

# Ontology Extraction Tools: an Empirical Study with Educators

Marek Hatala, Dragan Gasevic, Melody Siadaty, Jelena Jovanovic, Carlo Torniai

**Abstract**—Recent research in Technology-Enhanced Learning (TEL) demonstrated several important benefits that semantic technologies can bring to the TEL domain. An underlying assumption for most of these research efforts is the existence of a domain ontology. The second unspoken assumption follows that educators will build domain ontologies for their courses. However, ontologies are hard to build, especially for not-tech-savvy users. Tools for Ontology extraction from text aim to overcome this problem. We have conducted an empirical study with educators, both from Information Technology (IT) and non-IT domains, where they used current ontology extraction tools to build domain ontologies for their courses from their existing course material. Based on the obtained study results we have drawn conclusions about the existing ontology extraction tools and provided recommendations for their future development so that they can be beneficial for the TEL domain.

**Index Terms**—I.2.4 Knowledge Representation Formalisms and Methods; N.5.c Ontologies; H.5.2.e Evaluation/methodology; K.3.0 General



## 1 INTRODUCTION

For over a decade, an increasing number of TEL (Technology-Enhanced Learning) researchers have been exploring the potentials of Semantic Web technologies and more recently, the Linked Data paradigm for the development of advanced e-learning solutions [1],[2],[3]. Due to their inherent capacity to add explicit semantics to diverse kinds of data, as well as to interlink and/or integrate data originating from disparate sources [4], these technologies have been recognized as a promising technical foundation for the realization of personal learning environments [5]. In addition, by adding a reasoning layer on top of the semantically rich and interlinked data, different kinds and levels of adaptation and personalization in traditional e-learning systems are introduced [6],[7].

Ontologies, a mean for formally expressing the shared semantics of a certain domain, are the cornerstone of any Semantic Web based solution. In our previous research, we have explored the advantages of ontology supported e-learning systems. In particular, we demonstrated how a combined use of content structure ontology, content type ontology and domain ontology could significantly improve the search over learning content repositories [8]. We have also shown that if these three kinds of ontologies are complemented with a user model ontology and an ontol-

ogy formally specifying the learning path to be followed by a student, then advanced levels of learning content personalization can be achieved [9], as well. Furthermore, we demonstrated the relevancy of the integrated use of these different kinds of e-learning ontologies for providing online educators with reliable, fine grained and semantically rich feedback about the learning process [10],[11],[12]. Finally, in our most recent research efforts we made use of ontologies and Linked Data paradigm to develop a personal learning environment for collaborative learning of software design patterns [13].

The main problem with all the approaches that make use of ontologies to offer advanced TEL solutions is their assumption that the required ontologies are available or easy to develop. However, based on the experience of the TEL research community [1],[2] and our own experience from the abovementioned projects, this assumption is not realistic. We found that the major obstacle for widespread use of ontologies in e-learning systems lies in the complexity of the ontology development process, especially when considered from the perspective of educators who are typically unaware of ontology existence and its relevancy altogether. Although, recently, the Semantic Web research has shown a constantly increasing interest in automating ontology development and thus reducing the required human effort [14],[15], fully automatic ontology development is still in the distant future [8],[16],[17].

In this regard, domain ontologies, i.e., ontologies formally specifying concepts and relationships of a specific subject domain (e.g., a learning course), are the most challenging. Our experience, as well as the one of some other researchers in the field (e.g., [18],[19],[20]) have proven that unlike other kinds of ontologies relevant for e-learning, a domain ontology often cannot be directly reused even within the same subject domain. This is due to different requirements of the systems that make use of domain ontologies (e.g., different desired level of com-

- M. Hatala is with the School of Interactive Arts and Technology, Simon Fraser University, Surrey, Canada. E-mail: [mhatala@sfu.ca](mailto:mhatala@sfu.ca)
- D. Gasevic is with the School of Computing and Information Systems, Athabasca University, Athabasca, Canada. E-mail: [dgasevic@acm.org](mailto:dgasevic@acm.org)
- M. Siadaty is with the School of Interactive Arts and Technology, Simon Fraser University, Surrey, Canada. E-mail: [m.siadaty@gmail.com](mailto:m.siadaty@gmail.com)
- J. Jovanovic is with the Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia. E-mail: [jeljov@gmail.com](mailto:jeljov@gmail.com)
- C. Torniai is with Oregon Health & Science University, Portland, USA. Email: [carlotorniai@gmail.com](mailto:carlotorniai@gmail.com)

Manuscript received (insert date of submission if desired). Please note that all acknowledgments should be placed at the end of the paper, before the bibliography.

prehensiveness and/or complexity of the ontology, different ontology language) and/or different conceptualizations of the same domain by different educators. Therefore, domain ontologies often have to be either created anew or adapted to the needs of the specific context of their intended use. Another feature that makes domain ontologies distinct from other aforementioned ontologies relevant for e-learning is the need for their constant involvement, so that the semantics they capture do not lag behind the courses they are aimed to support.

Accordingly, a significant topic in our current research is to investigate how to reduce the efforts required for creating domain ontologies in educational systems, and thus implicitly enable easier and wider acceptance of ontology based systems among educational practitioners. Our first step towards achieving this goal was to explore the existing approaches for ontology development. Having done that, we were able to distinguish three general approaches: 1) handcrafting ontologies from scratch, 2) (semi-)automatic ontology development using ontology learning tools, and 3) search and retrieval of ontologies from online ontology libraries and Linked Open Data (LOD) cloud (<http://lod-cloud.net/>). We briefly report on these approaches and related tools in the following section (Literature Review).

The second step towards our research goal was to define the requirements that a domain ontology should satisfy in order to serve as a foundation for the development and functioning of an advanced e-learning system. These requirements were derived from our previous research work in the area of personalized learning supported by Semantic Web technologies (e.g., [5],[9], [10],[12],[13],[38]) as well as the work of other researchers in this area (e.g., [1],[2],[3],[6],[7]). In particular, we identified the following requirements:

- The ontology comprises all relevant concepts of the corresponding subject domain (i.e., the study course).
- The more detailed the ontology (i.e., the more domain specific concepts it has), the better, since an e-learning system would be able to provide better (more accurate) personalization [10].
- The ontology comprises relationships between domain concepts.
- The ontology provides good coverage of the entire course, meaning that it does not cover just one part of the course, while other parts are just barely covered or not covered at all.
- High semantic richness (in terms of ontology axioms) is a desirable but not a mandatory characteristic since advanced functionalities of an e-learning system (e.g., recommendation of learning resources, and provision of feedback) can be achieved with a rather simple domain ontology [5],[9],[11].

In the study that is the focus of this paper we investigated the (semi-)automatic ontology development. Specifically, the main goal of the study was to evaluate the approaches used by and readiness of the ontology extraction tools to be used for the development of domain ontologies that meet the identified set of requirements specific for advanced e-learning systems. Two research findings

guided our orientation towards ontology learning tools. First, a great majority of educators, being non-technically-savvy, would experience difficulties in using conventional ontology editors[21] and thus, may benefit from the suggested ontological elements that would provide them with some starting points in the process of ontology building[22]. Secondly, in our context of interest (i.e., e-learning), an extensive amount of domain related material is typically available in the form of electronic texts, slides, handouts, and the like, which can be used as an input for the ontology learning tools.

Even after a comprehensive literature review (including, for example, [23], [14]), we could not identify any research aimed at evaluating the level of adoption of tools for (semi-)automatic ontology development among end users and identifying the requirements for enabling their widespread use. The closest work we could find was the evaluation study done by Park et al. [16]. It was based on a comprehensive framework the authors proposed for the evaluation of ontology extraction tools. The experiment consisted of two parts: first, using the proposed evaluation framework, the authors themselves assessed four ontology extraction tools selected for the study; in the second part, four expert users assessed and ranked the considered tools using the Analytic Hierarchy Process method. However, the proposed evaluation framework and the experiment did not cater for the specific requirements of e-learning systems and their users. Furthermore, the assessment of the tools was done by ontology researchers and not real end-users (i.e., educators who are typically unaware of ontologies). Accordingly, we can say that the work presented in this paper is the first reported attempt to conduct an empirical study that evaluates ontology extraction tools taking into account both the requirements of e-learning practitioners and constraints imposed on ontologies by advanced e-learning systems. In this study, we specifically aimed at investigating how educators perceive the utility and usability of the ontology extraction tools as well as how they perceive the quality of the resulting ontologies. Our research questions and the overall study design are given in Section 3. In Section 4, we present the results of both quantitative and qualitative data analysis and discuss them in the context of our research questions. Section 5 concludes the paper with a summary of our findings and presents a set of recommendations to be followed when developing ontology extraction tools for the TEL domain.

## 2 LITERATURE REVIEW

Aiming to eliminate one of the major obstacles to wider acceptance and deployment of ontology-based software solutions, the Semantic Web community has devoted a significant attention to the development of tools that would facilitate and scaffold the ontology development task. The result of this effort is a wide range of tools<sup>1</sup> aimed for users with different levels of knowledge engineering expertise.

<sup>1</sup> Mike Bergman's ontology development tool list, <http://goo.gl/k6lt6>

Having explored and analyzed the available tools' support, we distinguished three general categories: 1) tools for handcrafting ontologies, 2) tools supporting reuse of existing ontologies and 3) tools for (semi)-automatic development of ontologies. In the following subsections, we overview these three categories of tools and explore the tools from the last category in more detail, as they are the focus of this paper.

## 2.1 Tools for Handcrafting Ontologies

Manual creation of ontologies using a specialized ontology editor (such as Protégé, NeON toolkit, or TopBraid-Composer) is still the predominant approach in ontology development. Its main drawback is the fact that the majority of ontology editors are suitable only for experts in the field of ontology development, since they assume a background in knowledge engineering and familiarity with ontology languages. One attempt at making the task of handcrafting ontology more convenient for non-technical users is the use of controlled language interfaces. In this approach, a tool provides structural elements, called templates, in a language close to natural language. The user fills in templates with domain specific elements. The obtained statements are then automatically translated into logic statements in the underlying ontology language. Prominent examples of tools that implement this approach include ROO (Rabbit to OWL Ontology authoring) [24] and AceView plug-in for Protege. However, a study by Dimitrova et al. [21] has shown that even with this kind of support domain experts without knowledge engineering skills had difficulty in producing good quality ontologies (with both tools the users created ontologies that matched just 50% of the axioms in the ontologies used as benchmarks).

Another attempt at facilitating the task of ontology construction consists of shifting the focus from 'heavy-weight' to 'light-weight' ontologies (often referred to as 'vocabularies') by relying on simpler underlying knowledge representation formalisms (e.g., supporting just RDFS Schema or a subset of OWL Light formalisms instead of OWL-DL). Such tools are often web-based and collaborative in nature, allowing ontology engineers and domain experts to work together on the ontology development tasks. Examples include Cupboard and Neologism. Hepp et al. [26] suggested Wikis' infrastructure and culture as an environment for constructing and maintaining consensual vocabularies for knowledge management. Though this seems to be an appealing solution from the perspective of knowledge engineers, it produces a collection of named conceptual entities with a natural language definition, and such an 'informal ontology' cannot address specific requirements of e-learning environments such as those discussed in the introduction.

## 2.2 Tools Supporting Reuse of Existing Ontologies

Created through contributions of the Semantic Web community members, ontology libraries such as Swoogle, BioPortal, and Watson offer a constantly increasing number of specific domain ontologies. However, supporting tools are needed to facilitate the searching process and evaluation of the retrieved ontologies [8]. In addition,

DBpedia, Freebase and other formally structured LOD knowledge bases could serve as domain ontologies. However, these knowledge bases are huge, covering a wide range of domains and topics and as such are not directly usable by educators. There is a need for tools that would allow for seamless extraction of segments of knowledge relevant for the educator's needs (i.e., covering the course he/she is teaching). In addition, tools should enable educators to adapt the existing knowledge models to fit their specific conceptualization of the courses and the topics they teach.

## 2.3 Tools Supporting (Semi-)Automatic Ontology Development

These tools aim at reducing the human intervention to supervision of the development process and refinement of the results[14],[25]. They offer users with suggestions for ontology elements, typically based on the analysis of existing domain documents. The two paradigms characterizing this category of tools are discussed in the following subsections.

### 2.3.1 Generate and Purge

The Generate and Purge Approach is exemplified by the Text2Onto tool [27] used in this study (Fig. 1). In this approach, a user first selects and loads a set of input documents (area 1). This results in a list of terms being extracted from the documents that may or may not be shown to the user (area 2). Next, one or more algorithm(s) is/are selected for extracting a particular ontological feature(s) (in area 3 of Fig. 1, algorithms for Concept and Relation algorithms are selected). After running the algorithms, the selected terms (or phrases) are presented to the user as suggestions, typically accompanied by some additional information such as the weight indicating the relevancy of the terms (area 4). The user is then asked to accept or reject these suggestions (area 5) to narrow them down to those to be included in the resulting ontology.

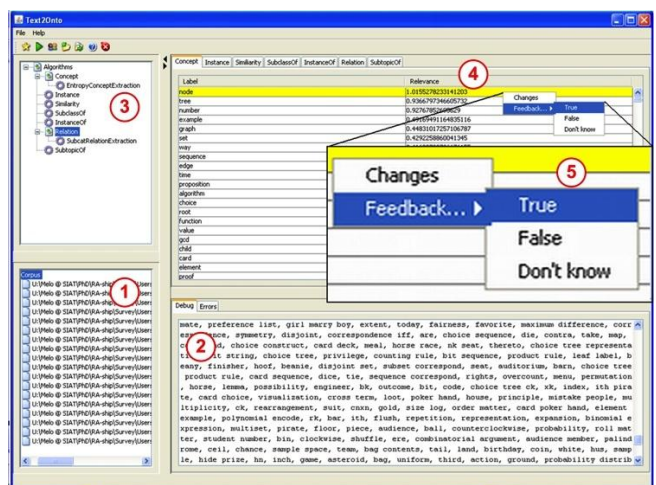


Fig.1 The Generate and Purge Approach as applied in the Text2Onto tool

Another well-known tool that belongs to this category is OntoLT [43] which was developed as a plug-in for the popular Protégé ontology development tool. It makes use of mapping rules to extract concepts and properties



from linguistically annotated text collections. However, OntoLT does not offer support for the linguistic annotation; for that, it depends on another proprietary tool. A number of mapping rules are included in the tool, and users can create new ones (provided that they master the precondition language that OntoLT uses for defining mapping rules). OntoBuilder [44] relies on the Generate & Purge approach to enable the extraction of ontology concepts and properties from Web pages. The extraction is based on the heuristic rules that the tool learns from a training set of HTML documents. After being given the URL of a Web page, OntoBuilder identifies the elements of the page, and then generates a dictionary of terms by extracting labels and fieldnames from the Web page; it also recognizes relations among the extracted terms.

### 2.3.2 Build Incrementally

In the Build Incrementally approach, a system helps users to build an ontology step by step by suggesting concepts and relations that can expand the selected node of the ontology based on the analysis of the underlying document corpus. OntoGen tool [28] (Fig. 2) implements this approach. The user starts by loading the documents into the tool (not shown in Fig. 2). The starting point for building an ontology is the root node. In the unsupervised approach, the user can specify how many concepts the tool should suggest (area 1) as sub-nodes of a selected node (area 2). The tool generates suggestions and for each one presents some information (e.g., keywords describing a suggested concept and the number of documents from the corpus that contain the respective concept) that should help the user to make a decision about whether to include the concept or not. In Fig. 2, four suggestions for the selected (highlighted) node have been accepted by the user and included in the ontology graph. In this way, starting from the root node, the user iteratively expands the nodes until he/she is satisfied with the ontology.

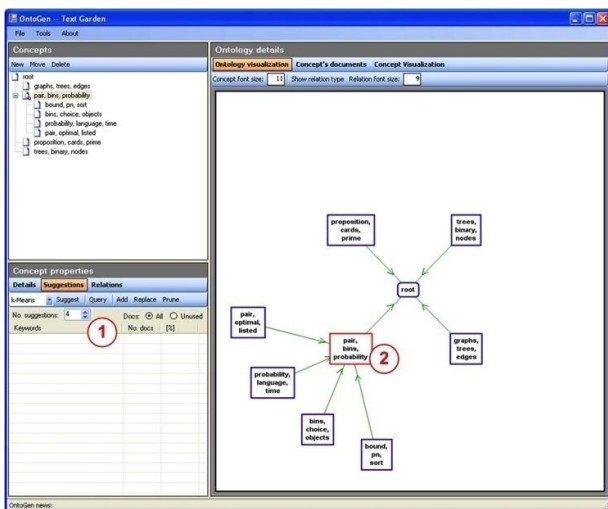


Fig.2 the Build Incrementally approach as applied in the OntoGen tool

DODDLE-OWL [45] is another example of an ontology extraction tool. Although the tool initially offers users a list of extracted concepts, similar to the Generate and

Purge approach, it actually relies on the Build Incrementally approach. After extracting a set of terms from the given domain specific corpus, the tool requires from the user to select relevant terms and identify their meaning by mapping each term to the corresponding concept in WordNet. Then, based on the user's input and by referring to the reference ontologies and documents, the tool generates an initial concept hierarchy and a set of concept pairs. The user then refines the initial ontology through the interactive support offered by the tool.

## 2.4 Methodologies for Ontology Development

To produce high quality result (i.e., ontology), an ontology development process has to be driven by a sound methodology. The Semantic Web research community has put a significant effort in the development of methodologies for ontology development; Corcho et al. [46] give a nice overview of the developed methodologies, whereas Simperl et al. [47] report on their actual usage in ontology development projects, based on an empirical study of a large number of both academic and industry projects. In the context of this paper, the most relevant is the methodology proposed in [48] since it targets the ontology development process based on the use of ontology extraction tools. The proposed methodology recognizes several phases in the ontology development process and for each phase gives a detailed description of input and output elements, activities to be performed, supporting tools to be used, and decisions to be made. It thus provides domain experts with detailed guidance for selecting and preparing information sources, applying ontology learning tools and evaluating the learned ontology.

## 3 METHOD

This section introduces the main components of the conducted study and the processes used in the preparation and execution of the study.

### 3.1 Research Questions

The study was driven by the following three main research questions (RQs):

*RQ1: How do educators perceive the utility of the existing ontology extraction tools?* To answer this rather broad RQ, we first identified the elements that contribute to the user's overall perception of the utility of an ontology extraction tool. Then, we investigated the perceived utility of the tools by exploring the identified elements:

- The users' perception of how intuitive and easy-to-follow the underlying ontology building approach is (incremental vs. generate-and-purge);
- the users' perceived value of the support the tool offers for active participation in and control over the ontology development process;
- the users' perceived value of the support the tool offers for comprehending the structure of the resulting ontology (this is related to the visualization of the resulting ontology);
- the users' perception of the tool's support for manipulating the resulting ontology, i.e., for applying the desired changes (e.g., by manipulating the visual rep-

- representation of the ontology);
- the users' perception of the guidance and feedback the tool offers in the ontology development process;
- the users' overall impression of the tool – to what extent it met their expectations.

RQ2: *How do educators perceive the usability of the existing ontology extraction tools?* We were primarily interested in the perceived ease of interacting with and operating the tool; also, the educators' ability to use the tool directly (i.e., without some form of training – ease of learning and use as per [29]); and the intuitiveness of the visualization(s) of the resulting ontology.

RQ3: *How do educators perceive the quality of the resulting ontologies?* We were particularly interested in this RQ, since the perceived quality of the resulting ontologies indirectly indicates the perceived usefulness of the tool for the end users. To answer this RQ, we first had to define the notion of quality of an ontology as well as the metrics to be used for assessing it. Unfortunately, the ontology quality research does not offer any empirically validated set of metrics, i.e., metrics empirically evaluated by following the well-established methods in software quality research. Zhan et al. [39] offer some metrics, but none of these metrics were empirically evaluated by following well-established methods. Accordingly, following the software quality research for similar type of software artifacts (i.e., UML class models and feature models) [40], [41], we considered the quality of an ontology in terms of its usability and maintainability, and also found that the best predictors of these quality dimensions are the number of features (i.e., concepts) and number of properties. By combining these metrics with the identified set of requirements for domain ontologies (see Section 1), we were able to define the characteristics of produced ontologies we wanted to elicit the educators' opinion about. These characteristics include:

- quantity of the concepts and relationships in the ontology;
- quality of concepts in the ontology (if the most important concepts were present);
- how effectively the ontology describes the domain.

For all three research questions we also wanted to know if there was a significant difference between the members of IT and nonIT groups (defined in Section 3.3.1 below) in the way they perceived the utility (RQ1) and usability (RQ2) of the tested tools, as well as the perceived quality of the resulting ontologies (RQ3).

## 3.2 Study Design

The study aimed at collecting and exploring the participants' perceived utility and usability of the existing ontology learning tools as well as their perceived quality of the resulting ontologies. To advance the existing research knowledge related to the ontology learning tools and their use by non-tech-savvy users in general and educators in particular (discussed in Sect. 2), we wanted to have a user study with a sufficiently large sample of educators and quantitative data. With such a sample size and data types, we were able to analyze the data in the way needed to address our research questions (Sect. 3.1).

## 3.3 Participants

### 3.3.1 Study Scoping

The survey aimed at two well-defined groups of teachers having low level of variability among their members. The first group (labelled as *IT* later) gathered teachers with Computer Science, Software Engineering, or Information Technology background. Hence, members of this group are representatives of those end users who are in general very familiar with complex software tools utilization and may have some notions related to ontologies. The other group (labelled *nonIT*) gathered non-Computer Science/Software Engineering teachers. Its members represent educators who are not aware of ontologies and knowledge representation and are less familiar with complex software tools. The rationale for having these two distinct groups of participants was to explore whether and to what extent their perception of ontology extraction tools and the quality of the produced ontologies differs.

During the study design phase, we conducted a simulation with the goal to estimate the sample size that, given the expected answers and variability of the population, can maximize the statistical power of the experiments. The target number of participants was set to 15 people for each group since it was a reasonable trade-off between statistical power (generally at least 80% for the expected outcomes) and the actual capability of recruiting a great number of participants. The simulations, performed using SAS software, assumed a latent normal distribution beneath the Likert-scale measurement and used Analysis of Variance (ANOVA) for comparing the tools.

### 3.3.2 Participant Selection

The participants were required to have at least a master's and preferably a PhD degree; plus at least three years of experience in teaching or course development. Also, they were required to have a substantial course material for the entire course they teach in electronic format: Word, PDF or PowerPoint documents or HTML pages.

The participants were recruited from the university faculty. Departments at authors' respective universities were targeted, as well as through other authors' connections, such as departments of our research partners from different post-secondary teaching and research institutions in Canada. The interested participants were screened by the research team for their background and the completeness level of their course material to guarantee the quantitative homogeneity of the basic input into the ontology building tools. The participant recruitment started in June 2008, though due to the summer season it was very low till September 2008, when we recruited most of the participants and started our experiments.

Total of 28 participants were recruited and retained for the study, 18 with the IT background, and 10 with non-IT background. However, due to some unforeseen time conflicts, one of the participants did not manage to finish the study, and thus did not provide us with fully completed questionnaire data. We did not consider the data of this user in our analysis. Even though we originally aimed at

a 50-50 split, some participants from the departments that indicated non-IT background were categorized as IT, due to their prior education in Computing (e.g., their MSc or PhD degree was in Education, but their BSc degree was in Computing).

### 3.4 Materials

The materials used in the study included: the course material provided by the participants, questionnaires and the ontology extraction tools selected for the study.

#### 3.4.1 Tool Selection

Information about the current state-of-the-art tools for the ontology learning was collected by exploring the literature (e.g., [30], [23], [14], [16]). However, out of almost a dozen tools mentioned in research papers, only four of them were publicly available on the Internet: OntoGen, Text2Onto and its predecessor TextToOnto, and OntoLT. Those were the tools that we managed to download and install. For the purpose of our study we decided to use two of them: Text2Onto and OntoGen. TextToOnto was discarded because it was superseded by Text2Onto. OntoLT was rejected as it depends on another proprietary tool which we did not have access to. Besides their availability, another reason for selecting Text2Onto and OntoGen was that they implement the two most widely researched approaches to ontology learning (Sect. 2.3).

**Text2Onto.** Text2Onto is an ontology-learning framework which supports automatic or semi-automatic generation of ontologies from textual documents. It combines machine learning with basic linguistic processing for learning atomic classes, class subsumption and object properties. The framework provides a Graphical User Interface allowing a user to define the document corpus from which the ontology will be created, then select the available algorithms to be applied for generating concepts/relations, and finally review the extracted concepts and relations. The tool is based on the “Generate and Purge” interaction paradigm (c.f. Section 2.3.1. In the publications related to this tool, we were not able to find any that discusses the evaluation of the tool.

**OntoGen.** OntoGen is oriented towards semi-automatic ontology construction. It is an interactive tool that aids a user during the ontology construction process by suggesting concepts, automatically assigning instances to concepts and providing visual representation of both the ontology and the corpus the ontology is built upon. To build an ontology, a user has to supply a set of documents that reflects the domain for which the ontology is to be built. The tool assists the user in every step of the concepts’ hierarchy development by suggesting sub-concepts of the currently selected concept. The interaction paradigm used by OntoGen is “Build Incrementally”, described in Section 2.3.2. The tool was evaluated [28] with two groups of students, one consisting of 48 Computer Science students and the other of 43 Psychology and Pedagogy students. After a brief introduction of the tool’s purpose and a demonstration of its functionalities, the students were asked to create ontologies using the corpus provided by the researchers. They also filled in questionnaires before and after using the tool (details of the ques-

tionnaires were not given). The authors report that the participants were mostly positive about the tool and its features. The tool is reported to be efficient, offering a lot of space for user intervention and handy visualizations. The identified drawbacks include: the lack of detailed instructions, the need to learn how to use it, and occasional slowness. However, since the design and results of the evaluation study are just briefly presented, it is not possible to derive any firm conclusions about the tool and its perceived usefulness.

#### 3.4.2 Questionnaires

We relied on two questionnaires (available at <http://www.sfu.ca/~mhatala/tlt2012/>). The first questionnaire was designed for collecting the participants’ perceived value of the tested tools and the created ontologies. It consisted of two groups of questions: 1) questions about the perceived utility and usability of the tested tools, and 2) questions about the perceived quality of the developed ontologies. This questionnaire was supposed to be filled in by the participants after they completed the ontology creation process with each tool. The questionnaire used 5-point Likert scale with values ranging from *Completely Disagree* to *Completely Agree*. The second questionnaire consisted of questions about participants’ experience in using the tested tools. More precisely, it included questions regarding the tools’ intuitiveness, ease of interaction, pros and cons, and participants’ expectation. All questions were open-ended with 200-word limit for the answers. This questionnaire was supposed to be filled in by the study participants after the ontology development task was done with both tools.

#### 3.4.3. Course Material Provided by the Participants

The selected tools require domain specific documents in plain text format as the input to the ontology extraction process. On the other hand, the great majority of course materials submitted by the users were in the form of (rich-) text documents, slide presentations, or HTML pages. To minimize the efforts of the study participants, we asked them to provide their course materials to the research team, and the team performed the required transformations of the participants’ materials into the format the tools could process. The same materials were used as input for both OntoGen and Text2Onto tools.

Table 1 gives the basic descriptive statistics about the documents we obtained from the participants. The data are grouped based on the type of the received documents (text documents, slides, HTML pages). The last row of the table contains the statistics about the overall number of words contained in the submitted material, once we converted them into text files to be used in the experiments.

### 3.5 Study Procedure

The participants submitted their course material to the research team upfront. The research team converted the materials into plain text format that was an accepted format for both tools. To avoid problems with tools installation on participants’ machines, a specific remote server was prepared with all the tools installed and the course material ready in the text format. The server was accessi-

ble via remote desktop and assistance was available via Skype communication software during the whole session. The assistance was provided by a research assistant with experience in using the tools selected for the study, along with a strong expertise to tackle all technical challenges that could have arose during the study (such as remote connection and software compatibility issues). A step-by-

**Table 1.** Descriptive statistics of course material: N – number of participants who supplied this type of material (N.B. some participants had mixed type of materials such as slides and HTML pages, while the overall number from whom we collect this data is 27 as reported in the last row); Min, Max, Mean, and SD –respectively minimal, maximal mean value, and standard deviation of numbers of the submitted units (i.e., files, pages, documents, words).

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>
Presentation slide files	10	6	36	14.6	8.6
Text Documents (PDF and MS Doc)	17	1	25	10.6	6.1
HTML pages	2	107	144	125.5	26.2
Number of words	27	4,345	293,555	59,372	78,000

During the session, the participants were asked to build an ontology describing their subject of expertise from the course material they had provided, and using the tools selected for the study. No training was provided for the tools, as we distributed the instructions on how to use the tools. There were no time constraints for the ontology creation process, with the majority of the sessions lasting between 2 and 3 hours.

After completing the task with the first tool (Text2Onto) the participants were asked to fill in the first questionnaire (the one consisting of questions about the tested tool and the developed ontology). Next, having used the second tool (OntoGen) for ontology development, the participants evaluated it as well (using the same questionnaire). Finally, they filled in the second questionnaire (the one about their experience in using the tools). The obtained results are discussed in Section 4.

### 3.6 Data Analysis

#### 3.6.1 Content Analysis

The study collected both quantitative and qualitative data. Qualitative data originated from the participants' answers to open-ended questions of the second questionnaire (Sec. 3.4.2).

Content analysis, generally used for systematically studying large amounts of communication content (e.g., articles, discussion forum messages, or blogs), was applied here for an interpretative inquiry. The process we followed was based on a combined method from [31] and [32]. Each answer was analyzed to identify the keywords associated with the participants' feedback and to understand its context.

The key characteristic of content analysis is that the analyzed content is categorized by researchers. Based on the participants' answers, we developed a coding scheme (see Sec. 4) for each of the open-ended questions. Three raters tested the scheme by applying it to five randomly selected questions and fine-tuned the coding manual. In the next step, the three raters applied the scheme independently to rate the answers. In the final step all the differences were resolved through a discussion in a meeting between the three raters. We should note that we wanted to report on the inter-rater reliability. However, Cohen's

step instructional package was made available to the participants a few days ahead of their scheduled session. It included: i) instructions on how to connect to the remote server and control the test environment; ii) information about and instructions on how to use the selected ontology building tools: Text2Onto and OntoGen.

kappa is applicable for two raters and only one category assigned per answer. Similarly, intra-class correlation coefficient is suitable for more than two raters, but again it requires one category assigned per answer only. We have not found a way to compute the measure for the situation with multiple raters and multiple categories per answer.

In the tables presenting the results of content analysis (Tables 3, 4, 6, 7 and 8), for each code of the developed coding schemes (one coding scheme for each question), the percentage of answers is given for the IT group, the nonIT group, and the total number of the code occurrence for each tested tool (more than one code could be assigned to a single answer). We tested for significance of the difference between the groups using chi-square statistics. If there is a significant difference in the code occurrence between the IT and nonIT groups, it is clearly indicated in the table and its caption, and we also address the issue explicitly in the discussion text.

#### 3.6.2 Methods and Test

To analyze the collected quantitative (Likert-scale) data we used standard descriptive statistics (as reported in [33] to be a common practice) including mean and standard deviation values. While there are two schools of thoughts on how Likert scales should be analyzed [34], there is a significant amount of empirical evidence that Likert scales can be used as interval data [35], [36]. Accordingly, having checked for the normality of the data distribution, we did the analysis of difference between the IT and nonIT groups by using parametric tests.

As already indicated, relations between the categorical data are tested by using the chi-square test.

## 4 RESULTS AND DISCUSSION

In this section we present the results of both quantitative and qualitative analysis of the collected data. The results are presented and discussed in the context of our research questions (Sec. 3.1).

### 4.1RQ1: The Perceived Utility of the Tools

In Table 2, we report on the results we got by analyzing the participants' responses to the questions related to the utility of the tested tools. As the table indicates, the find-



ings are predominantly negative. In what follows, we discuss these results as well as the results of the open-ended questions related to this RQ. The discussion is organized around the elements that have been identified as central to the educators' overall perception of the tools' utility (see RQ1 in Sec. 3.1).

**The perceived ability to actively participate in and have control over the ontology development process.** The participants' responses to question A.1 (Table 2) demonstrate that, for both tools, the participants felt that they would like to be engaged in the process of generating the ontologies. This was more pronounced in the case of Text2Onto where the user input is limited to removing concepts and relationships extracted from the supplied course materials. One participant complained on this limitation by saying: "I couldn't move concepts or elaborate some of my ideas further". Another user experienced a similar problem: "I wanted to bring it [deleted concept] back and say it was important but it had disappeared from the screen". OntoGen engages users more by letting them directly control which concepts in the ontology will be further expanded.

Question A.2 indicates that in the case of OntoGen, the participants felt that the tool is in control of the process. One participant echoed this negative aspect of the tool, saying: "I feel I have no control on OntoGen. It will generate sequences of keywords, which are not well-related. ...". In case of Text2Onto the participants were neutral. There was a statistically significant difference between the two tools, as shown in the last column (paired t-test) of Table 2. Importantly, in neither case the users felt that they were in control of the tool and the process, which negatively affects their attitude towards the tools.

**The perceived intuitiveness of the ontology building approach.** As indicated by question A.3 (Table 2), the participants were neutral with respect to how easy the process of obtaining the ontology was.

More significant feedback on this topic was obtained through the participants' responses to the open-ended question asking them to comment on the intuitiveness of the approach each tool is based upon. Table 3 summarizes the results of the content analysis based on the developed coding scheme. Four (out of seven) codes addressed the intuitiveness explicitly: INT - intuitive, LI - lack of intuitiveness, NVI - not very intuitive and GI - gradually intu-

itive. Although 25.9% of users found OntoGen and 37% found Text2Onto to be intuitive (col. INT), the scores for negative comments are even higher.

Text2Onto is much less intuitive with over 48% of participants describing it as lacking intuitiveness (col. LI) or not very intuitive (col. NVI). The following comments exemplify the overall impression on Text2Onto's intuitiveness: "Unclear since I didn't understand what the different algorithms would do"; "I had no idea what I was supposed to do or what the language it was using meant".

OntoGen was perceived as more or less unintuitive for 26% of the participants: 22.2% found it not very intuitive, and for 3.7% of users it fully lacked intuitiveness. One of its features which was particularly unintuitive for the users is the way it generates suggestions: "How 'suggestions' work needs to be better explained in the instructions"; "Not intuitive, esp for the part on how to generate suggestions."

For 14.8% of the participants OntoGen became gradually more intuitive with use (col. GI), as opposed to 7.4% for Text2Onto. Interestingly, both tools became gradually more intuitive for the larger number of nonIT participants. Some participants explicitly described one tool with respect to the other tool (Table 3, col. BOT), with 14.8% considering OntoGen a better tool than Text2Onto.

Overall, we consider the results for both tools to be rather negative. This indicates that the two ontology building approaches (Incremental Development and Generate-and-Purge), as implemented in the considered tools, are not suitable for use by educators without providing them with sufficient training (written guidelines as those provided in Sect. 3.5 are clearly not enough).

It might be interesting to mention that even though they were not explicitly asked about the visualization, when answering this question, the majority of participants commented on the visualization offered by the tools. In particular, 11.9% of the participants considered OntoGen providing a good visualization (Table 3, col. GV); whereas a small number of participants reported that a visualization was missing (Table 3, col. MV). This indicates that visualization is an important component of an ontology extraction tool as it obviously has an impact on the perceived intuitiveness of the overall ontology building approach.

**Table 2.** Text2Onto and OntoGen comparison based on participants' answers to the questionnaire. Both IT and nonIT participants are included. Values reported are from 5-value Likert scale (1-Completely Disagree, 2-Disagree, 3-Neither agree nor disagree, 4-Agree, 5-Completely Agree). (M, SD, N) values represent mean, standard deviation, and sample count. Paired t-test results are shown only for significant differences between Text2Onto and OntoGen.

Question	Text2Onto (M, SD, N)	OntoGen (M, SD, N)	Paired t-test
A.1 I prefer to participate in the process of creating an ontology	4.11, 1.01, 27	3.76, 1.05, 25	-
A.2 I felt in control of the process while obtaining the ontology	3.21, 1.06, 28	2.33, 1.03, 27	t(52.9)=3.10, p=0.003
A.3 I found the process of obtaining the ontology easy to accomplish	3.14, 1.07, 28	3.00, 1.51, 27	-
A.4 Visual representation of ontology helps the creation process.	4.17, 0.90, 28	3.44, 1.15, 27	t(49.2)=2.61, p=0.011
A.5 Being able to manipulate the generated ontology (e.g., add new elements, exclude those that I find unimportant, etc.) would improve my work.	4.25, 0.96, 28	3.88, 1.25, 27	-



A.6 It would have been good to have a sort of guidance during the creation process in order to know how the choices the tool provides would have affected the final result.	4.17, 0.98, 28	4.55, 0.84, 27	-
---	----------------	----------------	---

**Table 3.** Intuitiveness of the ontology building approach. Meaning of the codes: INT-intuitive, LI- lack of intuitiveness, NVI-not very intuitive, GI-gradual intuitiveness with the use, MV-missing visualization, GV-good visualization, BOT-better than the other tool.

		INT	LI	NVI	GI	GV	MV	BOT
OntoGen	IT	29.4%	-	23.5%	11.8%	5.9%	5.9%	11.8%
	nonIT	20%	10%	20%	20%	20%	-	20%
	Total	25.9%	3.7%	22.2%	14.8%	11.9%	3.7%	14.8%
Text2Onto	IT	35%	35%	17.6%	-	-	5.9%	-
	nonIT	40%	30%	10%	20%	-	-	10%
	Total	37%	33.3%	14.8%	7.4%	-	3.7%	3.7%

Finally, it is worth mentioning the participants' feedback on the ontology extraction approach that was indirectly collected through their responses to the open-ended questions related to the pros and cons of the tested tools. When commenting on the positive aspects of the tools (Table 6), over 55% of the participants stated that the biggest strength of Text2Onto is its ease of use (col. Ease) stemming primarily from its automatic generation of a large number of concepts and relationships ("I really like the output of the tool with respect to the very many concepts extracted from the course"). A large number of the nonIT participants (Table 6, col. Ease: 80%) valued this feature as positive vs. only half this number of the IT participants (41.2%). This indicates that for the nonIT users the generate-and-purge strategy for ontology building (used by Text2Onto) is more appealing than the incremental development approach (implemented by OntoGen).

**The perceived ability to comprehend the structure of the resulting ontology.** Related to the previous observation, results for question A.4 (Table 2) indicate that an ability to visualize the ontology is important. On average, participants expressed a significantly higher importance of visualization in case of Text2Onto than in case of OntoGen. This could have been expected since Text2Onto provides a long list of proposed concepts and relationships along with their weights in a tabular form, without any representation of the structure of the resulting ontology. The participants were "missing a visualization" to help them comprehend the ontology structure, or in the words of one user: "I have no clue what the ontology really looks like, sorry. Just looking at a list of relations didn't help me much".

**The perceived support for manipulating the resulting ontology.** In question A.5 (Table 2), the participants expressed that a support for manipulating the generated ontology is a desirable characteristic of any tool for this purpose. They felt more strongly about this in case of Text2Onto than in case of OntoGen, but the difference was not significant.

Explaining the difficulties he encountered in manipulating the result in Text2Onto, one participant wrote: "I found that I had to click so many times as I had such a wealth of concepts to consider. I am not sure how to change this but it was a great many clicks and almost had my wrist weary" (NB: this comment is from the only user who changed the ranking for all the concepts). The users were also struggling with OntoGen when trying to adapt the tool's output to their needs: "Difficult, because not flexible enough.

After suggestions at level 1, could not proceed down any more levels". Similar user perceptions were reported in [48].

**The perception of the guidance and feedback offered by the tool.** Having guidance during the process is perceived important (Question A.6, Table 2). Especially in case of OntoGen with a mean at 4.55, the participants were very uncertain how to proceed (e.g., one user wrote that it "could have been better if one could get new suggestions after adding new concepts and linking the appropriate files").

A number of participants mentioned the lack of feedback as a negative feature of both tools (Table 7, col. LF). Complaining on the lack of feedback in Text2Onto, one user wrote: "After click[ing] the run button, no feedback when it took a while to generate the result".

**Meeting expectations.** We were also interested to learn about the participants' overall impression of the tools' utility, i.e., whether the tools met their expectation or not. Table 4 summarizes the findings we got by analyzing the open-ended question asking the participants if and to what extent the tools met their expectations.

**Table 4.** Meeting Expectations. The codes have self-evident meaning. The values with a star show a statistically significant difference between IT and nonIT groups.

		Yes	No	Partly
OntoGen	IT	29.4%	35.3%	23.5%
	nonIT	30%	20%	30%
	Total	29.6%	29.6%	25.9%
Text2Onto	IT	-	*82.4%	5.9%
	nonIT	30%	*40%	20%
	Total	11.1%	66.7%	11.1%

For OntoGen, the opinions of the participants were split into three approximately same-sized groups, where OntoGen met expectations for 29.6% of the participants, met them partly for 25.9%, and did not meet for 29.6% of the participants. Those with unmet expectations were "expecting a more detailed ontology" and felt that the tool did not go well ahead of what they could have done without it ("It did provide a nice concept map with relations, but I'm not sure how this would add to what I could do myself"). They also found the suggestion of concepts rather poor ("The suggested first level and the sub-concept suggestion are helpful but not accurate at all."). In terms of meeting users' expectations, no major differences were noticed except a slightly higher proportion of dissatisfied IT participants than nonIT.

For Text2Onto the opinions were much more negative, especially from the IT participants. For 82.4% of the IT participants, Text2Onto did not meet their expectation as opposed to 40% of the nonIT participants. This difference was statistically significant with  $\chi^2(1, N=27) = 5.08$ ,  $p=0.024$ . The difference for answer “Yes” was also statistically significant where Text2Onto met expectations for 30% of the nonIT participants and for no IT participant ( $\chi^2(1, N=27)=5.74$ ,  $p=0.017$ ). The participants dissatisfaction came primarily from “too many concepts and relations [that]were learned” they felt “unable to really work with”. In fact, the following comments reflect the main reason the users felt that Text2Onto did not meet their expectations: “I was expecting to have a list of central concepts and relations but that didn't work” and “It reveals some relevant concepts and relations, but the concepts contain too many irrelevant

ones, while the relations revealed are true but kind of trivial”. These and similar comments indicate that unmet expectations were more related to the tools’ output (i.e., the produced ontologies) than to the interaction with the tools during the ontology construction process.

**Influence of Participant’s Background on Tool Evaluation.** We were also interested to know whether participants’ background has an influence on the perception of the tool’s utility. Therefore, we processed the responses to questions A1-A6 (Table 2) independently for each tool, while calculating and comparing the means of the two groups (IT and nonIT) using the paired t-test. Although none of the differences between the two groups were found to be significant, there are several cases where the participants’ background caused the shift in the response mean. Table 5 shows the comparison results.

**Table 5.** Comparison of tools’ evaluation based on the participants’ background. Answers were compared separately for Text2Onto and OntoGen. Paired t-test did not show any statistically significant differences between IT and nonIT groups. For question text and Likert scale values refer to Table 2.

Question	Text2Onto-IT (M, SD, N)	Text2Onto-nonIT (M, SD, N)	OntoGen-IT (M, SD, N)	OntoGen-nonIT (M, SD, N)
A.1	4.17, 1.07, 17	4.00, 0.94, 10	3.80, 1.14, 15	3.70, 0.94, 10
A.2	3.27, 1.07, 18	3.10, 1.10, 10	2.23, 0.97, 17	2.50, 1.17, 10
A.3	3.11, 1.07, 18	3.20, 1.13, 10	2.76, 1.34, 17	3.40, 1.77, 10
A.4	4.33, 0.76, 18	3.90, 1.10, 10	3.47, 1.00, 17	3.40, 1.42, 10
A.5	4.33, 0.97, 18	4.10, 0.99, 10	3.76, 1.39, 17	4.10, 0.99, 10
A.6	4.22, 1.00, 18	4.10, 0.99, 10	4.41, 1.00, 17	4.80, 0.42, 10

In the case of Text2Onto, the only difference worth commenting on is related to the question A.4 where the IT group felt a stronger need for the visual representation ( $M=4.33$ ) than the nonIT group ( $M=3.90$ ). This preference was not visible in the case of OntoGen.

The nonIT group found the process of obtaining ontology easier ( $M=3.40$ ) in case of OntoGen than the IT group ( $M=2.76$ ). However, both means are in the middle of the scale indicating neutral position on this question. The nonIT group also preferred more guidance ( $M=4.80$ ) for OntoGen than the IT group ( $M=4.41$ ). As we commented above, both values indicate a serious need for guidance during the process.

## 4.2 RQ2: the Perceived Usability of the Tools

To answer this research question, we made use of the participants’ answers to the open-ended questions. In particular, one of those questions directly addressed the usability aspect that was of primary importance for us: the ease of interacting with and manipulating the tools. We were also able to indirectly collect the participants’ feedback on other usability aspects from their answers to the questions related to the tools pros and cons.

**The ease of interacting with and manipulating the tool.** As can be seen in Table 8, the participants’ responses to this question were split in the middle for OntoGen: 40.7% considered it easy to use while 37% considered it not very easy to use. More nonIT participants considered OntoGen not very easy. Some of them felt that some common interaction elements were missing and that seriously affected the perceived usability of the tool (“Poor – couldn’t do standard stuff like select multiple items to move at

once, or use a drag-and-drop technique. Couldn’t use return to submit request. Couldn’t manually adjust diagram to see it better”).

In case of Text2Onto, 66.7% of all the participants considered it easy to use (“Fairly easy once you got over geek-speak”); this includes 90% of the nonIT participants and only 52.9% of the IT participants. This difference was statistically significant ( $\chi^2(1, N=27) = 3.89$ ,  $p=0.049$ ).

OntoGen’s visualization was described as hard to manipulate by 22% of the participants (“I like the visual approach, but ... almost impossible to edit”), against 7.4% in case of Text2Onto. A small number of the nonIT participants found both tools lacking feedback; while a few of the IT participants felt they had no control of the process (11.8% for Text2Onto).

Overall, the opinions on this question are split and the matter of usability of ontology building tools should be studied more carefully. However, an interesting pattern can be observed for the Text2Onto tool from the users’ responses to this question and the one on intuitiveness of the ontology building approach (Table 3). With its simpler interface that hides the structural aspects of the ontology, the nonIT group have found Text2Onto easy to use (90%; Table 8, col. Easy), although 40% reported that it was not intuitive (Table 3, col. INT) or it gradually became intuitive (20%; Table 3, col. GI).

**Additional insights on usability.** We were able to get additional insights on the tools usability by analyzing the content of responses to the open-ended questions about the tools’ pros (Table 6) and cons (Table 7). As Table 6 indicates, only 26% of the participants described ease of use (col. Ease) as a positive characteristic for OntoGen. One user found it particularly difficult to handle: “I would have

had a hard time if I had not read the manual before starting the experiment. Even then, I referred back to it several times and asked the Assistant for clarifications". However, the participants valued OntoGen's visualization aspects, with over 74% explicitly identifying visualization (col. Viz) as a positive characteristic ("It was nice to see the visual ontological representation arise as I worked"). While no user explicitly identified visualization as a feature of Text2Onto, 20% of

the nonIT participants were positive about the ranking of concepts it offers ("Easy and insightful to see the concepts which ranked the highest"). Obviously, visualization is a highly valued characteristic ("Visual organization of course content – helps in analyzing of what's there and what's not"; "Visualization of concepts is definitely nice to have") and should be provided in tools for ontology building to help boost its usability.

**Table 6.** Pros of the tools. The codes have the following meaning: Ease- Ease of use /automatic generation of ontology, Viz- Visualization, UC- User control, Nthng-nothing good, Rnk- ranking of concepts. The values with a star show statistically significant difference between the IT and nonIT groups.

		Ease	Viz	UC	Nthng	Rnk
Ontogen	IT	23.5%	70.6%	17.6%	-	-
	nonIT	30%	80%	30%	-	-
	Total	25.9%	74.1%	22.2%	-	-
Text2Onto	IT	*41.2%	-	11.8%	23.5%	-
	nonIT	*80%	-	-	10%	20%
	Total	55.6%	-	7.4%	18.5%	7.4%

**Table 7.** Cons of the tools. The codes have the following meaning: Miss- missing concepts and relationships, TME- too many generated concepts, nGUI- not friendly GUI, LF-lack of feedback, LC-lack of user control, Mac-Mac incompatibility, and Crash.

		Miss	TME	nGUI	LF	LC	Mac	Crash
OntoGen	IT	29.4%	5.9%	29.4%	11.8%	17.6%	11.8%	11.8%
	nonIT	30%	-	40%	30%	-	-	10%
	Total	29.6%	3.7%	33.3%	18.5%	11.1%	7.4%	11.1%
Text2Onto	IT	23.5%	29.4%	23.5%	-	5.9%	17.6%	-
	nonIT	30%	-	10%	20%	20%	-	20%
	Total	25.9%	18.5%	18.5%	7.4%	11.1%	11.1	7.4%

**Table 8.** The ease of interacting with and manipulating the tool. The codes have the following meaning: Easy- easy to use, NVE- not very easy to use, VP- hard to manipulate the visualization, LF- lack of feedback, NC- user has no control over the process. The values with a star show statistically significant difference between the IT and nonIT groups.

		Easy	NVE	VP	LF	NC
OntoGen	IT	41.2%	29.4%	29.4%	-	11.8%
	nonIT	40%	50%	10%	20%	-
	Total	40.7%	37%	22.2%	7.4%	7.4%
Text2Onto	IT	*52.9%	29.4%	-	-	5.9%
	nonIT	*90%	10%	20%	10%	-
	Total	66.7%	22.2%	7.4%	3.7%	3.7%

Among the negative aspects of the tools (Table 7), the participants mentioned the not-user-friendly GUI (col. nGUI): 33.3% for OntoGen and 18.5% for Text2Onto. Mac incompatibility (col. Mac) and frequent crashes (col. Crash) were also reported.

#### 4.3 RQ3: The Perceived Quality of the Ontology

Table 9 provides some basic descriptive statistics regarding the ontologies that the participants produced. These statistics clearly indicate that the two evaluated tools produced ontologies of different sizes and complexity. The participants' perceptions of the quality of the produced ontologies are given in Table 10.

As Table 10 indicates, all the participants' responses fall in the lower end of the scale. The participants perceived that OntoGen produces significantly worse ontologies than Text2Onto from the perspective of how effec-

tively it describes the domain (Question B.2). However, it outperformed Text2Onto in the appropriate number of concepts that describe the domain (Question B.4). Both differences were statistically significant. The participants were in general dissatisfied with the generated ontologies; the level of dissatisfaction was higher (but not statistically significant) for OntoGen. Moreover, whereas the quality of generated concepts was considered fair (Question B.5), the participants expressed a need for more relationships in the generated ontologies (Question B.3).

These findings are further supported by the responses to the open-ended questions about the tools' pros and cons. In particular, the users identified missing ontology elements as a negative aspect of both tools (Table 7, col. Miss): 29.6% of them considered this as a negative characteristic of OntoGen ("I expected a more detailed ontology"), and 25.9% had the same opinion about Text2Onto ("I like

the way that the tool works, but it does not seem to grasp the main concepts of the course”). Commenting on OntoGen, one

user stated that the “Basic idea is nice. If it generated a more complete graph I would be interested in experimenting with it.”

**Table 9.** Descriptive statistics of produced ontologies: *N* – number of participants; *Min*, *Max*, *Mean*, and *SD* – respectively minimal, maximal, mean value, and standard deviation of numbers of the concepts or relations (N.B. number of words is from Table 1)

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>
Number of words	25	4,345	293,555	59,372	78,000
Concepts (OntoGen)	22	3	26	8.6	5.9
Relations (OntoGen)	22	0	13	3.2	4.0
Concepts (Text2Onto)	22	228	4,602	1,494.0	1,015.0
Relations (Text2Onto)	19	0	155	53.4	44.4

**Table 10.** Comparison of the resulting ontologies built using Text2Onto and OntoGen. Both the IT and nonIT participants are included. Values reported are from a 5-value Likert scale: for question B.1 – B.3, the scale is 1-Completely Disagree, 2-Disagree, 3-Neither Agree nor Disagree, 4-Agree, 5-Completely Agree. For question B4 the scale is 1-Not enough even for a rough description of the domain, 2-Enough for a rough description of the domain, 3-Enough for a fair description of the domain, 4-Enough for a detailed description of the domain, 5-Too many even for a detailed description of the domain. For question B5 the scale was 1-Poor, 2-Fair, 3-Good, 4-Very Good, 5-Excellent. Paired t-test results are shown only for significant differences between Text2Onto and OntoGen.

<i>Question</i>	<i>Text2Onto</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )	<i>OntoGen</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )	<i>Paired</i> <i>t</i> -test
B.1 I am happy with the resulting ontology	2.96, 1.03, 28	2.44, 0.93, 27	-
B.2 The ontology describes effectively the domain it is built for	3.10, 1.06, 28	2.48, 0.93, 27	t(52.5)=2.31, p=0.024
B.3 I would like to have some additional relations in the generated ontology.	3.59, 0.97, 27	3.48, 0.89, 27	-
B.4 The number of concepts in the ontology are...	2.29, 0.99, 27	3.15, 1.46, 26	t(43.8)=-2.48, p=0.016
B.5 The quality of concepts in the ontology are (i.e. the most important concepts are included) ...	2.50, 0.92, 28	2.15, 1.00, 26	-

**Table 11.** Evaluation of produced ontologies based on participants’ background. Answers were compared separately for Text2Onto and OntoGen. Paired t-test did not show any statistically significant differences between the IT and nonIT groups. For the question text and Likert scale values refer to the caption of Table 10.

<i>Question</i>	<i>Text2Onto-IT</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )	<i>Text2Onto-nonIT</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )	<i>OntoGen-IT</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )	<i>OntoGen-nonIT</i> ( <i>M</i> , <i>SD</i> , <i>N</i> )
B.1	2.88, 1.02, 18	3.10, 1.10, 10	2.29, 0.84, 17	2.70, 1.05, 10
B.2	3.22, 0.94, 18	2.90, 1.28, 10	2.29, 0.91, 17	2.80, 0.91, 10
B.3	3.77, 1.00, 18	3.22, 0.83, 9	3.70, 0.77, 17	3.10, 0.99, 10
B.4	2.27, 1.07, 18	2.33, 0.86, 9	2.87, 1.45, 16	3.60, 1.42, 10
B.5	2.38, 0.97, 18	2.70, 0.82, 10	2.06, 0.85, 16	2.30, 1.25, 10

Too many elements generated (Table 7, col. TME) was also identified as a negative feature, mainly for Text2Onto (18.9% of the participants). Users commented that the tool generates “Too many concepts, most irrelevant” and that “It was almost as though one were reading a dictionary of random words”. Interestingly, too many elements generated were a concern of the IT group only. Another important observations regarding Text2Onto is the ambiguity of the extracted concepts (“Some of the terms used for concepts are difficult to understand without referring to other sources”).

These results indicate that ontologies produced by the current ontology extraction tools are not at the level of quality required for the deployment of semantic web technologies in e-learning. Yet, useful directions for future research on ontology extractions are obtained and summarized in the conclusions section.

**Influence of Participant’s Background on Tool Evaluation.** We also analyzed the results for the produced on-

tologies by both tools separately to find out whether the participants’ perceptions are influenced by their background. Although the average perceptions between the IT and nonIT groups differ in some cases, none of these differences proved statistically significant using the paired t-test (see Table 11).

For both Text2Onto and OntoGen, the IT group preferred to have more relationship identified in the ontology. In case of the ontology produced by OntoGen, the nonIT group considered the number of concepts generated to provide more detailed description than the IT group.

## 5 CONCLUSIONS AND RECOMMENDATIONS

This paper presented an empirical study of educators using two ontology extraction tools to build domain ontologies for their courses from their course materials. The two tools implement the two most widely explored and adopt-



ed approaches to ontology generation from documents, namely, Generate-and-Purge (Text2Onto) and Build Incrementally (OntoGen) (Sect. 2.3).

The study results show that the current state of the tools for developing domain ontologies by educators is unsatisfactory. Although some differences between the IT and nonIT groups become visible in the survey data, the results demonstrate that both groups were equally dissatisfied with both tools. However, several conclusions can be made and suggestions given with respect to the explored approaches and desirable features of the tools.

First, there is an appeal for the approach that generates a large number of suggestions for ontology concepts and relationships that are then 'weeded out' by the user. This approach, applied by the Text2Onto tool, was especially favored by the nonIT group. However, having examined the produced ontologies from the perspective of the requirements of advanced e-learning technologies, we found their utility rather minimal. This is due to the following reasons: 1) the quality of the extracted concepts is rather low, as they do not represent the domain well; 2) despite the previously stated fact, users tend to keep an extremely large number of concepts as the process of eliminating irrelevant concepts is rather tedious; 3) a very small number of relationships (among concepts) is extracted. Fortunately, some latest research in the area of filtering and ranking of concepts and relationships has shown some promising results. Graph based metrics (e.g., Betweenness Centrality) for concept and relationship filtering increase the precision [15] as compared to traditional ranking metrics (e.g., TF-IDF) used in tools like Text2Onto.

Second, in the case of both tools, the study participants felt that they would like to be more engaged in the process of generating ontologies. So, there is a clear need for tools that would enable users to manipulate the generated ontology. In addition, our study participants reported that they were often uncertain how to proceed with the ontology development process. This clearly indicates that tools should provide users with appropriate feedback and guidelines in the process of ontology development. The reliance on an appropriate ontology development methodology could facilitate the process, but as reported in [48], the availability of a methodology does help, but still, end users need assistance and guidance when using ontology extraction tools.

Third, there is a clear need for a good ontology visualization capability that can be easily manipulated by the users. The study has shown that visualization has a significant impact on the perceived intuitiveness of the overall ontology building approach. Therefore, it should be provided in any tool for ontology building to help boost its usability. Yet, visualization should allow for interacting with ontologies through operations such as editing, removing, or adding new ontology elements. Such an interactive visualization has already been appreciated by educators in ontology maintenance tasks [37]. The requirements for visualization and editing of ontologies generated using ontology extraction tools, point to the need for comprehensive ontology development environments that would offer all those functionalities to the end-users. This

kind of development environment already exists (e.g., NeON toolkit, <http://neon-toolkit.org>) and it would be important to explore the end-users', specifically educators' perceptions of and ability to deal with such an environment in an efficient and effective way.

Finally, an ontology building tool should have built-in evaluation metrics for the quality of the ontology being developed, so that it can provide users with immediate feedback on how good their ontology is and some guidance for improving it. In this study, we based the definition of ontology quality and the metrics for assessing it on the findings of software quality research for similar type of software artifacts. We believe that it was a well informed decision as it was based on a huge amount of research work done in the area of software quality metrics. To validate this approach, we intend to conduct at least one and more likely a family of experiments [42].

We believe that these findings will be beneficial not only for our future research but also for the future work of other researchers in the area. We intend to work on a research prototype and consequently, to study empirically the effects of the recommendations listed in this sections.

## Acknowledgment

We would like to thanks Prof. Thomas M. Loughin who helped us properly design the study. This study is funded in part by Athabasca University's Mission Critical Fund, Athabasca University's Associate Vice President Research's special project, and NSERC.

## REFERENCES

- [1] Devedzic, V.: *Semantic Web and Education*. Springer, Berlin Heidelberg New York, USA, 2006.
- [2] Dicheva, D., Mizoguchi, R. Greer, J. (Eds.): *Semantic Web Technologies for e-Learning*, IOS Press, The Netherlands, 2009.
- [3] Dietze, S., d'Aquin, M., Gasevic, D., Sicilia, M. (Eds.): *Proc. of Linked Learning 2011: the 1st Int'l WS on e-Learning Approaches for the Linked Data Age*, co-located with the 8th Extended Semantic Web Conference, Heraklion, Greece. (2011). Available at: <http://goo.gl/rGlyC/>
- [4] Allemang, D., & Hendler, J.: *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann, (2008)
- [5] Jeremic, Z., Jovanovic, J., Gasevic, D.: *Personal Learning Environments on Social Semantic Web*. *Semantic Web Journal* (2012). (accepted). Available at: <http://goo.gl/XU72M>
- [6] Naeve, A., Lytras, M.D., Nejdl, W., Balacheff, N., Hardin, J. (Eds.): *Special issue on Advances of the Semantic Web for e-learning: expanding learning frontiers*. *British J. of Edu. Technology*, Vol. 37, No. 3 (2006)
- [7] Dolog, P., Nejdl, W.: *Semantic Web Technologies for the Adaptive Web*. *The Adaptive Web*, pp. 697-719 (2007)
- [8] Gasevic, D., Jovanovic, J., Devedzic, V.: *Ontology-based Annotation of Learning Object Content*. *Interactive Learning Environments*, Vol. 15, No. 1, pp. 1-26 (2007)
- [9] Jovanovic, J., Gasevic, D., Devedzic, V., "TANGRAM for Personalized Learning Using Semantic Web Technologies," *Journal of Emerging Technologies in Web Intelligence*, Vol. 1, No. 1, 2009, pp. 6-21.
- [10] Jovanovic, J., Gasevic, D., Brooks, C., Devedzic, V., Hatala, M., et al.: *Using Semantic Web Technologies for the Analysis of Learning Content*. *IEEE Internet Computing*, Vol. 11, No. 5, pp. 45-53 (2007)
- [11] Jovanovic, J., Knight, C., Gasevic, D., Richards, G.: *Ontologies for Effective Use of Context in e-Learning Settings*. *Educational Technology & Society*, Vol. 10, No. 3, pp. 47-59 (2007)

- [12] Ali, L., Hatala, M. Gasevic, D., Jovanovic, J., "A Qualitative Evaluation of Evolution of a Learning Analytics Tool," *Computers & Education*, Vol.58, No.1, pp. 470-489, 2012.
- [13] Jeremic, Z., Jovanovic, J., Gasevic, D.: An Environment for Project-based Collaborative Learning of Software Design Patterns. *International Journal on Engineering Education*, Vol.27, No.1, pp.41-51 (2011).
- [14] Cimiano, P., Völker, J., Studer, R.: Ontologies on Demand? - A Description of the State-of-the-Art, Applications, Challenges and Trends for Ontology Learning from Text Information. *Wissenschaft und Praxis*, Vol. 57, No. 6-7, pp. 315-320 (2006)
- [15] Zouaq, A., Gasevic, D., and Hatala, M.: Towards open ontology learning and filtering. *Information Systems*, Vol. 36, No.7 (2011), pp. 1064-1081.
- [16] Park, J., Cho, W., & Rho, S.: Evaluating ontology extraction tools using a comprehensive evaluation framework. *Data & Knowledge Engineering*, Vol. 69, pp. 1043-1061 (2010).
- [17] Siorpaes, K., & Simperl, E.: Human Intelligence in the Process of Semantic Content Creation. *World Wide Web*, Vol. 13, No. 1-2, pp. 33-59 (2010).
- [18] Stojanovic, Lj., Staab, S., Studer, R.: eLearning in the Semantic Web. In: *Proc. World Conf. on the WWW and the Internet*, pp. 325-334 (2003)
- [19] Mohan, P., Brooks, C.: Engineering a Future for Web-based Learning Objects. In: *Proc. of Int'l Conf. Web Engineering*, pp. 120-123 (2003)
- [20] Brase, J., Nejd, W.: Ontologies and Metadata for eLearning. In: S. Staab & R. Studer (Eds.) *Handbook on Ontologies*, pp. 555-574, Springer (2004)
- [21] Dimitrova, V., Denaux, R., Hart, G., Dolbear, C., Holt, I., & Cohn, A.: Involving Domain Experts in Authoring OWL Ontologies. In *Proceedings of the 7th Int'l Semantic Web Conference*, pp. 1-16 (2008).
- [22] Zouaq, A. & Nkambou, R.: Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 11, pp. 1559-1572 (2009).
- [23] Gomez- Perez, A., Manzano-Macho, D.: An overview of methods and tools for ontology learning from texts. *The Knowledge Engineering Review*, Vol. No. 19, pp.187-212 (2004)
- [24] Denaux, R., Dolbear, C., Hart, G., Dimitrova, V., & Cohn, A. G.: Supporting domain experts to construct conceptual ontologies: A holistic approach. *J of Web Semantics*, Vol. 9, No. 2, pp. 113-127 (2011).
- [25] Zouaq, A. & Nkambou, R.: A Survey of Domain Ontology Engineering: Methods and Tools, In Nkambou, Bourdeau & Mizoguchi (Eds): *Advances in Intelligent Tutoring Systems*, Springer, pp. 103-119 (2010).
- [26] Hepp, M., Siorpaes, K., & Bachlechner, D., "Harvesting Wiki Consensus: Using Wikipedia Entries as Vocabulary for Knowledge Management," *IEEE Int. Comp.*, 11(5), (2007), 54-65.
- [27] Cimiano, P., Voelker, J.: Text2onto - a framework for ontology learning and data-driven change discovery. In: *Proc. 10th Int'l Conf. on Applications of Natural Language to Information Systems* pp. 227-238. (2005)
- [28] Fortuna, B., Grobelnik, M., Mladenic, D.: *OntoGen: Semi-automatic Ontology Editor*. HCI International 2007, July 2007, Beijing
- [29] Mayhew, D. J. *The usability engineering lifecycle: A practitioner's handbook for user interface design*. San Francisco, CA: Morgan Kaufmann Pub. (1999).
- [30] Shamsfard, M., Barforoush, A.A.: The state of the art in ontology learning: a framework for comparison. *The Knowledge Engineering Review*, Vol. 18, No.4, pp. 293-316 (2003)
- [31] Renner, M. & Taylor-Powell, E., T.: Analyzing qualitative data. University of Winsconsin: Programme development and evaluation. (2003). Retrieved 01 March 2010 from <http://goo.gl/OoVjU>
- [32] Graneheim U. H. & Lundman B.: Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today* 24, 105-112 (2004).
- [33] Blaikie N.: *Analysing Quantitative Data*. London: Sage Publications. (2003)
- [34] Carifio, J. & Perla, R.: Resolving the 50-year debate around using and misusing Likert scales. *Medical Edu.*, Vol. 42, No. 12, pp. 1150-1152. (2008).
- [35] Carifio J. & Perla R.: Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of the Social Sciences*, Vol. 3, No. 3, pp.106-116. (2007)
- [36] Carifio, J.: Measuring vocational preferences: ranking versus categorical rating procedures. *Career Education Quar.*, Vol. 3, No. 1, pp. 34-66. (1978)
- [37] Gasevic, D., Zouaq, A., Torniai, C., Jovanovic, J., Hatala, M., "An Approach to Folksonomy-based Ontology Maintenance for Learning Environments," *IEEE Trans. Learning Techn.*, Vol. 4, No. 4, pp. 301-314 (2011)
- [38] Siadaty, M., Jovanovic, J., Pata, K., Holocher-Ertl, T., Gasevic, D., Milikic, N., "A Semantic Web-enabled Tool for Self-Regulated Learning in the Workplace," In *Proc. 11th IEEE Int'l Conf. Advanced Learning Technologies*, 2011, pp. 66-70.
- [39] Zhang, H., Li, Y. F., & Kuan Tan, H.B.: Measuring design complexity of semantic web ontologies. *J. Syst. and Softw.*, Vol.83, No.5, 2010, 803-814.
- [40] Manso, M. E., Genero, M., & Piattini, M.: No-redundant metrics for UML class diagram structural complexity. *Proc. 15th international conference on Advanced information systems engineering* (2003) 127-142.
- [41] Bagheri, E., & Gasevic, D.: Assessing the maintainability of software product line feature models using structural metrics. *Software Quality Control*, Vol. 19, No.3 (September 2011), 579-612
- [42] Basili, V., Shull, F., Lanubile, F.: Building Knowledge through Families of Experiments. *IEEE Trans. Software Eng.*, Vol. 25, No. 4, (1999), 435-437.
- [43] Buitelaar, P., Olejnik, D., Sintek, M.: OntoLT: A protégé plug-in for ontology extraction from text. *Proc. Int'l Semantic Web Conf.*, pp.31-44. (2003)
- [44] Gal, A., & Modica, G.: OntoBuilder: fully automatic extraction and consolidation of ontologies from Web sources, *Proceedings of the 20th International Conference on Data Engineering*, 2004.
- [45] Morita, T., Fukuta, N., Izumi, N., & Yamaguchi, T.: DODDLE-OWL: a domain ontology construction tool with OWL, *Lecture Notes in Computer Science* 4185 (2006) 537-551.
- [46] Corcho, O., Fernandez, M., and Gomez-Perez, A.: Methodologies, Tools and Languages for Building Ontologies: Where is the Meeting Point? *Data & Knowledge Engineering* Vol. 46, pp. 41-64 (2003)
- [47] Simperl, E., Mochol, M., & Burger, T.: Achieving Maturity: the State of Practice in Ontology Engineering in 2009. *International Journal of Computer Science and Applications*, Vol. 7, No. 1, pp. 45 - 65, (2010).
- [48] Simperl, E., Tempich, C., & Vrandečić, D.: A Methodology for Ontology Learning. In *Proc.2008 Conf. Ontology Learning and Population: Bridging the Gap between Text and Knowledge* (2008), 225-249.

**Marek Hatala** is an Associate Professor and a Graduate Chair at the School Interactive Arts and Technology at Simon Fraser University. His research interests are in the areas of knowledge management, artificial intelligence, distributed systems, user modeling, interoperability, security and trust policies, e-learning and collaborative systems. More details about Marek can be found at <http://www.sfu.ca/~mhatala>

**Dragan Gašević** is a Canada Research Chair in Semantic Technology and an Associate Professor at Athabasca University. His research interests are in semantic technologies, software language engineering, technology-enhanced learning, learning analytics, and services computing. He can be reached at <http://dgasevic.athabascau.ca>

**Melody Siadaty** is a PhD Candidate at the School of Interactive Arts and Technology at Simon Fraser University, and a Research Assistant at Athabasca University. Her research interests include semantic web technologies, social web (web 2.0), technology-enhanced learning, workplace/organizational learning, learning analytics and learning theories. She can be reached at <http://www.sfu.ca/~msa70>

**Jelena Jovanovic** is an Assistant Professor of Computer Science with the Department of Software Engineering, FOS - School of Business Administration, University of Belgrade, Serbia. Her research interests are in the areas of semantic technologies, Web technologies, technology-enhanced learning, and knowledge management. She can be reached at <http://jelena.jovanovic.net>

**Carlo Torniai** is currently an Assistant Professor at Oregon Health & Science University in Portland, Oregon. His research interests include best practices of ontology development for multimedia and biomedical resource annotations and exploring interactions between Social and Semantic Web in learning environments.