A Framework for Assessing Commensurability of Semantic Web Ontologies

Liam Magee
RMIT University, Melbourne, Australia
liam.magee@rmit.edu.au

Abstract: The Semantic Web proposes a framework for establishing a “web of data”, analogous to the “web of documents” of the World Wide Web. It envisions a series of interconnected ontologies, underwritten by formal languages such as OWL and RDF. The problem of co-ordinating disparate ontologies has led to the development of various ontology matching approaches. However, as these approaches are algorithmic they cannot make use of background or tacit information about the ontologies they examine - information only available in the broader social context in which ontologies are created and used. In many practical knowledge management scenarios, such information is vital in understanding the costs, feasibility and scope of ontology alignment projects. Prior to undertaking the detailed task of concept-to-concept mapping between two ontologies, it is therefore useful to ask: are these ontologies broadly commensurable? This paper presents a framework for describing and comparing cultural information about ontologies, developed as part of a joint project conducted by RMIT University and FujiXerox Australia, “Towards the ‘Semantic Web’: Standards and Interoperability across Document Management and Publishing Supply Chains”. The framework is intended for practitioners to use as a tool to arrive at better estimates and assessments of the scope of work required to develop an adequate translation between two or more ontologies. The framework has been piloted as an online software toolkit, which is presented to a small group of participants. After using the software, participants complete an evaluation, which elicits quantitative and qualitative feedback on both the framework and the software. The paper presents the results of the pilot testing process, along with some considerations of how the framework might be further improved.

Keywords: ontology matching, commensurability, knowledge-producing cultures

1. Introduction

A key component of contemporary organisational strategies for knowledge management involves making knowledge explicit in canonical and reusable form (Härtwig and Böhm, 2006). As organisational boundaries become increasingly porous, knowledge also need to be transmittable across information networks (Choi and Whinston, 2000). By representing knowledge in formal languages with clearly defined semantics, it is possible to infer new facts from existing data sets and knowledge bases (Baader and Nutt, 2003). These three conditions – explicit, networkable and formally defined representations of knowledge – are met by the Semantic Web, a proposal for establishing a ‘web of data’, analogous to the ‘web of documents’ created by the World Wide Web (Berners-Lee et al., 2001; Shadbolt et al., 2006). By building on the successful distributed architecture of the Internet, the Semantic Web promises to connect existing information threads into a seamless tapestry of knowledge. From a knowledge management perspective, the Semantic Web offers a prospective ‘circuit breaker’ to conventional informational silos, leading to better utilisation of knowledge assets.

Ontologies, formal representations of knowledge developed in OWL and RDF languages, are at the core of the Semantic Web vision (Manola et al., 2004; McGuinness et al., 2004). Ontologies developed for particular domains can be shared, re-used and reasoned over, leading to a network of linked data with common conceptual foundations. Research in ontology modelling has been active in a number of domains, including upper-level ontologies (Niles and Pease, 2001; Masolo et al, 2001) and biological ontologies (Golbreich, C. and Horrocks, I, 2007; Smith, B. et al, 2007). In industrial contexts, however, shared ontological foundations can be difficult to arrive at. In the ten years since the initial publication of Semantic Web specifications, the beneficial network effects of ontological standardisation have been countered by high incipient costs of cross-organisational ontology development and curation, the complexities of managing conceptual relations between ontologies (Halevy, 2005) and the relative simplicity of alternative, purely syntactic information definitions, such as XML Schema and Microformats (Fallside and Walmsley, 2004; Khare and Çelik, 2006). Despite the recent emergence of a number of commercial Semantic Web “success stories” (Prevost, 2008), these factors have inhibited industry uptake (Shadbolt et al., 2006).

Considerable research has been dedicated to solving at least one of these problems, the co-ordination of disparate ontologies across common or related domains, by developing alignments. Ontology matching approaches seek to exploit different linguistic and semantic properties of

Reference this paper as
Magee, L.(2010) “A Framework for Assessing Commensurability of Semantic Web Ontologies” *Electronic Journal of Knowledge Management Volume 8 Issue 1 (pp91 - 102), available online at www.ejkm com*

ontologies to output a series of matches between different ontological concepts. As these approaches are invariably algorithmic they do not make use of background or tacit information about the ontologies they examine - information only available in the broader social context in which ontologies are developed and used. This study aims to augment such techniques by providing a framework for understanding the tacit assumptions behind ontologies, in order to assess their general degree of fit, or commensurability.

2. Background

Ontology matching aims to generate a series of conceptual matches between two ontologies (Shvaiko and Euzenat, 2005). A match consists of a tuple <*id,e,e',n,R*>, where *id* is the identifier of the match, *e* and *e'* are the two concepts from the two respective ontologies, *n* is the (optional) level of confidence in the match, and *R* is the relationship between the concepts (Shvaiko and Euzenat, 2005). Techniques can be broadly classified into syntactic and semantic approaches (Giunchiglia et al, 2007). Syntactic approaches examine, for example, the string edit distance (how similar two concept tokens are) or tree edit distance (where two concepts are located within a concept hierarchy). Semantic approaches look for logical relations between concepts, such as synonymy and subsumption (Shvaiko and Euzenat, 2005). Syntactic approaches typically generate probabilistic outputs for each candidate match; semantic approaches generate the logical relations themselves, sometimes with a corresponding confidence measure (Giunchiglia et al, 2007; Shvaiko and Euzenat, 2005).

A common feature of both syntactic and semantic techniques is that candidate terms are examined pair-wise to determine the probability and logical relationship of a match. While external sources are consulted in some approaches – for example WordNet is used to disambiguate term senses in the S-Match algorithm (Schvaiko, 2005) – this is to help improve precision of individual concept matches. What motivates the selection of concepts and relations in an ontology, and how they are used in ontological applications – the contextual conditions of ontology development and use – is a supplementary problem for the analyst conducting the alignment. While current state-of-the-art algorithms have demonstrated impressive precision and recall results against humanly-engineered mappings (Marie and Gal, 2008), there remain substantial challenges as Shvaiko and Euzenat (2008) have identified. Three of those challenges involving the *context* of ontologies (“discovering missing background knowledge”, “user involvement”, “social and collaborative ontology matching”) motivate the development of the framework presented here.

The role of context in the production and interpretation of meaning is widely acknowledged across a range of disciplinary fields, including artificial intelligence (Giunchiglia 1992; Akman and Surav 1996; Bouquet et al. 2003), social studies of science (Almklov 2008) and knowledge management (Loyola 2007). The notion of ‘context’ itself differs widely. The specific form of context of interest here is assumed to be objective (actually existent), social (comprised of social actors, structures and relations) and extrinsic to a system of meaning (not directly inferable from within the system of meaning itself). Transposed to specific knowledge artefacts such as ontologies, “context” can be understood to be the cultural conditions, practices and beliefs which inform their creation and use.

The *goal* of understanding the context of ontologies is to aid an analyst in the planning, estimation and scoping of an ontology alignment undertaking. Concept-to-concept mapping is generally an iterative, time-intensive and error-prone task, even with the assistance of algorithmic and other tooling approaches (Halevy, 2005). Background contextual information has been found to be useful in managing complexity and fostering “scalable community-ground ontology engineering” (De Moor at al, 2006). However the sort of detailed ethnographic studies of knowledge-producing cultures which might yield a deep understanding of context often involve multiple researchers or durations of months or years (Latour and Woolgar, 1986; Sommerlund, 2006; Almklov, 2008). The motivation of the present study, then, is whether it is possible to use lightweight, agile methods to capture, model and utilise such information within the time and cost constraints imposed on practical projects.

The proposed framework is introduced for *profiling* contexts, to draw out the tacit cultural assumptions which underpin ontologies. These profiles in turn serve the purpose of helping assess the general cultural fit or *commensurability* of ontologies. This approach adopts a form of conceptual or semantic *holism* (Brandom, 2000), in which specific relations between concepts belonging to different schemes are interpreted against an understanding of the broader commensurability of the schemes. This contrasts with the focus of ontology matching techniques, which generate a set of *atomic* – term-by-term – matches, given equivalent semantic properties such as the same necessary and sufficient conditions for class membership. Broadly, these techniques adopt a *correspondence* theory of truth – two terms are equivalent just if their corresponding extensions are judged to be equivalent. The approach advocated here aims to supplement these techniques by ushering in *coherentist* and *consensual* truth notions – ontologies as a whole are commensurable to the extent their merging or alignment entails no incoherent beliefs, and their users agree on this coherence. Practically, the profiling framework supplies a means for an analyst engaged in ontology matching to estimate the scope of a matching exercise, and to validate an existing match against coarse-grained commensurability indicators. The approach draws heavily on both phenomenological (Heidegger (1962, pp. 194-5), Gadamer (2004, p. 269)) and pragmatist (Rorty, 1982; Brandom, 2000) philosophical traditions. In particular, the reciprocal relationship of specific – atomistic – interpretation and general – holistic – understanding is articulated at higher levels of abstraction within phenomenology and hermeneutics. An equivalent reciprocity between the outputs of ontology matching algorithms and this framework is suggested here.

3. Modelling ontology commensurability

A model of an idealised ontology alignment scenario is presented below. Firstly the main concepts of the model are described; then the model is formalised, to generate a quantitative commensurability value. This value in turn can be used as an indicator of time, cost, required resources or other tangible project metrics.

3.1 Descriptive model

The main concepts used in the model are:

PROJECT: A project is the context in which ontology matching takes place. A schema-matching project will involve at least two schemas, and the usual features of projects: a designated purpose; one or more stakeholders; resources, human and otherwise, to accomplish the project tasks; and a set of environmental constraints.

ANALYST: Within a project, an individual or group is typically tasked with the ontology matching exercise itself. Such a person or group is called the *analyst* - to describe a specific role as a *systems* analyst, and to capture a more general sense of analysing the background information of schemas.

SCHEMA: While ontologies are the canonical knowledge representation form, the framework can applied to other representation forms, including XML schemas, relational and UML models. Such representations are termed *schemas*, to emphasise the relation between the explicit representation itself to an underlying implicit *cultural conceptual scheme* – a shared set of concepts and relations held by a community of practice.

CULTURE: *Schemas* in turn are created and used by groups of people. A *culture* here is designed to capture the beliefs and practices of a potentially amorphous collective who create or use a schema over time and space.

DIMENSION: Both the schemas and the cultures who produce and consume them are described through a two-tier taxonomy of dimensions (see Appendix 1). The *grouping* dimensions are used to describe schemas in *qualitative* terms; the *item* dimensions are used to describe schemas in *quantitative* terms. Dimensions can describe either a feature of schema (such as its size), or the relationship between two schemas (such as their degree of interdependency).

PROFILE: A *profile* is the collection of the dimension values (both qualitative and quantitative) for a particular schema, produced through an analysis of the schemas themselves and associated background data sources.

COMMENSURABILITY*:* The schema profiles form the basis of an analysis of the conceptual fit of schemas, or their *commensurability*. "Commensurability" is a term used heavily by Kuhn in his seminal discussion of scientific paradigms (Kuhn, 1970), to which this framework is heavily indebted. Here commensurability is applied not to broad historical periods of scientific enterprise but to local, specifically cultural orientations which form around the production and consumption of schemas. Moreover - reflecting pertinent criticisms of Kuhn's model (Davidson, 2006) - *in*commensurability is not taken here as indicating a *quality* of untranslatability of schemas, as they do for scientific paradigms. Rather, commensurability indicates a *quantitative degree* of conceptual fit. Transposed to a practical situation, it is an indicator of the scope of work, cost or resources required for schema translation. A basic depiction of the model is provided in Illustration 1 below.

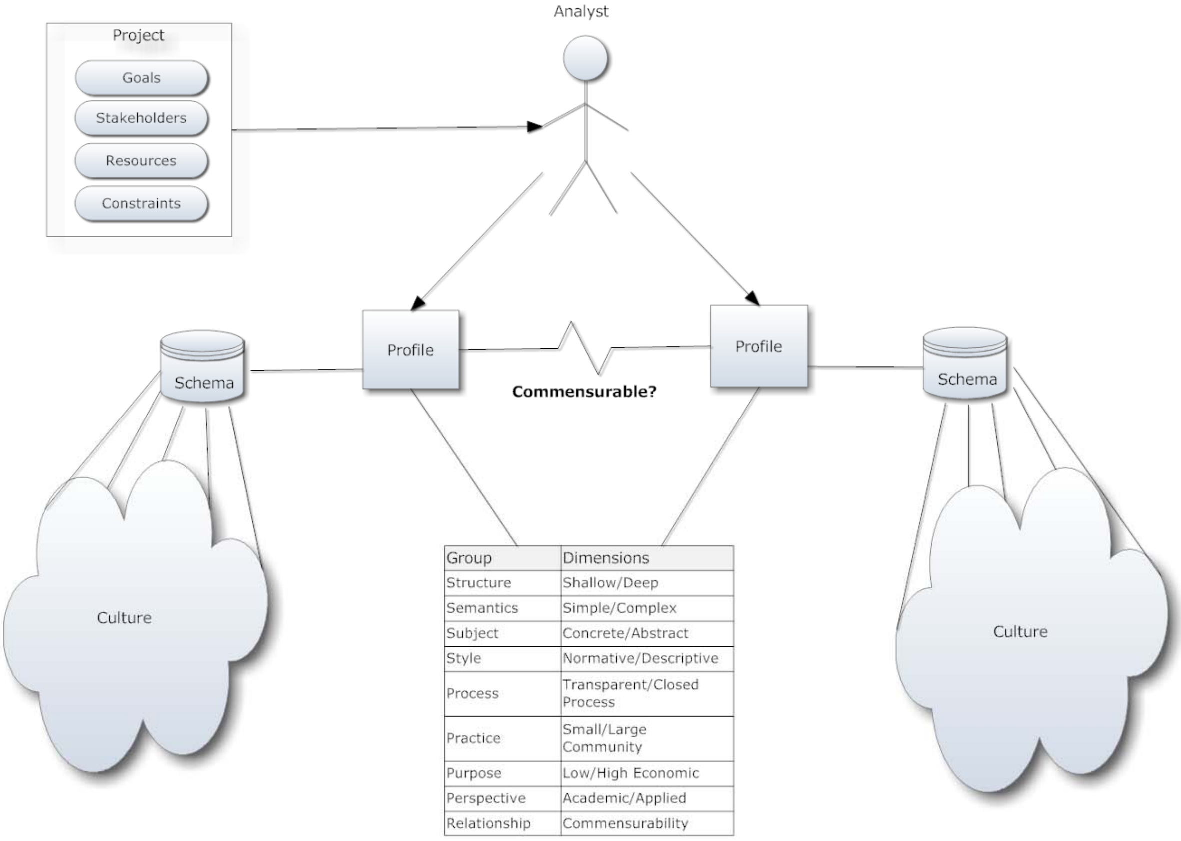


Figure 1: Commensurability model

3.2 Formal model

It is possible to formalise the descriptive model described above to provide a quantitative value for commensurability. The output, the *commensurability factor (CF),* represents the *relative* effort required to match two schemas, *S1* and *S2*. It is a heuristic measure which can be incorporated into a project planning or assessment activity.

The model distinguishes two kinds of dimensions, those which describe a schema directly, and those which describe the relationship between two schemas. The *absolute differences between the values* of the schema dimensions represent one set of inputs, while the *actual values* of the relationship dimensions represent another set. Let DF = $\{df_1, df_2...df_n\}$ be the set of differences for values of the schema dimensions, where *n* is the number of schema dimensions, and let R = $\{r_1, r_2...r_m\}$ be the set of relationship dimensions, where *m* is the number of relationship dimensions, for *S1* and *S2*.

Since the dimensions can be weighted, let SW = $\{sw_1, sw_2...sw_n\}$ be the respective weights for each of the schema dimensions, and let RW = $\{rw_1, rw_2...rw_m\}$ be the weights of the relationship dimensions. The commensurability factor, *CF*, is then calculated as the arithmetic weighted average of both schema and relationship values as follows:

$$CF = \frac{\sum_{i=1}^{n} df_i sw_i + \sum_{j=1}^{m} r_j rw_j}{\sum_{i=1}^{n} sw_i + \sum_{j=1}^{m} rw_j}$$

4. A framework for commensurability

The framework builds upon the preceding model by adding a) a set of dimensions and b) a lightweight methodology for assigning values to the dimensions.

4.1 Dimensions

The framework presents dimensions in a two-tiered taxonomy, composed of general dimension groups and the specific dimensions themselves. (In terms of the commensurability algorithm presented above, *group* weighting can be easily accommodated). The default set of dimensions attempts to characterise generally salient properties of ontologies, those properties which are most relevant to compare. Selection of appropriate generic dimensions has been to some extent guided by existing literature on ontology metrics (Burton-Jones et al., 2005; Vrandei and Sure, 2007) and broader comparative analyses of schemas (Schmitz and Leukel, 2005). In practice, in addition to relative weightings, the groups and dimensions may need to be customised to contextually-salient conditions.

The groups are loosely ordered, from intrinsic to extrinsic dimensions. The intrinsic dimensions characterise the internal properties of an ontology. These are grouped into four categories, as shown in *Table 1*. Ontology metrics, obtained from libraries such as OWLAPI [cite], can be used to derive values for some of the dimensions.

Table 1: Intrinsic properties

Group	Description
Structure	How the ontology is organised and modelled
Semantics	How the ontology relates to things in the world
Subject	What sorts of things the ontology models
Style	How the ontology is written

The extrinsic dimensions describe the sociological, theoretical and perspectival features of ontologies, as outlined in *Table 2*.

Table 2: Extrinsic properties

Group	Description
Process	How the ontology has been developed
Practice	How ontology is used in practice
Purpose	Why the ontology exists, and what motivates its development
Perspective	What general perspective is adopted by the ontology

A final category, *Relationship*, covers how two ontologies are related – for example, whether they cover similar areas, or whether they are logically compatible.

Together the intrinsic, extrinsic and relationship dimensions aim to profile how an ontology is modelled and used. Quantitative and qualitative valuations of two ontologies against the dimensions form the basis for an assessment of their commensurability. The dimensions were developed iteratively in conjunction with several case studies exploring the application of the framework to upper-level ontologies (BFO, GFO, PROTON, SUMO, DOLCE), document schemas (OOXML, ODF) and biological ontologies (GeneOntology, SNOMED-CT). The default set of dimensions (listed in full in Appendix 1) were chosen since they cover generic yet salient aspects of ontology *context*; however they can be further weighted, adjusted or extended to suit particular profiling scenarios – where, for instance, how ontologies are used practically might be more relevant than other factors.

4.2 Methodology

The methodology component of the framework is designed to guide an analyst in assigning values to the dimensions and developing an evaluation report on the commensurability of ontologies. It applies a redacted form of online discourse analysis (Schneider and Foot, 2004).

The steps below outline the procedure for developing an evaluation of commensurability. The steps are not “lock-step”, in that any step may be done in any order, and also re-done iteratively. However their basic organisation suggests a relatively linear process. The steps are:

- Configure the project: describe the project in terms of its purpose, constraints, resources and stakeholders.

- Modify the dimensions: weight the default dimension set, and optionally, add and remove dimensions from the set.
- Describe method and sources: locate background documentary or other kinds of evidence which can be used to develop profiles of the ontologies, and optionally, code the sources in terms of relevant user-defined categories. For example, a source may indicate that a particular ontology is a well-defined and well-supported standard, but excludes users from involvement in the development process.
- Profile the schemas: evaluate the schemas in terms of the qualitative and quantitative dimensions constructed in step 2, and using the sources added in step 3.
- Develop an analysis: analyse the commensurability of the profiles produced in step 4. The analysis may consider questions like potential scoping estimates for ontology alignment, specific problem areas and possible solutions.
- Construct a commensurability report: brings together all of the preceding elements into a report, which can be used for further evaluation and dissemination.

5. Evaluating the framework

The framework has been operationalised in the form of an online software toolkit. The toolkit includes the default set of dimensions and the methodology discussed above. A pilot evaluation of the software has been conducted with a hand-selected participant sample. The intention of the pilot is to refine various aspects of the framework and software prior to introducing it to a larger audience. As such, the pilot results are suggestive only, indicating problem areas with both the framework and the software implementation. A further, more extensive evaluation is planned once the pilot results have been analysed and suggestions have been incorporated into the toolkit.

In the pilot participants are invited to review a sample project comparing two commonly used tutorial ontologies, the Pizza (Rector et al., 2004) and Wine (Noy and McGuinness, 2001) ontologies. The project presents a scenario in which these ontologies need to be merged into a knowledge base of a new restaurant point-of-sale system. A very basic form of discourse analysis has been conducted into the background literature describing the cultural perspectives of these ontologies, and a quantitative evaluation against the default dimension set is also supplied. Finally a finished commensurability report is presented. A screen-shot of the software, demonstrating how values are applied to dimensions for the ontologies, is shown in Appendix 3.

The pilot test was conducted with a small sample of 13 participants, recruited non-randomly from RMIT University and affiliated organisations. The participants were selected because they had some background with data integration, Semantic Web and ontology matching technologies, or alternately because they were familiar with the social science research methods presented in the toolkit.

Participants are asked to browse both the interim steps and the completed report, and then to complete an evaluation survey. The survey consists of a series of Likert scale and open-ended questions, designed to elicit feedback on the utility, ease of use and relevance of the following elements:

- dimension model – the default dimensions used to measure the schemas
- methodology – the process of profiling the schemas
- reporting mechanism – the final analysis of commensurability
- software system – the implementation of the framework
- overall framework – how well the parts integrate into the whole

5.1 Quantitative results

The participants were asked to respond to a total of 24 Likert items. Three of these related to degrees of familiarity with, respectively, data integration, Semantic Web and ontology matching technologies. The remaining 21 items relate to the following attitudinal scales:

Area	No. of items
Default dimension set	4 items
Methodology	4 items
The commensurability analysis	4 items
The software system	4 items
Overall	5 items

All of the items were rated from “Strongly Disagree” to “Strongly Agree”, and coded 1 to 5 accordingly. As the summary statistics in Appendix 2 show, responses to items were favourable, with median and mode values of 4 (“Agree” - adjusted for reversed scales) for 81% and 76% of the attitudinal scales respectively. Chi-square tests were also conducted on each of the items, with 90% of items significant at the $p > 0.05$ level and 57% significant at the $p > 0.01$ level. These results suggest the toolkit rated well generally.

5.2 Qualitative results

In addition, participants were asked open-ended questions about how the model, methodology, report and analysis, software and overall toolkit could be improved, and invited to add further comments. Nine out of the thirteen participants offered some kind of written feedback. Although this feedback was varied, a common criticism focussed on the lack of explanation of how and why the system could be applied in practice, especially in relation to the dimension model and methodology.

In relation to the model, one participant offered:

> *Participant C: The model and its supporting documentation seem extremely easy to use and clear if someone already understands the "why" of doing this sort of comparison. If the tool were intended for professionals doing data integration projects on the ground (rather than people involved directly in more formal discussions of ontologies), it would probably need more upfront discussion of how the tool can be used in more everyday professional practice.*

Several other participants found the application of the model to the example scenario confusing:

> *Participant A: How to make things simple would be of great importance to the use of the model*

> *Participant B: I was confused as to what the model actually does also; how the values/weights are used.*

> *Participant G: The default model captures a vast array of possible dimensions on which to compare ontologies but perhaps only a few are actually important for any particular project.*

Two participants suggested simplifying or reorganising the dimensions:

> *Participant D: Some of the heads for qualitative commentary could be consolidated - there are perhaps too many, and more scope could be given to users to create their own heads.*

> *Participant F: One improvement might be to break up the the default model into several 'brackets' that might be more or less suited to different kinds of mapping challenges*

Participants also found the methodology instructions too broad:

> *Participant C: A specific case that ties into the specific example provided, would help "lay" users visualise how they might arrive at the point where they would develop a specific methodology and fill out other components of the tool, so that it would be easier to recognise what I think is a very practical, useful and intuitive environment for foreseeing potential complexities in a project.*

> *Participant F: The process of gathering background information to the mapping work is a little hazy... it might be possible to be far more directive and specific about the work that needs to done to prior to commencing the weighting against the default model.*

> *Participant G: The methodology is too broad and lacks focus.*

In spite of these reservations, feedback about the overall toolkit and framework were positive:

> *Participant C: The underlying conceptual framework and toolkit seem like they would be quite useful in professional practice*

> *Participant D: Nevertheless, scaffolds like this are very helpful*

> *Participant F: Your default model grounds the work of mapping in concrete ways and this I think is a great invention.*

> *Participant H: Conceptually fascinating - and I can see how the questions posed in the process of designing a project could act as a point of dialogue between different components of a project team*

5.3 Discussion

The quantitative results indicated statistically significant positive results on the majority of the items presented. Qualitative feedback suggested a more complex picture, with criticisms largely directed towards the framework model and methodology. Overall, both forms of feedback indicated that the toolkit would be useful in improving *understanding* and *accuracy* of schema matching projects, although the results were less conclusive as to whether it would help lower *time* and *cost* components.

Some of the qualitative criticisms can be addressed through more specific documentation and user interface improvements. Others suggest that both the dimension model and methodology need to be tailored to practical scenarios – this could be achieved by developing “templated” models and associated activities for specific use cases, perhaps on a vertical industry basis or by identification of particular business “patterns” for ontology comparison and alignment. After these criticisms have been addressed, it is expected that further evaluations of the framework will be conducted via a series of case studies on organisations involved in real-world ontology matching.

6. Conclusion

The problem of ontology matching has been well documented in the technical literature. The proposed framework applies a sociological approach to understanding the background cultural context of ontologies. It is intended that the framework will augment use of ontology matching algorithms, which focus on atomistic matching at the concept level. The paper makes several contributions: it has presented a) a generalised model of an ontology alignment scenario; b) a framework which supplies a set of dimensions and a methodology; c) a software toolkit which operationalises the framework, and d) results of a pilot study assessing the utility of the toolkit.

The pilot study shows promising results for the utility value of the framework. However more work needs to be done simplifying the description of both the dimensions and methodology components of the framework. The focus of the pilot has been to evaluate the framework at a general, abstract level, further work is required to evaluate the tool-kit in relation to real-world projects.

As progressive layers of the Semantic Web reach maturity, attention has turned to subsidiary problems of ontology alignment, curation and management. Frameworks oriented towards contextual understanding will be increasingly helpful for organisations introducing ontologies into their knowledge management strategies.

7. Appendix 1 - default dimensions

Dimension Group	Dimensions	Description
Structure	Small—Large Self-contained—Derivative Shallow—Deep	Describes features inherent in the schema itself: for example, does it contain a relatively small or large

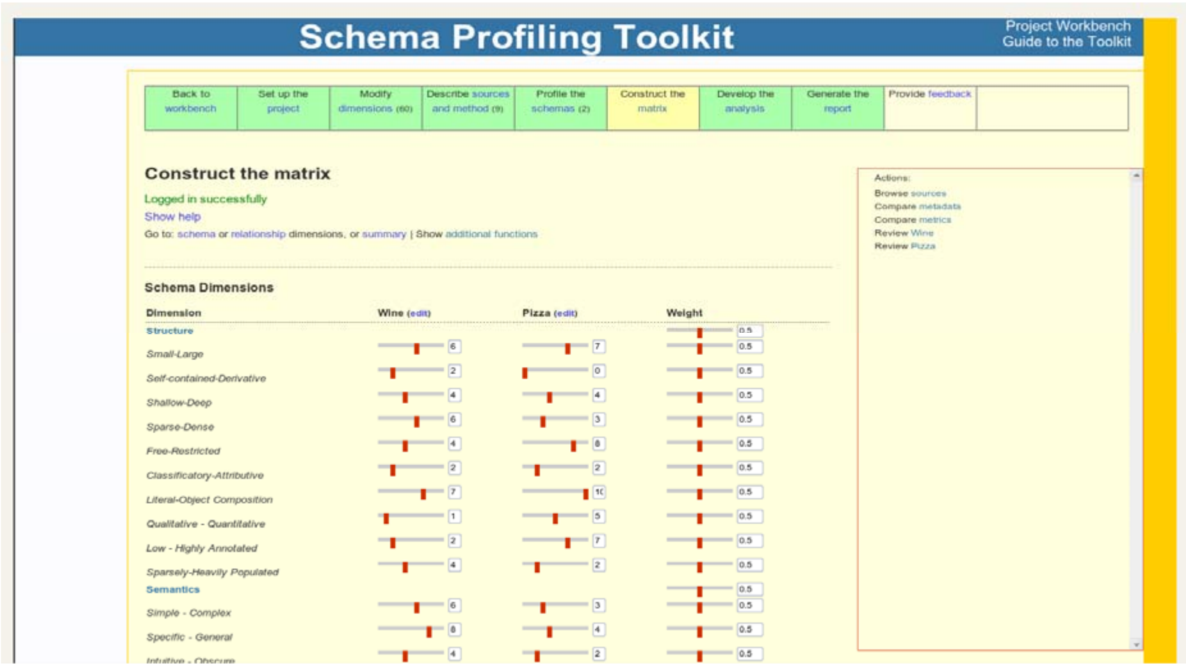
	Sparse—Dense Free—Restricted Classificatory—Attributive Literal—Object Composition Qualitative—Quantitative Lowly—Highly Annotated Sparsely—Heavily Populated	number of constructs?
Semantics	Simple—Complex Specific—General Intuitive—Obscure Dispersed—Concentrated Random—Coherent Inaccurate—Accurate Incomplete—Complete	Describes the relationship of the schema to things in the world: for example, is it a simplified view of objects in the domain?
Subject	Concrete—Abstract Spatial—Temporal Natural—Social	Describes the subject matter of the schema (in general terms): for example is the subject matter concrete or abstract?
Style	Light-hearted—Serious Normative—Descriptive Tentative—Committed	Describes the general manner in which the schema is written: for example, is it light-hearted or serious in tone?
Process	Lowly—Highly Representative of Users Distributed — Central Design Transparent — Closed Process Informal — Formal Decision Making Harmonious — Conflictual Design Process Implicit — Explicit Assumptions Ad hoc — Rigorous Design Method	Describes how the schema has been developed: for example, is it developed centrally (within an organisation), or in a distributed way?
Practice	Small — Large Community Low — Highly Active Community Declining — Increasing Adoption Rate Backwards Incompatible — Compatible De Facto Standardisation De Jure Standardisation Low or High Industry Support Low or High Availability of Documentation Low or High Levels of Satisfaction Recently published or mature	Describes how schema is used in practice: for example, is it used by a small or large community?
Purpose	Low—High Social Motivation Low—High Political Motivation Low—High Economic Motivation Low—High Scientific Motivation Low—High Technological Motivation Low—High Philosophical Motivation Low—High Educational Motivation Low—High Cultural Motivation Low—High Legal Motivation Low—High Environmental Motivation	Describes why the schema exists: for example, is it motivated by social, economic, political, technological factors?
Perspective	Every-day — Scientific Pragmatic — Idealistic Academic — Applied Grounded — Speculative Dependent — Independent	Describes the general perspective adopted by the schema: for example, does it adopt an everyday or scientific orientation towards the objects in the domain?
Relationship	Degree of conceptual overlap in domains Degree of conceptual translation required Logical consistency Mutual awareness Competitiveness Commensurability of perspectives	Describe the *relationships* between two or more schemas: for example, do they cover similar domains?

8. Appendix 2 - survey results

Item	Min	Max	Median	Mode	Chi-square
Familiar with data integration	1	5	4	4	0.122
Familiar with the Semantic Web	1	5	3	2	0.337
Familiar with ontology matching	1	5	3	2	0.337
Model understandable and clear	2	5	4	4	0.001
Model useful for this project	3	5	4	4	0.000
Model useful for other projects	2	5	4	4	0.085
Need to modify *model* to capture relevant dimensions	2	3	3	3	0.000
Methodology clear, understandable and easy to apply	2	5	4	4	0.001
Methodology complex and confusing *	1	4	3	2	0.040
Methodology useful for this project	3	5	4	3	0.028
Methodology useful for other projects	1	5	4	4	0.174
Analysis clear and understandable	2	5	4	4	0.004
Analysis complex and confusing *	1	4	2	2	0.040
Analysis useful for this project	3	5	4	4	0.003
Analysis useful for other projects	2	5	4	4	0.004
Software intuitive and easy to use	2	5	3	3	0.009
Software well documented	2	5	4	4	0.040
Software useful for this project	3	5	4	4	0.003
Software useful for other projects	2	5	4	4	0.028
Overall: Improve understanding of data schemas and ontologies	3	5	4	4	0.000
Overall: Useful for data integration projects	3	5	4	4	0.013
Overall: Reduce time data integration projects	3	5	3	3	0.000
Overall: Reduce costs of data integration projects	3	5	4	3	0.006
Overall: Improve accuracy of data integration projects	3	5	4	4	0.013

* Scales were reversed for these items

9. Appendix 3 – toolkit screen-shot



References

Akman, V. and Surav, M. (1996) “Steps toward formalizing context”, AI magazine, Vol. 17, No. 3, pp. 55–72.
Almklov, P. (2008) “Standardized Data and Singular Situations”, Social Studies of Science, Vol. 38, No. 6, p. 873.
Baader, F. and Nutt, W. (2003) “Basic Description Logics”. In The Description Logic Handbook: Theory, Implementation, and Applications, Cambridge University Press, New York.
Berners-Lee, T., Hendler, J. and Lassila, O. (2001) “The Semantic Web”. Scientific American, Vol. 284, No. 5, pp. 34–43.
Bouquet, P., Giunchiglia, F., Van Harmelen, F., Serafini, L. and Stuckenschmidt, H. (2003) “C-owl: Contextualizing ontologies”, Tech. Rep. 0306-06, Istituto per la Ricerca Scientifica e Tecnologica.
Brandom, R. (2000) Articulating Reasons: An Introduction to Inferentialism. Harvard University Press.

Burton-Jones, A., Storey, V., Sugumaran, V. and Ahluwalia, P. (2005) “A semiotic metrics suite for assessing the quality of ontologies”, Data & Knowledge Engineering, Vol. 55, No. 1, pp. 84–102.
Choi, S. and Whinston, A. (2000) “Benefits and requirements for interoperability in the electronic marketplace”, Technology in society, Vol. 22, No. 1, pp. 33–44.
Davidson, D. (2006) The Essential Davidson. Oxford University Press.
Dean, M., Schreiber, G. et al. (2004) “OWL Web Ontology Language Reference”, W3C Recommendation.
De Moor, A., De Leenheer, P. and Meersman, R. (2006) “DOGMA-MESS: A meaning evolution support system for interorganizational ontology engineering”, LNCS, Vol. 4068.
Fallside, D.C. Walmsley, P. (2004) “XML Schema Part 0: Primer Second Edition”, W3C recommendation, World Wide Web Consortium.
Gadamer, H. (2004) Truth and method. Continuum International Publishing Group.
Giunchiglia, F. (1992) “Contextual reasoning”, Tech. Rep. 9211-20, Istituto per la Ricerca Scientifica e Tecnologica.
Giunchiglia, F., Shvaiko, P. and Yatskevich, M. (2004) “S-match: an algorithm and an implementation of semantic matching”. In Proceedings of ESWS, pp. 61 – 75.
Giunchiglia, F., Shvaiko, P. and Yatskevich, J. (2007) “Semantic Matching”, The Knowledge Engineering Review, Vol. 18, No. 3, pp. 265 – 280.
Golbreich, C. and Horrocks, I. (2007) “The OBO to OWL Mapping, GO to OWL 1.1!”. In CEUR Workshop Proceedings, Vol. 258.
Halevy, A. (2005) “Why Your Data Won’t Mix”, Queue, Vol. 3, No. 8, pp. 50–58.
Härtwig, J. and Böhm, K. (2006) “A Process Framework for an Interoperable Semantic Enterprise Environment”. Electronic Journal of Knowledge Management, Vol. 4, No. 2, pp. 39–48.
Heidegger, M. (1962) Being and Time. Blackwell Publishing.
Khare, R. and Çelik, T. (2006) “Microformats: a Pragmatic Path to the Semantic Web”, in “Proceedings of the 15th international conference on World Wide Web”, pp. 865–866. ACM, New York, NY, USA.
Kuhn, T. (1970) The Structure of Scientific Revolutions. University of Chicago Press.
Latour, B. and Woolgar, S. (1986) Laboratory life: The construction of scientific facts. Princeton University Press.
Loyola, W. (2007) “Comparison of Approaches toward Formalising Context: Implementation Characteristics and Capacities”, The Electronic Journal of Knowledge Management, Vol. 5, pp. 203–215.
Manola, F., Miller, E. McBride, B. (2004) “RDF primer. W3C recommendation”, World Wide Web Consortium.
Marie, A. and Gal, A. (2008) “Boosting Schema Matchers”, Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part I on On the Move to Meaningful Internet Systems, pp. 283–300.
Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. and Schneider, L. (2001), “The WonderWeb library of foundational ontologies”, LADSEB-Cnr, Padova, IT, Preliminary Report D, Vol. 17.
McGuinness, D. and Van Harmelen, F. et al. (2004) “Owl web ontology language overview”, W3C recommendation, Vol. 10, pp. 2004–03.
Niles, I. and Pease, A. (2001) “Towards a standard upper ontology”. In Proceedings of the international conference on Formal Ontology in Information Systems, Vol. 2001, pp. 2-9.
Noy, N. and McGuinness, D. (2001) “Ontology Development 101: A Guide to Creating Your First Ontology”, Tech. Rep. KSL-01-05, Stanford Knowledge Systems Laboratory.
Rector, A., Drummond, N., Horridge, M., Rogers, J., Knublauch, H., Stevens, R., Wang, H. Wroe, C. (2004) “Owl Pizzas: Practical Experience of Teaching OWL-DL: Common errors and Common Patterns”, Lecture notes in computer science, pp. 63–81.
Rorty, R. (1982) Consequences of Pragmatism, University of Minnesota Press, Minneapolis.
Schmitz, V. Leukel, J. (2005) “Findings and Recommendations from a Pan-European Research Project: Comparative Analysis of E-Catalog Standards”, International Journal of IT Standards and Standardization Research, Vol. 3, No. 2, pp. 51–65.
Schneider, S.M. and Foot, K.A. (2004) “The Web as an Object of Study”, New Media Society, Vol. 6, No. 1, pp. 114–122.
Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) “The Semantic Web Revisited”, IEEE Intelligent Systems, Vol. 21, No. 3, pp. 96-101.
Shvaiko, P. and Euzenat, J. (2005) “A Survey of Schema-Based Matching Approaches”, Journal on Data Semantics IV, Vol. 3730, pp. 146–171.
Shvaiko, P. and Euzenat, J. (2008) “Ten Challenges for Ontology Matching”, LNCS, Vol. 5332, pp. 1164 – 1182.
Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, Susanna-A., and Scheuermann, R. H., Shah, N., Whetzel, P. L. and Lewis, S. (2007) “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration”, Nature Publishing Group, Vol. 25, No. 11.
Sommerlund, J. (2006) “Classifying Microorganisms: The Multiplicity of Classifications and Research Practices in Molecular Microbial Ecology”, Social Studies of Science, Vol. 36, No. 6, pp. 909–928.
Vrandei, D. and Sure, Y. (2007) “How to Design Better Ontology Metrics”. In “ESWC ’07: Proceedings of the 4th European conference on The Semantic Web”, pp. 311–325. Springer-Verlag, Berlin, Heidelberg.