

# Ontology based Qualitative Case Studies for Sustainability Research\*

**Dong Han, Kilian Stoffel**  
University of Neuchâtel  
Neuchâtel, Switzerland  
{dong.han, kilian.stoffel}@unine.ch

## Abstract

The overall objective of this paper is to describe the development of methodologies and to present an implementation for supporting researchers in the domains of sustainability. The methodologies are inspired by the approach of qualitative case studies as it is gradually gaining a large acceptance in these domains. In this paper, key elements of qualitative case study research are analysed and the state of the art relevant to these elements is presented. Ontologies, as the core methodology, are proposed as the basic tool to set up and represent knowledge from different perspectives. Original textual data is annotated with ontologies and information is retrieved from these ontologies by different ways for learning, inferring and proposing. Experiments and applications of sustainability scenarios are presented and an outlook to some further design and development possibilities are given.

## 1 Introduction

These years, qualitative case study has become an important methodology for scientific research and industrial applications in many fields, such as management, economics and environment. Particularly, in the domains of sustainability, it is getting an increasing attention due to its inherent advantages in many aspects. In this paper, we are going to present an ontology based qualitative case study approach for the domains of sustainability including energy and environmental management. This methodology as well as its implementation aim to extract information from original data and analyse it in multiple ways to provide feedback to domain experts. It was motivated by the need in the domains of sustainability but can be extended to other similar domains. The data we are using in this paper is collected in an ongoing interdisciplinary research project and consists of environmental data. The data stems from documents produced over several years in the context of the *Global Report Initiative*[GRI, 2011] and

---

\*This work is supported by Swiss National Science Foundation(SNSF) project "Formal Modelling of Qualitative Case Studies: An Application in Environmental Management" (Project No. CR21I2\_132089/1).

is analysed from a practical point of view by sustainability experts based on the approaches originated in computer science proposed in this paper.

In accordance with its characteristics, the case study approach provides many advantages which make it very attractive particularly in domains like sustainability management. First, the subject topic can be highly complex while lacking of corresponding theories as for example the study of new service design and environmental issues[Stuart *et al.*, 2002]. Furthermore, our research objective is mainly focusing on the question why the phenomenon emerge which is also well covered by the case study approach[Yin, 2009]. All these characteristics are arguments for using the qualitative case study approach in the domain of sustainability management.

Environmental issues have drawn increasing global attention. Not only international organisations which focus specially on the environment, but IT and financial companies with business closely related to environment, have started focusing on environmental issue as a decisive factor for their plans and operations. One of the usual problems they are confronted with is the large number of documents and reports containing information about environmental issues. Besides, since a high proportion of the original data is presented in a textual form, quantitative analysis is not sufficient. There is a clear need for an effective approach for supporting the analysis of their qualitative information. Qualitative case study, hence, is one of the appropriate methodologies for supporting sustainability management.

Because of their undeniable advantages for knowledge representation, ontologies are adopted in this paper as the core formalism for storing and managing information and knowledge. They allow for representing knowledge from nearly all domains and provide support for learning and reasoning processes. In this paper, we use ontologies for formally modelling the knowledge from the original data. Additionally, we use ontologies to formalise the grounded theory approach — one of the most important methodologies in the qualitative research domain. It enables us to incorporate into our system external knowledge from domain experts as well as from database.

The contribution of this paper focuses on providing a generic methodology and the description of a prototype for qualitative case studies in the domain of sustainability research. As stated in the previous paragraphs, the most sig-

nificant advance in our work is to allow for analysing very complex data, with a rich and diverse vocabulary. These characteristics caused the lack of standardised approaches readily available. Given this concrete condition, our objective is to successfully extract information and knowledge from the data, to set up a formal model to define a systematic approach, and to provide meaningful results to domain experts. Consequently, we highlight the advantages of using ontologies during our research process while selecting it as the basic tool for knowledge representation. We are thus able to investigate several integrated techniques to provide generic methodologies for the sustainability domains. Our implemented system has been used to contribute to the understanding of environmental management and its impact on financial service firms.

The paper is organised as follows: the second section introduces the state of the art; the third section presents the research methodology; the fourth and fifth sections explicate the system implementation, some experiments and use cases, while the final part is presenting the future work and conclusion.

## 2 State of the Art

In this section, we are going to present a detailed review on the existing scientific literature related to the main ideas, namely qualitative research, and the concept of ontology as well as their combined utilisation.

### 2.1 Qualitative Research

Qualitative research [Taylor and Bogdan, 1998] is an academic field dealing with the acquisition and analysis of non-standardised data. Data used in qualitative research can be provided using a wide range approaches, such as historical and phenomenological research. In the context of this paper we will essentially concentrate on the grounded theory approach [Glaser, 1992; Strauss, 1987]. In grounded theory the starting point is the data, which is marked by a series of *codes* which are then grouped together into *concepts* in order to reduce the dimensionality. These concepts are again grouped together to form *categories* which are finally used to build a *theory*. This contrasts the standard scientific method where the starting point consists of a hypothesis which is then tested against specially designed experiments. Based on grounded theory we will develop a formal framework for case study research in the domains of sustainability.

### 2.2 Ontologies in Computer Science

According to [Buitelaar *et al.*, 2005], in computer science an ontology is an explicit, formal specification of a shared conceptualisation in a domain of interest, where *formal* implies that the ontology should be machine-readable and *shared* so that it is accepted by a group or community. Mostly, they build upon a hierarchical backbone and can be separated into two levels: *upper level ontologies* (which describe the most general entities, contain very generic specifications and serve as a foundation for specialisation) and *domain ontologies* (which aim at describing a subject domain and where entities and relations of a specific domain are sometimes expressed directly in the texts belonging to it).

## 2.3 Ontology Learning

Ontology learning is concerned with knowledge discovery in different data sources and with its representation through an ontological structure and, together with ontology population, constitutes an approach for automating the knowledge acquisition process [Buitelaar *et al.*, 2005]. According to [Maedche and Staab, 2001], there are two important aspects in ontology learning: *availability of prior knowledge* — if knowledge exists, it must be used to create a first version of the ontology, which is then extended (semi-)automatically through learning procedures — and the *type of input* — according to its nature, input data is classified as structured data, semi-structured data (dictionaries, user tags [Benz and Hotho, 2007]) and unstructured data. The nature of data of interest for this paper is intrinsically non-structured and consists of natural language texts in the form of PDF documents.

## 3 Research Methodology

Based on our analysis and state of the art given above, the detailed research methodology is presented in this section.

### 3.1 Ontologies as Knowledge Representation Tool

Grounded Theory (GT) is probably the most systematic methodology currently used in qualitative research. In this section we will describe how we intend to integrate ontologies into this methodology and what the improvements are given this integration, especially from a formal point of view.

The basic idea of GT is to generate a theory starting from data. In our ongoing project, as the way of how we coordinate the research collaboration, GT is essentially regarded as the bridge between domain experts and computer scientists in that it reveals the principles of both specialised sustainability management and knowledge representation and discovery. The two sides supplement each other by fulfilling the implementation of GT from different aspects. The cooperative experts set up their theories by leveraging our methodologies and prototype whilst we take the reference from their fields to establish the domain knowledge. This combination works effectively due to our conformity with the substance of GT and its implementation in the form of ontologies.

#### Formal Upper Level Ontology

In the first phase of this paper we have developed a formal specification of the GT process.<sup>1</sup> Ontologies are well suited for describing processes (see e.g. [Hepp and Roman, 2007]). The process we have to describe here is very simple compared to a complete business process. However the big challenge and the goal of this ontology would be to formally define the process that describes the three coding schemes used in GT, namely: *open coding*, *selective coding* and *theoretical coding*. Open coding refers to a process where the user is free to select whatever codes he judges relevant to the task under consideration. Selective coding limits the usable codes to a

<sup>1</sup>A preliminary analysis has shown that formal knowledge representation aspects can be done using ontologies. We have decided to encode all ontologies in the context of this paper in OWL. Depending on the usage of the ontologies we can use the *lite*, *DL* and *full* dialect.

subset of previously (in open coding) identified codes. Theoretical coding is a top down approach where codes have to be selected in accordance with a previously established theory.

### Lower Level Ontologies

In this paper we are developing the formal framework that is needed to specify the lower level ontologies which will allow the implementation of a concrete case study research project in the sustainability domains. The set of lower level ontologies is composed of three types of ontologies, namely project, reference and coding ontologies.

The first set of ontologies (project ontologies) defines which GT approach will be used as well as the set of reasoning mechanisms that are allowed. Following this schema, all qualitative case studies will have the same structure even if they are following different views in respect to the GT approach. This will guaranty a much higher comparability between different projects even if they would have all their own structure. This set of lower level ontologies will define a concrete instance of the upper level ontologies which are only defining the formal constraints.

The second set of ontologies (reference ontologies) contains ontologies that are used as reference systems. Domain experts may want to impose a specific view on a given subject and impose a codification based on this reference system. As a minimum an environmental ontology is included as we are focusing on environmental management. This would lead to a *theoretical coding* approach.

The third set of ontologies (coding ontologies) is created by the users while coding the raw data. In our approach the results of this coding process will also be stored in ontologies, as a simple fragment of the ontologies shown in Figure 1. This allows for a direct application of the reasoning mechanisms implemented in the higher level ontologies. Coding ontologies are direct successors of project ontologies as well as reference ontologies and they are thereby inheriting their properties. In this way we guaranty that coding is done accordingly to the constraints imposed by the project and reference ontologies which in their turn have to conform to the upper level ontology.

### Ontologies of Sustainability

As far as sustainability is concerned, inheriting from the previous part, a number of ontologies are incorporated as references specifying the domain knowledge. Taking the example of environmental performance indicators, a group of indices to evaluate the performance issues related to environment, we have established ontologies including energy, emissions, materials and transportation. Each indicator, for example, energy, comprises a list of subcategories, such as gas, water, electricity, etc. The knowledge represented in the form of an ontology is readable for both domain experts and machines. It identifies the most important concepts, in the form of a taxonomy, and a hierarchical structure of the sustainability domains. It is contributing to the interface operations, learning, inferring and proposing. Ontologies of the sustainability domains participate seamlessly in all the phases of a project and demonstrate the advantages of our methodologies for a concrete case.

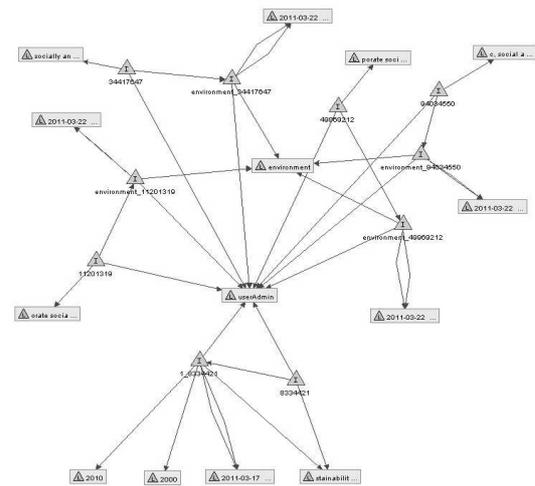


Figure 1: Ontology fragment

With the ontologies of sustainability, the process of case studies is facilitated and amended for two aspects: on the one hand, domain experts are able to employ the knowledge from these ontologies and together with the system we have implemented to do analysis for their case studies in depth. More patterns of the data, which are not intuitive for humans, can be revealed and presented. On the other hand, as their investigation to the cases gets more advanced, the experts are going to perceive more elements from different cases and these elements will be accordingly converted to ontology based knowledge to be navigated by the system.

### 3.2 Annotating with Ontologies

Based on the formal specification acquired as described previously we have developed a prototype implementation with a solid foundation completely based on an ontological approach. This prototype has been used to implement most of advanced functionalities outlined before.

As benchmark for the prototype we chose to focus on systems currently used by our project partners to analyse "Green Services". Derived from the idea of grounded theory, a number of commercial software products have been released for qualitative applications. Among them ATLAS.ti [ATLAS.ti, 2011] is the most prominent one. It emphasises elementary concepts and their implicit relationship.

ATLAS.ti has been designed on the basis of a *group of concepts*. Beyond all, primary documents are the original data source explored though the domain knowledge. A marked section or piece of text, named quotation, represents the significant or valuable parts of primary documents. These quotations can be exploited by adding codes with initiative names selected by the users or with the words in the text. Another concept, called memo, can be used to add comments about concepts using annotations.

In our prototype implementation, an operational graphical interface providing the functionalities for coding and exporting of information exists together with the mechanisms for

learning and inferring. The whole process follows the design of ontologies on different levels in the way described above as a consistent integration by linking knowledge representation and functional operations.

### 3.3 Ontology Learning

Ontology learning deals with discovering and acquiring new ontology elements, and integrating them in an existing ontology in a consistent and coherent way. In this paper, ontology learning is the basic process for constructing the coding ontologies. This important functionality is implemented based on the following assumptions:

- Input data for the learning process is represented by unstructured, natural language documents.
- The ontology concepts are represented by latent topics.
- Each document is annotated by the users with a set of labelling concepts (or codes), which may either be provided from the process/reference domain ontologies or from a predefined user language (in the form of a "bag-of-concepts").

As Figure 2 illustrates, the learning process starts from the annotation phase, followed by the term extraction phase. The matrices created using labels, documents and terms become the input for the label-topic model, and the tags (or codes) are analysed to estimate possible relations and to construct the hierarchical structure.

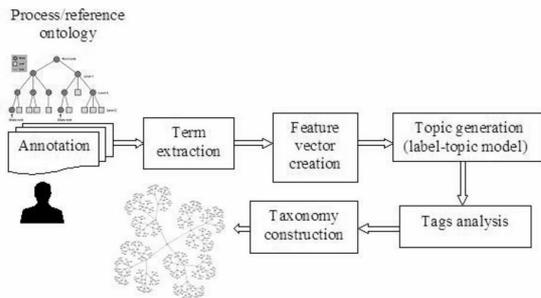


Figure 2: Learning process

More specifically, the stages of the ontology learning process are proposed as follows:

1. *Annotation* - The marked sections of the primary documents are coded (labelled) by the users, according to the constraints imposed by the project and reference ontologies. More than one label can be used to tag a piece of text (considered as an annotated document).
2. *Term extraction* - Each document of the corpus of annotated documents is treated as a bag-of-words. Stop-words are detected using simple, language independent statistical techniques and are then removed.
3. *Feature vector creation* - For each document  $d$ , the vector  $(w_d, l_d)$  is created. Two matrices, one recording the frequency of each term in each document, and the other recording the frequency of each label in each document, are calculated and used as input for the label-topic model.

4. *Topic generation* - The number of topics ( $K$ ) is fixed (but approaches using a variable number of topics may also be considered) and a generative probabilistic model is applied to model the correspondence between the documents and the labels.
5. *Tags analysis* - The possible relations between labels, seen as distributions over topics, are estimated using different divergence measures. The considered relations are Subordinate (or hypernym-hyponym relation), Merge (or synonymous relation) and Keep. The divergence measures are calculated based on the KL-divergence or on the conditional independence.
6. *Taxonomy construction* - The hierarchical structure of the tags (codes) is constructed by applying different operations, such as merging two similar tags or creating the hypernym relation between two tags. In this hierarchy, the result of a minimising or a keep operation is always a virtual concept (tag). The resulting taxonomy must be created by minimising the information loss.

### 3.4 Proposition System

The coding process, whatever schema is used, requires the decision of a user regarding the "best" code for annotating a given document. One of the most important advantages of the label-topic model is its capacity to predict the topic distribution of a new, unannotated document, and — based on this distribution — to find the most likely labels adapted for this document. Concretely, after the model is trained, we can infer information from the existing knowledge in the form of ontologies. The procedure of fold-in is fast, since it only deals with a single document instead of the whole corpus. In this context, the proposition system can be used to estimate the quality of the coding process for a given user, by calculating the "distance" between the codes selected by the user and the "best" ones proposed by the system.

## 4 System Implementation

Based on the methodologies proposed in the previous section, we have implemented a prototype as our application platform enforcing a systemic utilisation of our approach by the sustainability practitioners.

### 4.1 Architecture

As shown in Figure 3, the system, named *Qualogier*, is composed of several components - a PDF tool, an ontology manager, an inference engine, a proposition system and an exporting tool.

### 4.2 PDF Tool

The Portable Document Format(PDF) is an open standard widely used and most of the other data formats, such as Word and web pages, can easily be converted to PDF. This is the reason why we selected PDF as the default format for the original documents. Consequently, assigning annotation and all other essential functionalities for supporting qualitative case study research have to be added to basic PDF documents. We use ICEPdf[ICEpdf, 2011], a Java-based project, as the tool for PDF operations and extend it with extra functionalities. In ICEPdf, basic PDF operations are provided. Users can

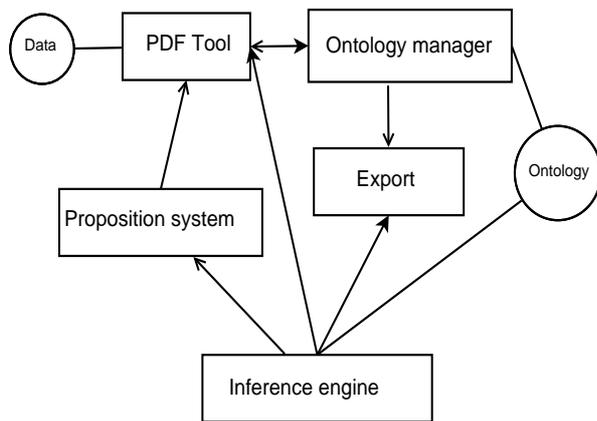


Figure 3: System architecture

open a PDF document, turn pages in different ways, zoom and print. Based on these utilities, as shown in Figure 4, the following functionalities have been added and are implemented in the prototype described in this paper:

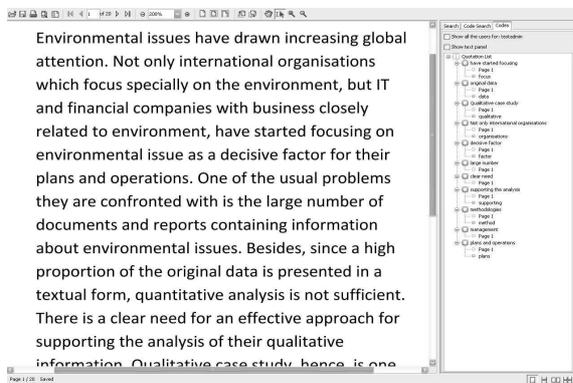


Figure 4: System screenshot

1. *Profile and project management* - For the sake of multiple users and projects, each active instance in the system is located in a scope characterised by its owner and the containing project. User name and project name are displayed at the menu bar of the system as a key. In Figure 5 all the projects previously created are listed for the users to select.



Figure 5: Project list

2. *Quotation* - Users are able to choose some of the words, sentences or paragraphs from the primary document as a quotation. Each quotation contains as information its text, creation and modification time and date and the text size. Figure 6 indicates a quotation highlighted in the original text.

Environmental issues have drawn increasing global attention. Not only international organisations which focus specially on the environment, but IT and financial companies with business closely related to environment, have started focusing on environmental issue as a decisive factor for their

Figure 6: Quotation in the text

3. *Coding* - As demonstrated in Figure 7, once a user has selected some quotations from the document, he can add codes to the quotations as tags. As an important definition in grounded theory, codes are the key elements later used in the analysis provided for qualitative case studies.

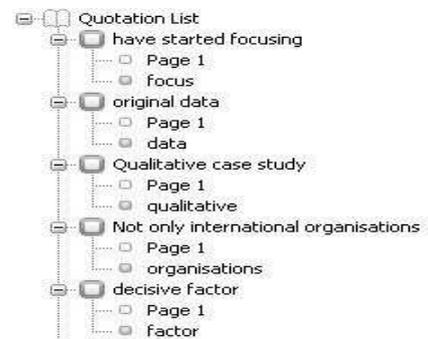


Figure 7: Code tree

4. *Code Search* - A user can specify interesting codes. Likewise he can also exclude words he does not want to find in the result set. Multiple key words are supported.
5. *Code Family* - In some cases, several codes are linked due to their inherent relationships, defined as a code family. Users are able to create a code family with existing codes, and then modify or delete them.

An ontology manager is used to store all the knowledge defined above in different ontologies. They are providing the necessary information for further analysis.

### 4.3 Inference Engine

We use Jena[Jena, 2011], a Java-based ontology API containing a rule engine, as our ontology inference system. Supported inference mechanisms are forward chaining, backward chaining and hybrid chaining. With the inference engine, we can learn rules and ontologies in order to generate new facts. The results of the inference process can be used in order to make propositions or they can be used by the exporting tool.

### 4.4 Proposition System

Since we have already extracted the knowledge from annotations in the original documents, the next step is to carry out

ontology learning and exploit inference to provide extended functionalities for qualitative case studies.

Once some codes are given by the users manually, they should be able to get proposition according to the manual coding in an automatic way. Different mechanisms can be used to provide this functionality. Let's assume a code is given (named *codeN*) to a quotation. In our system, a proposition is generated in several ways: first, the system can propose codes which appear together with *codeN* in other quotations. Second, other codes from the same code family as *codeN* are proposed. Third, if the quotation itself contains some key words which are of interest, they are also proposed.

Moreover, data mining algorithms are applied to the data in our system. Currently we have integrated association rule approach, which builds all the item sets by learning and inferring from the ontologies and then adopts a-priori and predictive-a-priori algorithms to generate rules. With these rules, propositional codes are produced for the users according to what they have already coded. The proposition system is a typical case in which ontology learning and inferring are made use of.

#### 4.5 Export and Visualisation

In qualitative research programs it is often important to be able to export the collected data to allow further analysis e.g. in statistical tools. In the standard GT approach this is difficult to realise as codes, concepts, categories and theories are stored in different places. The ontology approach overcomes this difficulty because all the information is stored in one single system and in a common format. Therefore we are able to define and implement an export mechanism based on ontological inference. This is provided by using the OWL query language which will allow a user to specify dynamically which elements of the ontology he wants to export. The data is exported in a portable spread sheet format in order to ensure an easy interoperability with other analysis tools.

Another form of export is to visualise the existing knowledge. For instance, users of the system can give certain ratings to the quotations and codes in each page. They will then be visualised in the form of a rating distribution over the whole document in terms of page numbers. In Figure 8, the X-axis lists all the pages in the document and the Y-axis illustrates the ratings provided by the users. The full mark of the rating can be modified in the configuration file of the system.

#### 4.6 Comparison

As stated above, ATLAS.ti is the most popular software package used in a case study setting and it serves also as the research benchmark of our system. Here we are going to compare the main features of ATLAS.ti and of our system.

From the comparison in Table 1 we can observe that Qualogier has the advantages of learning extracted knowledge to produce analytical results and output them. These are global advantages frequently used by the users and furthermore all these features can be personalised to meet each user's specific needs.

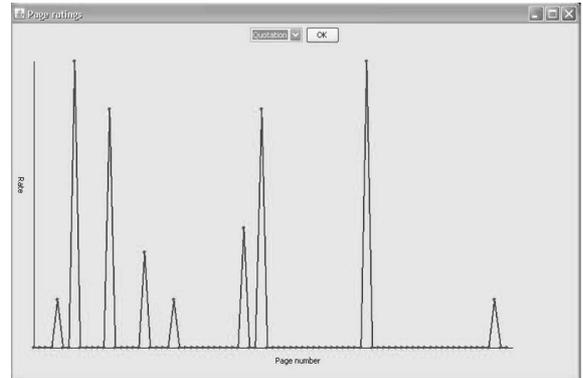


Figure 8: Rating distribution

Features	ATLAS.ti	Qualogier
Basic PDF functionalities	Yes	Yes
Quotation and code labelling	Yes	Yes
Code rating	No	Yes
Code time range	No	Yes
Code proposition	No	Yes
Code family	Yes	Yes
Flexible export	No	Yes
Node network	Yes	No

Table 1: Comparison table

## 5 Experiment and Use Cases

One of the big challenges in using ontologies is the issue of scalability. When data is getting larger, computational complexity is increasing accordingly. In order to test the efficiency of the system, we have carried out an experiment for adding new codes. A group of ontology files is generated containing a number of quotations varying from one to one thousand. We are going to add new codes to these files to verify the processing time required by the system for automatically iterating over all the files.

While adding a new code, we need first to search for its corresponding quotation in the file, insert the relation between this code and its quotation, and then add the code instance. For each file, we added one hundred codes and calculated the average time required. The experiment was conducted on a PC running Windows XP SP3 with a 2.80GHz CPU and in total of 4GB physical memory. In Figure 9, the X-axis represents the number of existing quotations in each file and the Y-axis indicates the processing time used for adding a new code in millisecond.

From the experimental result, we can see that in the case where a large amount of existing knowledge is already stored in the ontology files, the system still runs efficiently for supporting new coding.

Currently, Qualogier has been deployed to the research team specialised in the sustainability domains, with a focus on environmental analysis. They are conducting research case studies and experiments using this system to facilitate data organising, inferring and exporting. Compared to the

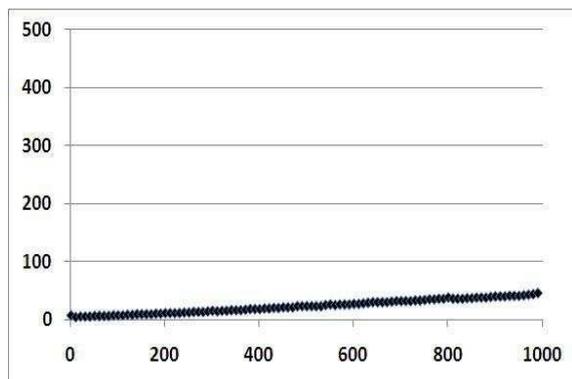


Figure 9: Coding experiment

traditional ways which was used before, ontology shows its advantages in that it is readable for both machines and humans, resulting a more efficient integration between domain analysis and computational output. Data from the original documents is well structured for different types of usage and the codes and concepts established can be maintained in a more consistent way. The system has been proven to be entirely satisfactory for their project.

Furthermore, it is discovered that the methodologies proposed in this paper are specifically feasible and practical for the environment-like topics. As we know, the field of environment contains a large group of concepts with multiple branches. Each of them have influence on each other, either in an implicit or explicit way. The existing way, like relational database, is not sufficient to represent this kind of characteristics. Also, the phenomenon of environmental topics are often resulted by chains of factors. With the reasoning mechanisms of ontologies, the analysis to these processes is more effective. At the same time, different forms of reports and outputs are demanded by the experts for cooperative or conclusive purpose. The query and export capacity of ontologies fits well these utilities.

## 6 Future Work and Development

As an ongoing interdisciplinary project, it is far-reaching to have reasonable design of the development plan in the coming steps to improve the methodologies and their implementation from different perspective. Based on our previous analysis and description, we will focus on the following aspects:

First, regarding the mechanisms of ontology learning and proposition, we have found that a trade-off is significant between efficiency and accuracy. Since the data is undoubtedly large and diverse, computational duration is usually long. On the other hand, the users of the system desire precise results transparent from the data structure and scalability. We are thus going to find the balance between these two factors to improve our proposed methodologies.

Then, from the technical point of view, a more dynamic and flexible architecture of the system is in demand. Currently the prototype is designed without consideration of reusability and adaptation. As the applications evolve, we need the system suitable for more scenarios not only limited to sustainability

domains and particular data structure. Compositional modularisation is of necessity for the next development.

In addition, it is also significant to set up the evaluation criteria to compare our system and its benchmark. This methodology of the evaluation is supposed to be generic and feasible for all the qualitative software. Besides, we need to design an entire and complete business flow to support and demonstrate how our methodologies are integrated into the processes involving data, system and experts.

Moreover, we will formulate a more advanced research framework that permits for describing and interpreting the links between environmental causes and effects. More domain knowledge will be brought into our project from different sources.

Finally, we will also set up more sophisticated approaches for output evaluation and validation. This is in great necessity because we are supposed to know how the methodologies proposed in this paper and their implementation meet the demand of the users, especially domain experts. We are currently designing some experiments to fulfil this initiative.

## 7 Conclusion

The goal of this paper, on the one hand, is to make a contribution in the field of case based studies by adding ontologies into the overall process, and on the other hand to integrate the approach of building theories from case studies to the field of ontologies in computer science. The approach would then be validated using real case study research, projected to develop an overall understanding of environmental management orientations and their impact on financial service firms.

In order to achieve this goal we first develop a formal model for case study research based on grounded theory. This model is formalised in the form of ontologies. The ontology approach allows for the validation of a user's work in accordance with the project, domain experts and grounded theory ontologies which are used as constraining reference systems. Based on this model, a prototype implementing the ontology layers and supporting supplementary functionalities has been created and used as a working tool for the case studies in environmental management. In general, we are applying ontologies to enable and facilitate the complete theory development, especially in sustainability domains.

## Acknowledgments

The authors wish to thank Swiss National Science Foundation (SNSF) for supporting this work and the reviewers for their time and efforts for reviewing this paper. We would like to thank our cooperative research team for their collaborative work. Thanks to Paul Cotofrei and Christophe Künzi for their help and advice to this work.

## References

- [ATLAS.ti, 2011] ATLAS.ti. ATLAS.ti: The Qualitative Data Analysis Software. <http://www.atlasti.com>, 2011.
- [Benz and Hotho, 2007] Dominik Benz and Andreas Hotho. Ontology learning from folksonomies. In *Lernen -*

- Wissen - Adaption, Workshop Proceedings (LWA 2007)*, pages 109–112, Halle, Germany, September 2007. Martin-Luther-University Halle-Wittenberg.
- [Buitelaar *et al.*, 2005] Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini. Ontology learning from text: An overview. In *Ontology Learning from Text: Methods, Applications and Evaluation*, pages 3–12, Amsterdam, Netherlands, July 2005. IOS Press.
- [Glaser, 1992] Barney G. Glaser. *Basics of grounded theory analysis*. Sociology Press, Mill Valley, California, 1992.
- [GRI, 2011] GRI. Home Page of Global Reporting Initiative. <http://www.globalreporting.org>, 2011.
- [Hepp and Roman, 2007] Martin Hepp and Dumitru Roman. An ontology framework for semantic business process management. In *Proceedings of Wirtschaftsinformatik 2007*, pages 423–440, Karlsruhe, Germany, February–March 2007. Universitaetsverlag Karlsruhe.
- [ICEpdf, 2011] ICEpdf. Home Page of ICEpdf. <http://www.icepdf.org>, 2011.
- [Jena, 2011] Jena. Jena Semantic Web Framework. <http://jena.sourceforge.net>, 2011.
- [Maedche and Staab, 2001] Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2):72 – 79, March 2001.
- [Strauss, 1987] Anselm L. Strauss. *Qualitative analysis for social scientists*. Cambridge University Press, Cambridge, United Kingdom, 1987.
- [Stuart *et al.*, 2002] I. Stuart, D. McCutcheon, R. Handfield, R. McLachlin, and D. Samson. Effective case research in operations management: a process perspective. *Journal of Operations Management*, 20(5):419–433, September 2002.
- [Taylor and Bogdan, 1998] Steven J. Taylor and Robert Bogdan. *Introduction to Qualitative Research Methods*. Wiley, Hoboken, New Jersey, 1998.
- [Yin, 2009] Robert K. Yin. *Case study research: Design and methods*. Sage Publications, Inc, Thousand Oaks, California, 2009.