

SEMANTIC TECHNOLOGY INSTITUTE (STI)



# A SURVEY ON DATA INTERLINKING METHODS

Stephan Wölger      Katharina Siorpaes  
Tobias Bürger      Elena Simperl      Stefan Thaler  
Christian Hofer

STI TECHNICAL REPORT 2011-03-31

MARCH 2011

SEMANTIC TECHNOLOGY INSTITUTE (STI)

**STI Innsbruck**  
University of Innsbruck  
Technikerstrasse 21a  
Innsbruck, Austria  
[www.sti-innsbruck.at](http://www.sti-innsbruck.at)





STI TECHNICAL REPORT  
STI TECHNICAL REPORT 2011-03-31, MARCH 2011

## A SURVEY ON DATA INTERLINKING METHODS

Stephan Wölger<sup>1</sup>, Katharina Siorpaes<sup>1</sup>, Tobias Bürger<sup>2</sup>, Elena Simperl<sup>3</sup>,  
Stefan Thaler<sup>1</sup>, Christian Hofer<sup>1</sup>,

**Abstract.** As Linked Open Data promises to enrich a user's web experience in various ways like having computers read information that is shared by users or by being able to query resources on the web in an advanced way, research puts more and more effort into the development of linking tools. In this paper we summarize several linking tools regarding their most important aspects by applying a template to each of them. This includes a general description for each of the tools as well as technical aspects. To provide a comprehensive overview we also compare the tools using a table with important aspects. Furthermore we add additional information regarding linking techniques for completeness.

**Keywords:** interlinking, semantic web, ontologies, ontology merging, ontology alignment, ontology mediation, ontology matching.

---

<sup>1</sup>Semantic Technology Institute (STI) Innsbruck, University of Innsbruck,  
Technikerstraße 21a, A-6020 Innsbruck, Austria.

E-mail: {stephan.woelger, katharina.siorpaes, stefan.thaler, christian.hofer}@sti2.at

<sup>2</sup>Capgemini, Carl-Wery-Str. 42, D-81739 Munich, Germany. E-mail: tobias@tobiasbuerger.com

<sup>3</sup>Karlsruhe Institute of Technology,

B. 11.40 KIT-Campus Süd D-76128 Karlsruhe, Germany. E-mail: elena.simperl@kit.edu

# 1 Introduction

In 2007 the Linking Open Data (LOD) community project<sup>1</sup> started an initiative which aims at increased use of Semantic Web applications. Such applications on the one hand provide new means to enrich a user's web experience but on the other hand also require certain standards to be adhered to. Two important requirements when it comes to Semantic Web applications are the availability of RDF datasets on the web and having typed links between these datasets in order to be able to browse the data and to jump between them in various directions.

While there exist tools that create RDF output automatically from the application level and tools that create RDF from web sites, interlinking the resulting datasets is still a task that can be cumbersome for humans (either because there is a lack of incentives or due the non availability of user friendly tools) or not doable for machines (due to the manifoldness of domains). Despite the fact that there are more and more interlinking tools available, those either can be applied only for certain domains of the real world (e.g. publications) or they can be used just for interlinking a specific type of data (e.g. multimedia data).

In this paper we want to provide a comprehensive overview of several linking tools with respect to aspects like required input, resulting output, considered domain, used matching techniques and others. We also want to compare the tools among each other and summarize the obvious drawbacks and issues to be taken care of in future research.

The rest of this paper is structured as follows:

In section 2 we summarize different linking tools by applying a specific template to them to provide a comprehensive overview for each of the tools. At the end of section 2 we furthermore compare the tools in terms of the most important aspects of linking tools in comparison tables.

Section 3 then discusses different aspects related to the field of interlinking for completeness before concluding in section 4.

---

<sup>1</sup><http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

## 2 Linking Tools

In this section we provide the reader with an overview of several linking tools by applying a specific template to each of them. Such an overview includes a short description of the tool under consideration as well as the most important aspects of linking tools, evaluated for each of them. The template looks as follows:

**description** The description is a short paragraph containing name, purpose and general information about a tool. If available we also provide a reference to a paper, book chapter or similar containing detailed information about the tool.

**degree of automation** This part focuses on the degree of automation which is about whether a tool needs human input as well as about the type of and requirements to the input.

**human contribution** Human contribution refers to the way a user has to contribute in order to be able to use the tool and to what extend a user has to contribute.

**domain** Some tools are tied to a specific domain, others are domain independent.

**used matching techniques** This part is about the mechanisms a tool uses to match data. Such mechanisms e.g. can be string matching, semantic indexing and so on.

**ontologies** Here we want to know whether the tool takes ontologies associated to the dataset into account or whether the tool allows to interlink datasets described according to different ontologies. In case the ontologies differ from each other, possible ontology alignment support can be of interest.

**input** Which input does the tool require and in which format?

**output** What kind of output does the tool produce (owl:sameAs links, VoiD linkset, other types of links)? Does the tool propose to merge the input datasets?

**post-processing** Does the tool provide any post-processing capabilities (consistency checking and inconsistency resolution)?

**data access** How can the data be accessed?

**use cases** In which experiments/use cases was the tool used and is there any remarkable contribution to existing datasets?

## 2.1 RDF-AI

**description** RDF-AI[14] focuses on the integration of RDF datasets. Providing modules for pre-processing, matching, fusing, interlinking and post-processing, RDF-AI provides different outputs depending on the module creating the output. An alignment of two input datasets can be used to merge those datasets or to create a linkset. In addition to the datasets under consideration the tool requires parameters for performing various algorithms in the different modules. Due to this the user can influence the output of RDF-AI to fit his specific requirements.

**degree of automation** semi

**human contribution** provide two datasets and parameters to be considered by the tool

**domain** domain independent

**used matching techniques** string matching, word relation matching (synonyms, taxonomical similarity)

**ontologies** multi

**input** two datasets and parameters

**output** merged dataset or entity correspondence list (owl:sameAs triples)

**post-processing** consistency checking

**data access** Jena framework

**use cases** System tested with following datasets: AKT EPrints archive and Rexa datasets and the works of Johann Sebastian Bach in two different datasets

## 2.2 Silk

**description** Silk[17] is a link discovery framework using a declarative language for searching relationships between various datasets. This language on the one hand can be used for specifying the type of RDF links to be searched for. On the other hand the user can specify conditions that must hold in order to interlink two entities.

**degree of automation** semi

**human contribution** specification of resources/link type, specification of conditions/comparison technique, (certain transformations to increase efficiency)

**domain** domain independent

**used matching techniques** string matching, similarity metrics (numerical, dates, concept distance, sets)

**ontologies** n/a

**input** link specific alignment method (Silk-LSL in XML)

**output** linkset (owl:sameAs, RDF predicates, reified format)

**post-processing** no

**data access** SPARQL

**use cases** Linking of DBpedia to LinkedMDB, GeoNames to DBpedia and DrugBank to DBpedia

### 2.3 Knofuss

**description** Knofuss[9] is designed to enable data-level integration of semantic annotations. The tool produces merge sets of two RDF knowledge base resources by comparing them with respect to a specific ontology. This can be an ontology describing two resources of the same type or alternatively, in case the two resources are described by heterogeneous ontologies, an ontology alignment format can be used. In addition Knofuss provides means to verify the consistency of the resulting dataset.

**degree of automation** automatic

**human contribution** provide input

**domain** domain independent

**used matching techniques** string matching, adaptive learning

**ontologies** single

**input** source RDF knowledge base, target knowledge base, fusion ontology (OWL)

**output** set of statements, merge sets

**post-processing** inconsistency resolution

**data access** local copies

**use cases** System tested with following datasets (domain of scientific publications): AKT EPrints archive, Rexa dataset and SWETO DBLP dataset

## 2.4 LinQuer

**description** LinQuer (Linkage Query Writer)[3] is a framework for generating SQL queries used to semantically discover links between relational datasets. The framework includes the LinQL (a language used to specify requirements to be fulfilled), a web interface and an API which can be used to transform LinQL queries into SQL queries as well as an interface making it easier to write LinQL queries. Due to the fact that LinQuer is built modularly and generic it can be easily enriched with user defined link discovery algorithms.

The problem LinQuer addresses is the discovery of links between entities that represent the same or semantically related objects of the real world. In addition or in combination to/with that the tool also provides syntactical methods to find links between two datasets.

**degree of automation** The user has to define LinQL queries and hence has to do some manual work. The queries can be built on top of native link discovery methods or on custom ones.

**human contribution** The user has to provide LinQL queries (set of source data, set of target data, set of methods to use).

**domain** domain independent

**used matching techniques** The framework provides synonym, hyponym, and several string matching methods. Furthermore it enables the use of link clause expressions, which are boolean combinations of link clauses.

**ontologies** In case of using semantic link discovery methods the tool takes ontologies into account however there are no means for ontology alignment.

**input** The only restriction is that the input must be available in a SQL database.

**output** The tool converts the LinQL queries to SQL queries and hence extends the database tables with entries containing the generated links.

**post-processing** no

**data access** SQL queries

**use cases** LinQuer was showcased in the context of the Linked Clinical Trials (Linked CT) project. The framework was used to find links between real clinical trial data and several other data sources including patient data and Wikipedia articles.

## 2.5 Interlinking Distributed Social Graphs

**description** Interlinking Distributed Social Graphs[12] is a method for interlinking user profiles from different social platforms like Facebook, MySpace and Twitter. The idea is to generate RDF graphs (mapping of XML

schemas from platforms to RDF) which then can be interlinked based on the user identifiers of each of the platforms under consideration. The linking is based on graph matching and the presented tool uses different graph matching algorithms to derive similarity measures and to create owl:sameAs links.

**degree of automation** It is completely automatic (black box)

**human contribution** no human contribution required

**domain** The tool is built for matching Facebook, MySpace and Twitter user profiles. However, the tool probably can be extended to include other services if modified manually.

**used matching techniques** The tool is based on graph matching algorithms. The paper presents three matching techniques: Node/Edge Overlap, Node Mapping and Graph Reasoning (incl. Levensthein string metric for string matching).

**ontologies** The tool extracts XML data from social sites/services and maps them to the FOAF ontology. For geographical information the Geonames ontology and the Geolocations vocabulary are used.

**input** The data is extracted from the web in XML and then mapped to FOAF.

**output** The tool produces owl:sameAs links to interlink the user profiles in the social RDF graphs. It does not really merge input sets, but rather creates graphs from the XML data from different sites and then interlinks them. The graph then points to the different resources from the social services/platforms and does not include the whole data from the different profiles (privacy reasons).

**post-processing** no

**data access** RDF

**use cases** The tool was tested within a small experiment, no large use case yet.

## 2.6 Guess What?

**description** Guess What?![8] is a semantic game with a purpose. It focuses on creating ontologies by presenting users with concept fragments from Linked Open Data in order to have the users guessing a described concept. In addition each user can judge on the concept choice of an opponent in order to reach consensus on the final concept to use. Starting point is a seed concept which serves as basis for crawling DBpedia, Freebase and OpenCyc for concept fragments. From then on the game continues round based. Each round provides a new concept for the ontology to be built.

**degree of automation** The tool automatically retrieves information fragments for a seed concept from DBpedia, Freebase and OpenCyce. Users then have to provide a concept which is described by the fragments and evaluate the concepts given by other players.

**human contribution** Human contribution is required in terms of providing concepts which match the given fragments and in terms of evaluating the opponents' concepts.

**domain** The game uses URIs from DBpedia, Freebase and OpenCyce.

**used matching techniques** Natural language processing and graph mining. The final decisions are human guessing and evaluating.

**ontologies** n/a

**input** RDF from external data sources

**output** The paper does not specify the output format, however there is an example how the output could be formalized using owl.

**post-processing** no

**data access** SPARQL endpoint

**use cases** The paper mentions a small experiment, no large use-case.

## 2.7 Pool party

**description** PoolParty[13] is a thesaurus management tool which allows to create/enrich thesauri with concepts discovered on the one hand by analysing documents and on the other hand by gathering additional information from linked data sources (LOD). Due to the fact that domain experts very often are not semantic web experts at the same time, PoolParty focuses on keeping access barriers to contribution rather low. This is achieved by means of wiki-style interfaces.

**degree of automation** The tool allows to take the labels from the taxonomy as search parameters for connecting them to LOD data (e.g. DBpedia). It returns different concepts from which the user must select the right one.

**human contribution** 70%

**domain** The tool is restricted to constructing/maintaining thesauri using SKOS.

**used matching techniques** Probably string matching.

**ontologies** You can link your thesaurus concepts to other ontologies (e.g. DBpedia). The tool retrieves the matching candidates (probably based on string matching) and the user has to select the right ones.

**input** RDF triples or one can use the interface to create a thesaurus, (geographical coordinates).

**output** RDF triples with owl:sameAs relations, HTML Wiki enriched with RDFa

**post-processing** No

**data access** By means of the tool's web interface. Probably also directly via a SPARQL endpoint.

**use cases** KiWi and LASSO project

## 2.8 GUMF

**description** Grapple User Modeling Framework (GUMPF)[7] is a framework focusing on the linking of user profiles and user models. By aggregating, contextualizing and modelling user data, GUMF increases interoperability by eliminating the issue of schema mapping which may arise when incorporating data from heterogeneous sources. With GUMF's plug-in GDR the framework enables furthermore the use of reasoning capabilities for deducing implicitly given information about users from their profiles.

**degree of automation** The user has to provide Grapple Derivation Rules (GDR) in order to interlink datasets. The rules consist of premises and consequences. If the rules are applied, the GDR engine constructs the relationships between the datasets.

**human contribution** Grapple Derivation Rules

**domain** No restriction, however, data must be in RDF.

**used matching techniques** manual (creation of GDRs)

**ontologies** n/a

**input** Datasets in RDF (Sesame repository)

**output** RDF

**post-processing** n/a

**data access** SPARQL endpoint

**use cases** The extended GUMF was used in the e-learning domain in the context of the GRAPPLE project.

## 2.9 CaMiCatzee

**description** CaMiCatzee[4] is a multimedia data interlinking concept demonstrator using the UCI methodology. By using flickr's annotation mechanisms used by humans CaMiCatzee links flickr images to FOAF data, related topics and locations.

**degree of automation** automatic

**human contribution** The user has to provide a FOAF document, a person's URI or simply a name.

**domain** multimedia data

**used matching techniques** semantic indexer

**ontologies** n/a

**input** FOAF document, a person's URI or simply a name

**output** rdfs:seeAlso in XHTML+RDFa

**post-processing** n/a

**data access** user input, FOAF access

**use cases** n/a

## 2.10 ExpLOD

**description** ExpLOD[6] is a tool focusing on the exploration of RDF usage summaries. Using bisimulation contractions applied to RDF graphs in a hierarchical manner, ExpLOD provides means for summarizing RDF usage at different granularities. In addition to that the tool also provides interlinking capabilities.

**degree of automation** semi

**human contribution** provide a dataset, a bisimulation label scheme and neighbourhoods to be considered

**domain** any

**used matching techniques** labeled bisimulation contractions, intersection-automaton creation

**ontologies** n/a

**input** a dataset, a bisimulation label scheme, neighbourhoods to be considered

**output** RDF summaries

**post-processing** n/a

**data access** RDF

**use cases** n/a

## 2.11 GNAT

**description** GNAT[11] is a tool enabling to link personal music collections to the Musicbrainz dataset. The tool focuses on the use of metadata, however in case that the metadata is of low quality, audio fingerprinting can be used to match the music datasets. GNAT produces an RDF file containing `mo:available_as` links which can serve the user as starting point for exploring the internet regarding (music) data.

**degree of automation** automatic

**human contribution** none

**domain** music datasets

**used matching techniques** audio fingerprinting, metadata

**ontologies** n/a

**input** personal audio collection, Musicbrainz dataset

**output** RDF statements (`owl:sameAs`, `mo:available_as`)

**post-processing** n/a

**data access** n/a

**use cases** GNAT results were used as input for the GNARQL programm experiment

## 2.12 Comparison table

Tool	Automation	Human Contribution	Domain	Matching methods	Ontologies	Input	Output	Post processing	Data access
RDF-AI	semi	provide two datasets and parameters to be considered by the tool	any	string matching, word relation matching (synonyms, taxonomical similarity)	multi	two datasets and parameters	merged dataset or entity correspondence list (owl:sameAs triples)	consistency checking	Jena framework
Silk	semi	specification of resources/link type, specification of conditions/comparison technique, (certain transformations to increase efficiency)	any	string matching, similarity metrics (numerical, dates, concept distance, sets)	n/a	link specific alignment method (Silk-LSL in XML)	linkset (owl:sameAs, RDF predicates, reified format)	no	SPARQL

Table 1: Comparison table

Tool	Auto- mation	Human Con- tribution	Domain	Matching methods	Onto- logies	Input	Output	Post process- ing	Data ac- cess
Knofuss	automatic	provide input	any	string match- ing, adaptive learning	single	source RDF knowl- edge base, target knowl- edge base, fusion ontology (OWL)	set of state- ments, merge sets	in- consistency resolu- tion	local copies
LinQuer	semi	LinQL queries (set of source data, target data, methods to use)	any (SQL access)	synonym, hyponym and string matching, link clause expressions	multi, no align- ment	SQL database	SQL database entries	no	SQL queries
Interlinking Dis- tributed Social Graphs	automatic	none	Facebook, MySpace and Twit- ter profiles	graph match- ing, string matching	multi	XML	owl:sameAs	no	RDF
Guess What!?	semi	provide match- ing concepts, evaluation of opponents' concepts	use of URIs from DBpedia, Freebase and Open- Cyc	natural language processing, graph min- ing, human guessing and evaluation	n/a	RDF	could be formalized to OWL	no	SPARQL endpoint

Table 2: Comparison table

Tool	Automation	Human Contribution	Domain	Matching methods	Ontologies	Input	Output	Post processing	Data access
Pool Party	semi	70%	thesauri using SKOS	probably string matching	multi	RDF triples, web interface activity	RDF triples with owl:sameAs relations, HTML Wiki enriched with RDFa	no	tool's web interface
GUMF	semi	Grapple Derivation Rules	any	manual matching (creation of GDRs)	n/a	RDF (Sesame repository)	RDF	n/a	SPARQL endpoint
CaMi-Catzee	automatic	provide a FOAF document, a person's URI or simply a name	multimedia data	semantic indexer	n/a	FOAF document, a person's URI or simply a name	rdfs:seeAlso in XHTML + RDFa	n/a	user input, FOAF access
ExpLOD	semi	provide a dataset, a bisimulation label scheme and neighbourhoods to be considered	any	labeled bisimulation contractions, intersection-automaton creation	n/a	a dataset, a bisimulation label scheme, neighbourhoods to be considered	RDF summaries	n/a	RDF
GNAT	automatic	none	music datasets	audio fingerprinting, metadata	n/a	personal audio collection, Musicbrainz	RDF statements (owl:sameAs, mo:available_as)	n/a	n/a

Table 3: Comparison table

### 3 Related Aspects

The following section provides the reader with information related to the field of interlinking which includes e.g. different approaches for interlinking data. Here we do not focus on specific tools as in the previous section but for completeness we include aspects related to the field of interlinking.

#### 3.1 Manual Interlinking

*User Contributed Interlinking (UCI)*[2][5] defines a methodology for interlinking datasets based on information provided by a user. The advantage of this methodology used to interlink multimedia is that in general information provided by a user is qualitatively rather high compared to other approaches of interlinking. Yet there is the fact that manual contribution to interlinking data can be rather high in terms of effort. This is why UCI should be combined with *Game Based Interlinking (GBI)* approaches in order to provide additional incentives and benefit. UCI is implemented for example in the tool CaMiCatzee[4] and Henry.

#### 3.2 Game Based Interlinking

Due to the fact that manual interlinking can be very high in terms of effort UCI should be combined with GBI. The point here is to provide incentives for users to interlink data by playing games. An example for using this approach is Games With A Purpose (GWAP)[18]. Different games gather information about provided pictures by asking the user to annotate images or by asking the user to trace objects in images. Similarly OntoGame[15] provides incentives for users to interlink data without focusing on this specific task but having actually fun playing the games. SeaFish[16] is one of these games where the user has to separate flickr images which have the same name but represent different concepts of the real world.

#### 3.3 Semi-automatic Interlinking (of Multimedia)

Semi-automatic interlinking[5] is an approach which combines analysis techniques and human judgement. The first step is (in case we have multimedia data) to analyse multimedia data and corresponding metadata like user profiles. Based on the result human users have to make a choice which suggestions should be accepted, rejected or modified.

*Emergent Interlinking (EI)*[2][5] is about observing how multimedia data is used by using context-rich environments enabling to analyse user behaviour as well as the users themselves. This includes browsing paths, structured content placement on a web site and other types of behaviour.

### 3.4 Collaborative Interlinking

Based on semantic wikis interlinking can be carried out collaboratively. This can improve the quality of the interlinking results and lowers the effort of an interlinking project for a single member of the wiki. Such wikis are for example Ylvi, MultiMakna and Meta- VidWiki (MVW).

### 3.5 Automatic Interlinking (of Multimedia)

If there is metadata of a high quality available during a linking project, *Automatic Interlinking (AI)*[2][5] can provide good results when it comes to analysing global textual resources, however the more fine-grained the resources are the less accurate the results can become. This approach is used for example in Henry, a tool which analyses audio signals to segment them and in the following interlink them with concepts related to music.

### 3.6 Weaving SIOC into the Web of Linked Data

SIOC (Semantically Interlinked Online Communities)[1] refers to interlinking data from social media pages like blogs, forums and wikis, however SIOC is not a linking tool but a vocabulary to transform the mentioned media into RDF. There is a whole implementation list containing approximately 35 tools that are capable of creating SIOC data. The extracted metadata then can be linked with other datasets from the Linked Open Data. This can be done in both directions however it is up to the user to interlink e.g. a FOAF profile to a SIOC dataset via a `rdfs:SeeAlso` link or the other way round a SIOC dataset to a FOAF profile using an `owl:sameAs` or `rdfs:seeAlso` link.

### 3.7 Additional Tools

In addition to the tools discussed in section 2 we found interlinking tools that we did not include in section 2 since at the time writing this paper we did not find publications at all about them or just publications containing not sufficient information to apply our template. Yet, for completeness, we would like to list them here:

- RKB-CRS: Co-reference Resolution Service (CRS) is about finding URIs that seem to be equivalent in first place but are different actually. Such URIs just refer to the same concept of the real world. The service aims at enabling the management of URI co-reference lists in order to enable interoperability of various sources of Linked Open Data. The RKB-Explorer (<http://www.rkbexplorer.com/explorer/>) uses CRS for storing, manipulating, and reusing co-reference information. Yet there are many tools available not coping with the problem of URI co-reference.
- Henry[10] (<http://dbtune.org/henry/>) is a Music-related Semantic Web agent enabling on-demand extraction and interlinking.

- LD-Mapper (<http://motools.svn.sourceforge.net/viewvc/motools/dbtune/ldmapper/>) is a script programmed in Prolog which implements an interlinking algorithm focusing on Jamendo (<http://www.jamendo.com/de/>) and Musicbrainz (<http://musicbrainz.org/>).
- ODD-Linker (<http://melinda.inrialpes.fr/systems.html>)

## 4 Conclusion

Recently, research has put more and more effort into the development of linking tools in order to enrich a user's web experience in various ways. This focus on linking tools can be explained by the necessity of having tools that automatically interlink datasets or at least decrease the effort for manual interlinking either by semi-automatically interlink datasets or by providing incentives for users to interlink data while actually considering this task not so much as effort but rather as fun.

Despite the fact that there are more and more such tools available we recognized several drawbacks with these tools that have to be taken care of in future research.

In this paper we summarized several linking tools regarding their most important aspects by applying a template to each of them. This included a general description for each of the tools as well as technical aspects. To provide a comprehensive overview we also compared the tools using a table with important aspects. Furthermore we added additional information regarding linking techniques for completeness.

While some of the tools interlink data fully automatically and provide convenient (RDF, SPARQL) means to access the results, others need human input either by creating links manually, by having the user prepare the input data in a way it can be used by the tool under consideration or by even requiring dataset specific programs written by the user.

Most of these tools hardly provide incentives (additional to the incentive to get interlinked data) for users to do this sometimes cumbersome tasks. Game Based Interlinking seems to be a promising approach for coping with this issue. While a user actually plays a game the tool uses the gamer's input for interlinking data in the background.

Another issue regarding the tools under consideration is that some of them can be applied only for certain domains of the real world (e.g. publications) or that they can be used just for interlinking a specific type of data (e.g. multimedia data). There is no tool that can be used as general purpose interlinking tool. This is due to the fact that different domains require different types of links to interconnect related datasets or to link to datasets of other domains and to the fact that interlinking includes the task of matching concepts which is different depending on the type of data. While e.g. publications rather use various string matching techniques, multimedia data requires audio fingerprinting and metadata analysis. Other types of data require similarity propagation or semantic indexing.

On the contrary having a powerful tool providing interlinking capabilities for any kind of domain and type of data can be counterproductive in terms of introducing complexity of use to the human being having to use the different means to interlink data and having to make decisions based on the results provided by the tool. Even if a tool fully automatically interlinks datasets in terms of selecting appropriate matching techniques or selecting the types of links, most

of the results of such a tool would need human review since certain domains are too manifold for a machine to make the mentioned decisions. That is why a linking tool always has to be a compromise between fully automatic and manual interlinking, between domain independent and domain specific requirements and between expressive power in terms of available types of links and matching algorithms and usability.

## References

- [1] Uldis Bojars, Alexandre Passant, Richard Cyganiak, and John Breslin. Weaving SIOC into the Web of Linked Data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, April 2008.
- [2] Tobias Bürger and Michael Hausenblas. Interlinking multimedia - principles and requirements.
- [3] Oktie Hassanzadeh, Reynold Xin, Renée J. Miller, Anastasios Kementsietsidis, Lipyeow Lim, and Min Wang. Linkage query writer. *PVLDB*, 2(2):1590–1593, 2009.
- [4] M. Hausenblas and W. Halb. Interlinking Multimedia Data. In *Linking Open Data Triplification Challenge at the International Conference on Semantic Systems (I-Semantics08)*, 2008.
- [5] M. Hausenblas, R. Troncy, Y. Raimond, and T. Bürger. Interlinking Multimedia: How to Apply Linked Data Principles to Multimedia Fragments. In *WWW 2009 Workshop: Linked Data on the Web (LDOW2009)*, Madrid, Spain, 2009.
- [6] Shahan Khatchadourian and Mariano Consens. Explod: Summary-based exploration of interlinking and rdf usage in the linked open data cloud. *The Semantic Web: Research and Applications*, pages 272–287, 2010.
- [7] Erwin Leonardi, Fabian Abel, Dominik Heckmann, Eelco Herder, Jan Hidders, and Geert-Jan Houben. A flexible rule-based method for interlinking, integrating, and enriching user data. In Boualem Benatallah, Fabio Casati, Gerti Kappel, and Gustavo Rossi, editors, *ICWE*, volume 6189 of *Lecture Notes in Computer Science*, pages 322–336. Springer, 2010.
- [8] Thomas Markotschi and Johanna Völker. GuessWhat?! - Human Intelligence for Mining Linked Data. In *Proceedings of the Workshop on Knowledge Injection into and Extraction from Linked Data (KIELD) at the International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, 2010.
- [9] Andriy Nikolov, Victoria Uren, Enrico Motta, and Anne de Roeck. Integration of semantically annotated data by the KnoFuss architecture. In *16th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2008)*, Acitrezza, Italy, 2008.
- [10] Yves Raimond and Mark B. Sandler. A web of musical information. In Juan Pablo Bello, Elaine Chew, and Douglas Turnbull, editors, *ISMIR*, pages 263–268, 2008.
- [11] Yves Raimond, Christopher Sutton, and Mark Sandler. Automatic interlinking of music datasets on the semantic web, 2008.

- [12] Matthew Rowe. Interlinking Distributed Social Graphs. In *WWW 2009 Workshop: Linked Data on the Web (LDOW2009)*, 2009.
- [13] Thomas Schandl and Andreas Blumauer. Poolparty: Skos thesaurus management utilizing linked data. 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 421–425. Springer, 2010.
- [14] François Scharffe, Yanbin Liu, and Chunguang Zhou. RDF-AI: an Architecture for RDF Datasets Matching, Fusion and Interlink. In *Proceedings of the IJCAI 2009 workshop on Identity, reference, and knowledge representation (IR-KR), Pasadena, CA US*, 2009.
- [15] Katharina Siorpaes and Martin Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23(3):50–60, 2008.
- [16] Stefan Thaler, Katharina Siorpaes, David Mear, Elena Simperl, and Carl Goodman. Seafish: A game for collaborative and visual image annotation and interlinking. In Grigoris Antoniou, Marko Grobelnik, Elena Simperl, Bijan Parsia, Dimitris Plexousakis, Jeff Pan, and Pieter De Leenheer, editors, *To be published in ESWC 2011*, 2011.
- [17] Julius Volz, Christian Bizer, Martin Gaedke, and Georgi Kobilarov. Silk - A Link Discovery Framework for the Web of Data . In *2nd Workshop about Linked Data on the Web (LDOW2009)*, Madrid, Spain, 2009.
- [18] L. von Ahn. Games with a Purpose. *Computer*, 39(6):92–94, June 2006.