Semantic Web Journal 0 (2010) 1
IOS Press

Creating Knowledge out of Interlinked Data

Editor(s): Krzysztof Janowicz, Pennsylvania State University, USA; Pascal Hitzler, Wright State University, USA
Solicited review(s): Claudia d'Amato, Università degli Studi di Bari, Italy; Rinke Hoekstra, Universiteit van Amsterdam, The Netherlands
Open review(s): Prateek Jain, Wright State University, USA

Sören Auer, Jens Lehmann
Universität Leipzig, Institut für Informatik,
PF 100920, 04009 Leipzig, Germany
{lastname}@informatik.uni-leipzig.de

Abstract. Over the past 3 years, the semantic web activity has gained momentum with the widespread publishing of structured data as RDF. The Linked Data paradigm has therefore evolved from a practical research idea into a very promising candidate for addressing one of the biggest challenges in the area of the Semantic Web vision: the exploitation of the Web as a platform for data and information integration. To translate this initial success into a world-scale reality, a number of research challenges need to be addressed: the performance gap between relational and RDF data management has to be closed, coherence and quality of data published on the Web have to be improved, provenance and trust on the Linked Data Web must be established and generally the entrance barrier for data publishers and users has to be lowered. In this vision statement we discuss these challenges and argue, that research approaches tackling these challenges should be integrated into a mutual refinement cycle. We also present two crucial use-cases for the widespread adoption of linked data.

Keywords: Linked Data, Semantic Web

One of the biggest challenges in the area of intelligent information management is the exploitation of the Web as a platform for data and information integration as well as for search and querying. Just as we publish unstructured textual information on the Web as HTML pages and search such information by using keyword-based search engines, we should be able to easily publish structured information, reliably interlink this information with other data published on the Web and search the resulting data space by using expressive querying. The Linked Data paradigm has evolved as a powerful enabler for the transition of the current document-oriented Web into a Web of interlinked Data and, ultimately, into the Semantic Web. The term Linked Data here refers to a set of best practices [4] for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the past three years, leading to the creation of a global data space that contains many billions of assertions - the Web of Linked Data. However, in order to sustainably establish the Web of Data and to maximize the value of published data on the Web, we are facing four fundamental challenges: (1) we have to improve the performance of very large-scale RDF Data Management, (2) we have to increase and ease the interlinking and fusion of information, (3) algorithms and tools have to be developed for improving the structure, semantic richness and quality of Linked Data, (4) adaptive user interfaces and interaction paradigms have to deployed for authoring and maintaining Linked Data.

In the remainder of this vision statement we elaborate on these challenges, possible approaches for solving them and present two crucial use-cases for linked data.[1]

1. Improving the Performance of Large-Scale RDF Data Management

Experience demonstrates that an RDF database can be an order of magnitude less efficient than a relational representation running on the same engine[5]. This lack of efficiency is perceived as the main obsta-

[1]The interested reader may also want to have a look at a related article [15] in this issue, which poses similar challenges on dealing with Linked Data from a slightly different perspective.

cle for a large-scale deployment of semantic technologies in corporate applications or for expressive Data Web search. For RDF to be the lingua franca of data integration, which is its birthright, its use must not bring significant performance penalty over the much less flexible best practices prevalent today. The main reason for the difference in performance between triple stores and relational databases is the presence of fine-grained optimised index structures in relational systems in contrast to more flexible and extensible structures in triple stores. The performance gap between relational and RDF data management can be mitigated by developing adaptive automatic data indexing technologies that create and exploit indexing structures as and when needed, entirely based on received query workload.

The performance of knowledge stores can, for example, be significantly increased by applying query subsumption and view maintenance approaches to the RDF data model. Query subsumption can be based on analysing the graph patterns of cached SPARQL queries in order to obtain information on (a) whether the previously cached query result can be reused for answering a subsequent query and (b) which updates of the underlying knowledge base will change the results of cached queries and thus have to trigger invalidation. As queries are executed, intermediate results can be persisted and labeled for reuse. When a subsequent query is executed, it can reuse those results if it either subsumes the previous query or is subsumed by it. In the latter case the query can be performed on the persisted previous results and in the first case join operations between persisted and base data can be performed. This builds query shortcuts across the data, essentially making materialized views on demand. These are either invalidated or brought up to data as data changes and discarded if no longer needed. The cacheable operations are joins and inferences such as `owl:sameAs`, transitive property traversal, class and property hierarchies etc. Intermediate materializations may also cache aggregates. First steps in this direction were for example performed in [14,6]. In addition, caching and view materialization techniques should be able to handle implicit information commonly found in ontologies. In order for Linked Data to be successful in the Web at large and within enterprises in particular, such new RDF indexing technology must ultimately find its way in RDF processing systems.

2. Increase and Ease the Interlinking and Fusion of Information

While the sum of data published as linked data amounts already to billions of triples and grows steadily, the number of links between them is several orders of magnitude smaller and by far more difficult to maintain (cf. [12]).

The task of interlinking and supplementing the knowledge bases with information from external data sets, knowledge bases and ontologies can draw from previous work within different research communities: Interlinking has a long history in database research and occurs in the literature under a dozen of terms [19] such as Deduplication[9], Entity Identification [13], Record Linkage [7] and many more. Encountered problems are generally caused by data heterogeneity. The processes of data cleaning [16] and data scrubbing [22] are common terms for resolving such identity resolution problems. Elmagarmid et.al. (2007) [8] distinguish between structural and lexical heterogeneity and focus their survey on the latter. According to Elmagarmid et. al., a stage of data preparation is a necessary prerequisite to efficient record linkage and consists of a parsing, a data transformation and a standardization step. As a new challenge, RDF and OWL, alongside with the Linked Data paradigm and commonly published vocabularies, provide the means necessary to skip the data preparation step as they have already proliferated a shared structural representation of data as well as a common access mechanism. With the availability of large open data sets and links between them, the generation of benchmarks for interlinking becomes feasible and can add an edge to research in this area. The need for adaptive methods also arises in order to cope with changing data. Thus, automated reinforced approaches have to be developed that adapt themselves over time. (Both research directions are mentioned in Elmagarmid et. al.) Further reading can be found in a survey paper on Ontology Matching [18]. Although there has been extensive work on the topics of interlinking and ontology matching, the new situation creates new requirements and challenges and calls for an adaptation of existing methods.

The availability of large open data sets and accessibility via Linked Data pose the following requirements, currently insufficiently covered by research. Likewise, numerous specifics related to combined instance and schema matching in RDF and OWL w.r.t. timeliness are hardly addressed:

- ETL (Extraction, Transformation, Loading) of legacy data under the aspect of linking
- Lack of benchmarks for instance and schema mapping as well as an evaluation framework and metrics.
- As knowledge bases evolve, links and mappings have to evolve likewise. This poses special requirements on scalability and maintenance.
- Web data sources often mix terms from different RDF vocabularies and OWL ontologies. This aspect is not covered by previous work on database schema and ontology matching, which builds upon the assumption of the existence of a single schema or ontology.
- Database schemata impose harder restrictions on the instance structure compared to Semantic Web data where the open world assumption applies and information about instances is not assumed to be complete.
- The standardization of the representation format (RDF) allows the creation of links based on the availability of third-party knowledge bases from the LOD cloud such as DBpedia (similar to the star-like pattern in a mediation-based EAI).

A promising approach, which can respond to some of these requirements is to integrate schema mapping and data interlinking algorithms into a mutual refinement cycle, where results on either side (schema and data) help to improve mapping and interlinking on the other side. Both unsupervised and supervised machine learning techniques should be investigated for this task, where the latter enable knowledge base maintainers to produce high quality mappings. Further research is needed in the area of data fusion, i.e. the process of integrating multiple data items, representing the same real-world object into a single, consistent, and clean representation. The main challenge in data fusion is the reliable resolution of data conflicts, i.e. choosing a value in situations where multiple sources provide different values for the same property of an object.

The usefulness of a knowledge base increases with more (correct) links to other knowledge bases (network effect), since this allows applications to combine information from several knowledge bases. The advantage of de-referenceable URIs (URLs) as identifiers is two-fold. Contrary to a database id, URLs are unique identifiers on the web, which have a defined semantics and provenance. Furthermore, the available data identified by URLs are easily accessible via content-negotiation and standard retrieval mechanisms (i.e. the HTTP protocol). In this situation, linking brings immediate advantages, because it defines relations, e.g. equality, of Web identities and allows the convenient aggregation of data by following these links. The emerging Web of Data now faces several challenges and problems. a) How to find all links between two knowledge bases (high recall)? b) How to verify the correctness of found links (high precision)? This issue is most important in case of `owl:sameAs` links, which entail strict logical equivalence and therefore need to be very precise. c) How to maintain such a link structure with evolving knowledge bases? To solve those challenges, a constant evaluation of links between knowledge bases is necessary. Ideally, scalable machine learning techniques should be applied to generate links based on manually provided and maintained test sets. Although approaches and tools in this direction are developed, they still require higher usability and scalability to have a wider impact in the Web of Data.

After links are found and verified the next challenge is the fusing of data with respect to completeness, conciseness and consistency.

3. Improving the Structure, Semantic Richness and Quality of Linked Data

Many data sets on the current Data Web lack structure as well as rich knowledge representation and contain defects as well as inconsistencies. Methods for learning of ontology class definitions from instance data can facilitate the easy incremental and self-organizing creation and maintenance of semantically rich knowledge bases. Particularly, such methods can be employed for enriching and repairing knowledge bases on the Data Web.

The enrichment steps can be done by learning axioms, e.g. equivalence and inclusion axioms, whose left-hand side is an existing named class in the knowledge base. The task of finding such axioms can be phrased as a positive-only supervised machine learning problem, where the positive examples are the existing instances of the named class on the left hand side and the background knowledge is the knowledge base to be considered. The advantage of those methods is that they ensure that the learned schema axioms fit the instance data. The techniques should also be robust in terms of wrong class assertions in the knowledge base. We argue that those learning methods should be semi-automatic, i.e. a knowledge engineer makes the final decision whether to add one of the suggested axioms to the knowledge base. Once an axiom is added, it can be used by inference methods, for example, to populate the knowledge base with inferred facts, or spot and repair inconsistencies. One challenge is to be able to apply such machine learning methods to very large knowledge bases. Machine learning methods usually rely heavily upon reasoning techniques and currently it is not possible to reason efficiently over large knowledge bases which consist of more than 100 million RDF triples[2]. Therefore, extraction methods should be used to extract a relevant fragment of a knowledge base with respect to given individuals, which is sufficiently small to reason over it, while still containing sufficient information with respect to those instances to apply class learning algorithms. Experiments that employ such extraction methods against the DBpedia knowledge base have already shown promising results [10].

Another method for increasing the quality of Linked Data is semi-automatic repair. In particular large knowledge bases are often prone to modelling errors and problems, because their size makes it difficult to maintain them in a coherent way. These modelling problems can cause an inconsistent knowledge base, unsatisfiable classes, unexpected reasoning results, or reasoning performance drawbacks. We need algorithms, which detect such problems, order them by severity, and suggests possible methods for resolving them to the knowledge engineer. By considering only certain parts of a (large) knowledge base, those algorithms can be able to find problems in a relevant fragment, even if the overall knowledge base is not consistent. Repair approaches can be also used in combination with knowledge base enrichment algorithms: After learning a formal description of a class, problems in the knowledge base may be spotted. Those problems can then be repaired through the knowledge engineer by giving him or her possible suggestions for resolving them.

Two challenges have to be addressed in order to develop enrichment and repair methods as described above: Firstly, existing machine learning algorithms have to be extended from basic Description Logics such as ALC to expressive ones such as *SROIQ(D)* serving as the basis of *OWL 2*. Secondly, the algorithms have to be optimized for processing very large-scale knowledge bases, which usually cannot be loaded in standard OWL reasoners. In addition, we have to pursue the development of tools and algorithms for user friendly knowledge base maintenance and repair, which allow to detect and fix inconsistencies and modelling errors.

[2]Amongst others, the LarkC project (http://www.larkc.eu/) works on (incomplete) OWL reasoning.

4. Adaptive User Interfaces and Interaction Paradigms

All the different Data Web aspects heavily rely on end-user interaction: We have to empower users to formulate expressive queries for exploiting the rich structure of Linked Data. They have to be engaged in authoring and maintaining knowledge derived from heterogeneous and dispersed sources on the Data Web. For interlinking and fusing, the classification of instance data obtained from the Data Web as well as for structure and quality improvements, end users have to be enabled to effortlessly give feedback on the automatically obtained suggestions. Last but not least, user interaction has to preserve privacy, ensure provenance and, particularly in corporate environments, be regulated using access control.

The adaptive nature of the information structures in the Data Web is particularly challenging for the provisioning of easy-to-use, yet comprehensive user interfaces. On the Data Web, users are not constrained by a rigid data model, but can use any representation of information adhering to the flexible RDF data model. This can include information represented according to heterogeneous, interconnected vocabularies defined and published on the Data Web, as well as newly defined attributes and classification structures. Hence, the user interface components on the Web of Data should support the reuse of existing vocabularies and ontologies as well as the ad hoc definition of new ones and their refinement and evolution in concordance with the data over time. Different information structures should be seamlessly combinable in a provenance preserving way in a single visualization or authoring environment, even if the information to be visualized or authored is obtained or stored in various Linked Data sources ([3,21] are approaches going in that direction). The authoring tools should hide technicalities of the RDF, RDFS or OWL data models from end users, thus realizing WYSIWIG for the authoring of knowledge bases. Based on the information structures found, the most suitable authoring widgets should be automatically combined.

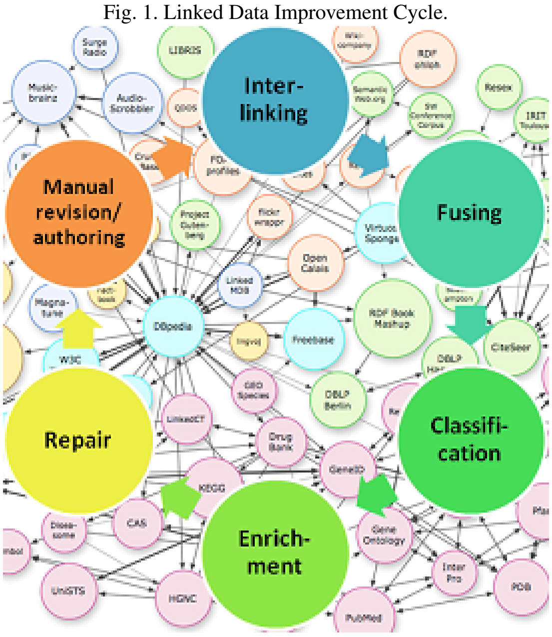
In particular for ordinary users of the Internet, Linked Data is not yet sufficiently visible and (re-) usable. Once information is published as Linked Data, authors hardly receive feedback on its use and the opportunity of realizing a network effect of mutually referring data sources is currently unused. On the social web, technologies such as *Refback*, *Trackback* or *Pingback* enabled the timely notification of authors once their posts were referenced. In fact, we consider these technologies as crucial for the success of the social web and the establishment of a network effect within the blogosphere. In order to establish a similar network effect for the Data Web, we should investigate how such notification services can be applied to the Web of Data. The Semantic Pingback method as described in [20], for example, can serve here as a technical foundation, but much more work is required to integrate such notification services in particular with adequate user interfaces into the fragmented landscape of ontology editors, triple stores and semantic wikis.

5. Making the Web a 'Washing Machine' for Linked Data

The four challenges presented in the previous sections should be tackled not in isolation, but by investigating methods which facilitate a mutual fertilization of approaches developed to solve these challenges. Examples for such mutual fertilization between approaches include:

– The detection of mappings on the schema level, for example, will directly affect instance level matching and vice versa.
– Ontology schema mismatches between knowledge bases can be compensated for by learning which concepts of one are equivalent to which concepts of the other knowledge base.
– Feedback and input from end users (e.g. regarding instance or schema level mappings) can be taken as training input (i.e. as positive or negative examples) for machine learning techniques in order to perform inductive reasoning on larger knowledge bases, whose results can again be assessed by end users for iterative refinement.
– Semantically enriched knowledge bases improve the detection of inconsistencies and modelling problems, which in turn results in benefits for interlinking, fusion, and classification.

Fig. 1. Linked Data Improvement Cycle.



– The querying performance of the RDF data management directly affects all other components and the nature of queries issued by the components affects the RDF data management.

As a result of such interdependence, we should pursue the establishment of an improvement cycle for knowledge bases on the Data Web - i.e. make the Web a Linked Data washing machine. The improvement of a knowledge base with regard to one aspect (e.g. a new alignment with another interlinking hub) triggers a number of possible further improvements (e.g. additional instance matches).

The challenge is to develop techniques, which allow to exploit these mutual fertilizations in the distributed medium Web of Data. One possibility is that, various algorithms make use of shared vocabularies for publishing results of mapping, merging, repair or enrichment steps. After one service published his new findings in one of these commonly understood vocabularies, notification mechanisms (such as Semantic Pingback [20]) can notify relevant other services (which subscribed to updates for this particular data domain) or the original data publisher, that new improvement suggestions are available. Given a proper management of provenance information, improvement suggestions can later (after acceptance by the publisher) become part of the original dataset.

6. Complementing SOA with Linked Data

Competitive advantage increasingly depends on business agility, i.e. becoming the "real-time enterprise." This entirely depends on dealing with a constant flood of information, both internal and external. Linked data is a natural addition to the existing document and web service or SOA based intranets and extranets found in large corporations. Enterprise information integration needs grow continuously. Mergers and acquisitions further drive diversity of IT infrastructure and the consequent need for integration. Simultaneously, enterprise data warehouse sizes have been more than doubling annually for the past several years, effectively outstripping Moore's law. The rapid development in the quantity and quality of structured data on the Internet creates additional opportunities and challenges. The main issues of integration are the use of different identifiers for the same thing and diversity in units of measure. Classifications, application of linked data principles for consistent use of identifiers in information interchange and making schema semantics explicit and discoverable, thus effectively rendering data self-describing, offer great promise with no disruption to infrastructure. An important distinction for the adoption of the Data Web paradigm in corporate scenarios is that the reuse of identifiers and vocabularies is not the same thing as making all data public. Corporate data intranets based on Linked Data technologies can help to reduce the data integration costs significantly and entail substantial benefits [11]. Linking internal corporate data with external references from the Data Web will allow a corporation to significantly increase the value of its corporate knowledge with relatively low effort. Examples include integration of product, customer/supplier, materials, regulatory, market research, financial statistics and other information between internal and external sources. The key to this is resolving the disparity of identifiers and associating explicit semantics to relational or XML schemes.

For example, the information integration with the Data Web will allow the sales database of a company to be enhanced with semantic descriptions of customers, products and locations by linking the internal database values with RDF descriptions from the LOD cloud (e.g. from the DBpedia, WikiCompany or Geonames data sets). The linking can be used to correct, aggregate and merge information. In addition to this, the reasoning and semantic mining tools can infer and generate new knowledge that only experts can provide. For example, using machine learning algorithms and the background knowledge from the Data Web, groups of customers can be better classified and described semantically, potentially leading to better targeting and market intelligence.

7. Publishing Public and Governmental Data on the Web

Besides employing the Linked Data paradigm in corporate environments another scenario with a great application potential is the publishing of public and governmental data (cf. [1,17]). Quite some governmental and administrative information in Europe, for example, is already publicly available in structured form. Unfortunately, this data is scattered around, uses a variety of different incompatible data formats, identifiers and schemata.

The adaptation and deployment of linked data technologies in this area will increase the ability of the public to find, download, and creatively use data sets that are generated and held by various governmental branches and institutions, be it supra-national (e.g. European) or national ones as well as regional governments and public administrations. In particular, for the case of Europe this will be very challenging due to the large organizational and linguistic diversity. This decentralization and diversity renders centralized and strictly top-down approaches less suitable and thus European governments and public administrations represent an ideal application scenario for Linked Data technologies. This scenario has been recognized [3], but should be explored much further.

In order to realize this application scenario, a number of different aspects have to be considered and different technologies should be deployed. For example, a portal as well as a network of decentralized registries should provide descriptions of the data sets (i.e. metadata), information about how to access the data sets, facilities for sharing views and reports, as well as tools that leverage government data sets. Based on research approaches tackling the above mentioned challenges, tools and services should be deployed to classify and interlink data sets automatically, to assess their information quality and to suggest enrichments and re-

[3] The following are pointers to published data sets:
http://www4.wiwiss.fu-berlin.de/eurostat/,
http://riese.joanneum.at/,
http://www.rdfabout.com/demo/census/,
http://www.govtrack.us/

pairs to the published data sets. Public participation and collaboration will be paramount in this application scenario. The public has to be engaged to provide additional downloadable data sets, to build applications, conduct analyses, and perform research. The published information can improve based on feedback, comments, and recommendations. As a result Linked Data has the potential to improve access to governmental data and expand creative use of those data beyond the walls of government by encouraging innovative ideas (e.g., mashups and web applications). The Linked Data paradigm can help to make governments more transparent and thus strengthen democracy and promote efficiency and effectiveness in governments.

8. Conclusions

While the past few years have been very succesful for the Linked Data initiative and the Web of Data, there has also been well-founded criticism [12]. As a consequence, we pointed out a number of challenges, which need to be solved in order to exploit the Web of Linked Data as medium for information integration and consumption. The four challenges center around the topics of query performance, data interlinking, data quality and user interaction. In some cases we provided future research directions to overcome these issues. We believe that the success in these research areas over the next few years is crucial for the Web of Data and its adoption by end users and enterprises.

Acknowledgement

We would like to thank the reviewers and the members of the LOD2 project consortium and in particular Orri Erling for the fruitful discussions, which contributed to this article.

References

[1] Harith Alani, David Dupplaw, John Sheridan, Kieron O'Hara, John Darlington, Nigel Shadbolt, and Carol Tullo. Unlocking the potential of public sector information with semantic web technology. In *Proc. of ISWC/ASWC2007, Busan, South Korea*, volume 4825 of *LNCS*, pages 701–714. Springer, 2007.

[2] Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache, editors. *The Semantic Web: Research and Applications, 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010, Proceedings, Part II*, volume 6089 of *Lecture Notes in Computer Science*. Springer, 2010.

[3] Sören Auer, Raphael Doehring, and Sebastian Dietzold. Less - template-based syndication and presentation of linked data. In Aroyo et al. [2], pages 211–224.

[4] Tim Berners-Lee. Linked data - design issues. web page, 2006.

[5] Christian Bizer and Andreas Schultz. The berlin sparql benchmark. *Int. J. Semantic Web Inf. Syst.*, 5(2):1–24, 2009.

[6] Roger Castillo, Christian Rothe, and Ulf Leser. Rdfmatview: Indexing rdf data for sparql queries. Technical report, Department for Computer Science, Humboldt-Universität zu Berlin, 2010.

[7] Abhirup Chatterjee and Arie Segev. Data manipulation in heterogeneous databases. *SIGMOD Record*, 20(4):64–68, 1991.

[8] Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, and Vassilios S. Verykios. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.*, 19(1):1–16, 2007.

[9] David Geer. Reducing the storage burden via data deduplication. *IEEE Computer*, 41(12):15–17, 2008.

[10] Sebastian Hellmann, Jens Lehmann, and Sören Auer. Learning of owl class descriptions on very large knowledge bases. *Int. J. Semantic Web Inf. Syst.*, 5(2):25–48, 2009.

[11] Afraz Jaffri. Linked data for the enterprise - an easy route to the semantic web. web page, March 2010.

[12] Prateek Jain, Pascal Hitzler, Peter Z. Yeh, Kunal Verma, and Amit P.Sheth. Linked data is merely more data. Technical Report SS-10-07, Menlo Park, California, 2010.

[13] Ee-Peng Lim, Jaideep Srivastava, Satya Prabhakar, and James Richardson. Entity identification in database integration. In *ICDE*, pages 294–301, 1993.

[14] Michael Martin, Jörg Unbehauen, and Sören Auer. Improving the performance of semantic web applications with sparql query caching. In Lora Aroyo, Grigoris Antoniou, Eero Hyvönen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache, editors, *ESWC (2)*, volume 6089 of *Lecture Notes in Computer Science*, pages 304–318. Springer, 2010.

[15] Axel Polleres, Aidan Hogan, Andreas Harth, and Stefan Decker. Can we ever catch up with the web? *Semantic Web - Interoperability, Usability, Applicability*, Vol. 0(1), 2010.

[16] Sunita Sarawagi. Letter from the special issue editor. *IEEE Data Eng. Bull.*, 23(4):2, 2000.

[17] John Sheridan and Jeni Tennison. Linking uk government data. In Christian Bizer, Tom Heath, Michael Hausenblas, and Tim Berners-Lee, editors, *Proceedings of the WWW2010 Workshop on Linked Data on the Web (LDOW 2010)*, 2010.

[18] Pavel Shvaiko and Jérôme Euzenat. A survey of schema-based matching approaches. *J. Data Semantics IV*, 3730:146–171, 2005.

[19] Andreas Thor. *Automatische Mapping"=Verarbeitung von Web-Daten*. Dissertation, Institut f"ur Informatik, Universit"at Leipzig, 2007.

[20] Sebastian Tramp, Philipp Frischmuth, Timofey Ermilov, and Sören Auer. Weaving a social data web with semantic pingback. In *Proceedings of the EKAW 2010 - Knowledge Engineering and Knowledge Management by the Masses; 11th October-15th October 2010 - Lisbon, Portugal*, 2010.

[21] Sebastian Tramp, Norman Heino, Sören Auer, and Philipp Frischmuth. Making the semantic data web easily writeable with rdfauthor. In Aroyo et al. [2], pages 436–440.

[22] Jennifer Widom. Research problems in data warehousing. In *CIKM*, pages 25–30. ACM, 1995.