klein, J.T. (1997). Scienceware's model-it: Technology to support authentic science inquiry. *Technological Horizons on Education*, *25*, 54–56.

Spiro, R.J., Coulson, R.L., Feltovich, P.J., & Anderson, D. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In V. Patel (Ed.), *Proceedings of the 10th annual conference of the cognitive science theory* (pp. 375–383). Mahwah, NJ: Erlbaum.

Stojanovic, L. (2004). *Methods and tools for ontology evolution* (PhD thesis). University of Karlsruhe, Germany.

Stuckenschmidt, H. (2008). *Debugging owl ontologies – A reality check*. Paper presented at the 6th workshop on evaluation of ontology-based tools (EON'08), Tenerife, Spain.

Tran, T., Cimiano, P., Rudolph, S., & Studer, R. (2008). Ontology-based interpretation of keywords for semantic search. In K. Aberer, K.S. Choi, N. Noy, D. Allemang, K.I. Lee, L. Nixon, . . . P. Cudré-Mauroux (Eds.), *Proceedings of the 6th international semantic web conference and 2nd Asian semantic web conference (ISWC/ASWC–07)* (pp. 523–536). Berlin Heidelberg: Springer-Verlag.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Woods, D.R. (1996). Problem-based learning for large classes in chemical engineering. In L. Wilkerson & H. Gijselaers (Eds.), *Bringing problem-based learning to higher education: Theory and practice* (pp. 91–99). San Francisco, CA: Jossey-Bass.