

Revyu: Linking Reviews and Ratings into the Web of Data

Tom Heath^{†‡} and Enrico Motta[‡]

† Talis Information Limited
Knights Court, Solihull Parkway
Birmingham, B37 7YB, United Kingdom

‡ Knowledge Media Institute, The Open University,
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

tom.heath@talis.com, e.motta@open.ac.uk

Fax: +44 (0)870 400 5001

Corresponding Author: Tom Heath

For Submission to: Journal of Web Semantics, Semantic Web Challenge Special Issue

Revyu: Linking Reviews and Ratings into the Web of Data

Tom Heath and Enrico Motta

Keywords: Semantic Web; Linked Data; Reviews; Ratings; Revyu

Abstract

Revyu is a live, publicly accessible reviewing and rating Web site, designed to be usable by humans whilst transparently generating machine-readable RDF metadata for the Semantic Web, based on user input. The site uses Semantic Web specifications such as RDF and SPARQL, and the latest Linked Data best practices to create a major node in a potentially Web-wide ecosystem of reviews and related data. Throughout the implementation of Revyu design decisions have been made that aim to minimize the burden on users, by maximizing the reuse of external data sources, and allowing less structured human input (in the form of Web 2.0-style tagging) from which stronger semantics can later be derived. Links to external sources such as DBpedia are exploited to create human-oriented mashups at the HTML level, whilst links are also made in RDF to ensure Revyu plays a first class role in the blossoming Web of Data. In this paper we document design decisions made during the implementation of Revyu, discuss the techniques used for linking Revyu data with external sources, and outline how data from the site is being used to infer the trustworthiness of reviewers as sources of information and recommendations.

1. Introduction

Reviews and ratings are widely available on the Web and are one major form of 'user-generated content' that has become associated with 'Web 2.0' [25]. However, despite the availability of reviews and ratings through APIs such as the Amazon Associates Web Service [2], this data remains largely isolated in 'silos', and described in formats that hinder its integration and interlinking with data from other sources. This presents considerable barriers to the aggregation of all reviews of a particular item from across the Web, as an item reviewed in one silo cannot easily be associated with the same item reviewed elsewhere. As has been recognised by previous authors [15, 16], the Semantic Web, or Web of Data, provides a technological platform with which to overcome this problem and Revyu is a significant and concrete step towards realising a solution.

Revyu is a live, publicly usable and well used reviewing and rating Web site, launched in November 2006 and available at <http://revyu.com/>. The site combines an approachable interface for the creation of reviews by human users with a range of APIs through which Semantic Web applications can access machine-readable data for reuse in third-party applications. The site has been developed using Semantic Web technologies and standards such as RDF [19] and SPARQL [26], and according to Linked Data principles [5] and best practices [7]. These features enable Revyu to readily consume data from external services for the creation of human-oriented

'mashups' while also seeding an ecosystem of interlinked review and rating data on the Web that is helping to bootstrap the Semantic Web as a whole. In the following sections we will describe Revyu in more detail, examine these human- and machine-oriented characteristics and discuss many of the underlying design decisions.

2. Revyu compared to conventional Web APIs

Revyu allows people to review and rate things simply by filling in a Web form. This style of interaction with the site will be familiar to those who have written reviews on sites such as Epinions [12], Amazon [1] or TripAdvisor [28]. Whilst this functionality is not especially novel, as a reviewing application Revyu improves significantly over other work in the area in a number of ways.

Firstly, Revyu is not a data 'silo' that locks data away for 'safe keeping'. Instead, reviews and ratings created within Revyu are exposed in a reusable, machine-readable data format, RDF. This contrasts with sites that collect data from users but only republish it in HTML, thereby masking the structure created by the author of the review and no doubt still represented in the underlying database. The RDF data model represents a more flexible mechanism for publishing structured data than approaches such as *microformats*¹, which are not supported by a common underlying data model and consequently present greater challenges for those wishing to parse and consume published data.

Secondly, the data access mechanisms provided by Revyu improve upon the APIs of sites such as Amazon. Publishing data in RDF allows for easier merging of data from disparate sources, as heterogeneous data can be combined in one document without the document as a whole needing to conform to a single schema. Golbeck's *FilmTrust* [15] is noteworthy as one of the first applications to make review data available on the Web in RDF. However it is limited to the domain of films and does not provide a query interface to the underlying data via languages such as SPARQL. The Revyu SPARQL endpoint provides developers with greater flexibility in querying the underlying data set than is generally possible with conventional Web APIs.

Thirdly, Revyu takes a Linked Data [5] approach to publishing reviews and ratings on the Web. The technical aspects of Revyu as a Linked Data application will be described in detail in later sections – at this stage it is sufficient to outline the benefits of the approach:

1. All entities within the Revyu site are addressable over the Web, allowing these to be referenced from other online data sets. In the context of Revyu, third parties may use this capability to indicate (dis-)agreement with a particular review, or where there may be a conflict of interest that compromises the credibility of a particular reviewer (as reported in [31]) this can be highlighted. Crucially, this additional information can be published elsewhere on the Web in RDF, simply referencing the appropriate items on Revyu. This avoids the creation of a single silo in which all information must be located.

¹ <http://microformats.org/>

2. Duplication of data across data providers is reduced. There is less need for Revyu to maintain local copies of data about reviewed items, and incur the associated data management overhead, as RDF statements can be used to connect review data with richer item descriptions in other locations on the Web. This has the effect of encouraging data to remain published and managed by the authoritative source.
3. Data integration can be performed once, and reused many times. A data publisher need only expend the effort once to link entities in their data with those in external data sets, and publish these links as RDF statements for consumption by third parties. This removes the need for each data consumer to perform their own data integration, the results of which may be locked within application code used to create 'mashups' and therefore not available for reuse by others.

Lastly, Revyu takes an open world perspective on the reviewing process by not constraining users to reviewing items from a fixed and pre-selected database. Anything a user can name can be reviewed, and consequently reviewers are not restricted to reviews and ratings in one domain. By supplying links related to the item they are reviewing, users enable disambiguation and linking of reviewed items through inverse functional properties such as `foaf:homepage`.

3. A Usable Semantic Web System

A major goal of Revyu was to create a Semantic Web application that could be used by non-specialist users, i.e. those with no experience or knowledge of Semantic Web technologies. This was achieved by making the creation and publication of RDF invisible to the reviewer, enabling users to contribute data to the Semantic Web through a familiar, Web 2.0-style mode of interaction. The Revyu home page is shown in Figure 1.

Figure 1.

As of May 2008 Revyu has attracted 837 reviews from 261 distinct reviewers. This corresponds to over 35,000 RDF triples publicly available on the Semantic Web. Whilst not a large figure by many standards, it is significant that many of these triples have been generated from direct user input, rather than by data mining, extraction from natural language, or conversion of existing databases.

3.1 Tagging in Revyu

When creating Revyu, a significant decision was taken to not require users to classify the item they were reviewing, but instead to associate keyword tags with the item. This decision was taken for several reasons: firstly there was not seen to be a sufficiently comprehensive classification available of items that users may want to

review; secondly, requiring all users to subscribe to a single classification scheme for reviewed items seemed unnecessarily constraining and against the open-world spirit of the Semantic Web; thirdly, providing a usable interface through which non-specialists could classify items using arbitrary types discovered in ontologies on the Semantic Web was seen as unfeasible; and lastly, even the coverage provided by multiple ontologies readily available on the Web was deemed insufficient to describe all items that might be reviewed, therefore potentially resulting in a more closed world of reviewed items.

Despite the availability of large-scale ontologies such as Yago [27], and their integration with DBpedia [3], we believe that these issues remain unresolved. For example, the Yago class that should be used to describe a relatively simple concept such as 'movie' remains unclear, and users cannot be expected to engage in such issues during review creation.

We believe that tagging retains the appropriate balance of usability whilst also providing data from which stronger semantics can be derived. At present we use tagging data in two ways: to identify basic semantic relationships between tags, and to derive type information about a reviewed item.

3.1.1 Identifying Co-occurring Tags

Tags that are frequently associated with the same item are assumed to be related in some way. In the HTML pages about each tag, tags that co-occur above a certain threshold are displayed to the user. This threshold is set low for HTML output, as human readers of the page are unlikely to infer erroneous information based on these relationships. In contrast however, relationships exposed in RDF descriptions of tags (using the `skos:related` property) are based on a more conservative threshold, in order to avoid erroneous inferences based on these assertions. In the future these co-occurrence relationships may also be described using the SCOT ontology².

3.1.2 Deriving Type Information

We currently derive type information from tagging data in two domains, books and films, relying on external data sources to help ensure accurate results. Firstly, where items are tagged 'book' we parse Web links provided by the reviewer that relate to the item, and attempt to extract ISBN numbers embedded in these links. Where we are able to extract an ISBN number in this fashion we conclude that the reviewed item is in fact a book, and add a corresponding `rdf:type` statement to the triplestore.

If an item has been tagged 'film' or 'movie', we execute a query against the DBpedia SPARQL endpoint in order to find any films that have the same name as the reviewed item. If a match is found then we conclude this item is in fact a film, and add an `rdf:type` statement to this effect to the triplestore. These type statements for both books and films are exposed in the RDF descriptions of items on Revyu, and also serve as the basis for showing additional relevant data in the HTML pages about an item, as detailed below.

² <http://scot-project.org/scot/spec>

4. Revyu Architecture and Implementation

Revyu is built on the same technologies that support many conventional Web applications – Apache, MySQL and PHP – but is also fundamentally a Semantic Web application from the ground upwards. Backend storage of RDF triples is provided by a de-normalised MySQL database. The application layer uses the RDF API for PHP (RAP) [24] for accessing, querying, manipulating and serializing RDF data.

All site content – reviews, data about reviewers, data about reviewed items and the tags reviewers assign to these – is published on the site simultaneously in HTML and RDF/XML. These HTML and RDF descriptions of resources are published as separate crawlable documents on the Revyu site but interlinked using RDF Autodiscovery techniques. The site uses the Review [4], FOAF [9] and Tag [23] ontologies to describe reviews, reviewers and tags respectively, as well as properties and classes from RDFS and OWL.

4.1. Human-oriented vs. Machine-oriented Mashups

The data exposed in HTML and RDF descriptions of resources is not always isomorphic. In the RDF descriptions of resources we choose to simply expose data from Revyu, complemented by links to additional data in external data sets. Data from external sources is not republished in Revyu RDF output. In contrast, HTML pages about items in Revyu do reproduce external data. This approach could be described as using Semantic Web data to produce Web 2.0-style mashups at the human-readable, HTML level, whilst simply linking data at the RDF level.

The rationale for this design decision is as follows: human-oriented HTML documents need to present coherent and comprehensive information to a user viewing that document with a conventional Web browser, without requiring her to navigate to many other pages unless she requires related information. Therefore, the publisher of a human-oriented mashup may choose to integrate data from various sources and provide a coherent view to the user that somewhat masks the provenance of the data, as the user can choose to trust the mashup (and its underlying data sources) based on the extent to which they trust the site on which it is published.

In contrast to this 'Web of Documents' perspective, the Semantic Web is designed to enable data to be integrated from multiple sources in potentially unanticipated ways. As a result, data publishers cannot make any assumptions about how the data in a single RDF document will be viewed, as different Semantic Web applications will approach this in different ways. We argue that browsers for the Web of Data should be able to follow RDF links between related data sources, integrate this distributed data dynamically, and present a unified, coherent view to the end user. In this scenario, replicating RDF data from one source in a different location is redundant and potentially obscures the origin of data, thereby complicating the process of reasoning about and conveying data provenance in user applications.

It should be noted that we do not claim that the Revyu Web 2.0-style mashups represent something that could not have been achieved using conventional Web 2.0 approaches. However, to clarify, the following features distinguish our approach: the simultaneous publishing of data-oriented and human-oriented mashups, so that the

data integration effort we have invested is not lost but can be reused by other parties; the ability to easily integrate additional heterogeneous sources using RDF; and the substantially reduced development costs in producing human-oriented mashups through use of Semantic Web technologies.

4.2. Programmatic Access to Revyu Data

Third parties can access Revyu data by performing HTTP GET requests on the URIs of RDF documents on the Revyu site. In addition, data held within the site can be queried programmatically via the Revyu SPARQL endpoint. The Revyu SPARQL endpoint uses the RAP SPARQL engine, which operates against the same MySQL-based triplestore as the rest of the Web site. This SPARQL endpoint allows third parties to access Revyu data for reuse in their own applications. Whilst in some ways analogous to Web 2.0 APIs that provide remote query capabilities, SPARQL endpoints afford many advantages to the developer: for example, common libraries can be used to query multiple RDF graphs yet return the results as one resultset, effectively allowing joins over multiple data sources; furthermore, developers are not limited to performing certain pre-selected queries defined by an API – instead queries can be performed to retrieve arbitrary elements of the underlying RDF graph.

5. Production and Consumption of Linked Data

5.1. Linkable Data in Revyu

Rather than simply publishing 'islands' of unconnected RDF data on the Web, Revyu was designed from the outset to adhere to the four principles of Linked Data, outlined by Berners-Lee [5]: using URIs as names for things, using HTTP URIs so people can look up those names, providing useful information when someone looks up a URI, and linking to other URIs so more things can be discovered. By following these principles and Linked Data best practices [7] the site ensures that reviews it hosts can be fully connected into a Web of Data.

All things represented on Revyu are assigned URIs: reviews, people, reviewed things, tags assigned to things, and even the bundles that represent tags assigned by one person at one point in time. Providing URIs for all reviewers and reviewed things gives many items a presence on the Semantic Web which they would not have otherwise, and enables any third party to refer to these items in other RDF statements. This 'linkable' data creates the potential for inward links to Revyu from other data sets.

All URIs in the Revyu URI-space can be dereferenced. Attempts to dereference the URIs of non-information resources receive an HTTP303 "See Other" response containing the URI of a document that describes the resource. This adheres to the W3C Technical Architecture Group's finding on the httpRange-14 issue [30], and serves to reinforce the distinction between a resource and a description of that resource, as each has a distinct URI. Content negotiation is also performed on the URIs of non-information resources, whereby the user agent receives a description of the resource in either HTML or RDF depending on the value of the `Accept:` header sent in the initial HTTP request.

5.2. Linking to other data sets

In addition to publishing data in a form that is amenable to inward linking on the Semantic Web, Revyu also aims wherever possible to create outward links from entities in the site to those in external, related data sets, in order to create a Web of Data rather than simply isolated islands of RDF.

At present there are outgoing links from Revyu to DBpedia [3], the RDF Book Mashup [6], the Open Guide to Milton Keynes [13], the Randomness Guide to London, papers from the 6th International Semantic Web Conference, and papers from the Linked Data on the Web (LDOW2008) workshop [8]. These links take the form of `owl:sameAs` statements in RDF, to assert that two URIs identify the same resource, and have been produced in two different ways, *retroactively* and *proactively*.

5.2.1. Retroactive Linking

Our initial approach to linking across data sets took items that had already been reviewed in Revyu, and attempted to match these with the same item represented in different data sets. This approach has been used in the domains of books and films. As described above, where an item has been tagged 'book' and an ISBN can be extracted from the related links provided by a reviewer, then an `owl:sameAs` link is set to the corresponding item in the RDF book mashup. Similarly, where a reviewed item has been tagged 'film' or 'movie' and a match found via the DBpedia SPARQL endpoint, an `owl:sameAs` link is set to that item in DBpedia. This approach has met with some success, but has a number of disadvantages: the book matching algorithm relies on users providing certain types of links when they review an item, which we have found to not always be a reliable assumption; the film matching algorithm is overly reliant on string matching, and constrained by the lack of string similarity functions in SPARQL; custom matching algorithms are required for each type of reviewed item, which is not very scalable when any type of item can be reviewed.

5.2.2. Proactive Linking

In order to overcome the limitations of retroactive linking, we have investigated a proactive approach based on priming Revyu with 'skeleton' descriptions of items that users may wish to review. These skeleton records simply include a text label for the item, a statement indicating the type of the item, a number of relevant keyword tags, and an `owl:sameAs` link to the same item in the external source data set. Not only does this provide a foundation upon which new reviews can be created, it also ensures that new reviews of these items are instantly part of the Web of Data. Compared to the retroactive approach, in this case linking becomes trivial and is guaranteed to succeed, as the skeleton record for an item is created based on data about an item whose URI in an external data set is already known.

This skeleton record approach has been followed when linking Revyu to data from the Open Guide to Milton Keynes and the Randomness Guide to London, both members of the Open Guides family of wiki-based city guides that expose data in RDF. Taking the Open Guide to Milton Keynes as an example, whilst some amenities in the city, such as pubs and restaurants, were already reviewed on Revyu, many more were listed in the Open Guide due to its longer history. Therefore, after identifying items

existing in both locations and making the appropriate mappings to avoid duplication, we created skeleton records in Revyu for the remaining items, setting links back to their Open Guide URIs.

This has enabled latitude and longitude data for many items to be retrieved from RDF exposed by the Open Guide, and used to show a Google Map of the items location (see Figure 2 for an example). The same approach can also be used to expose address, telephone, and opening time information held in the Open Guide.

Figure 2.

Our goal is to apply this proactive approach to many other domains, such as books, restaurants, academic papers and films (for which a set of 12,000 skeleton records has already been prepared based on DBpedia data). The major barrier to this approach at present is not the availability of source data sets from which to create skeleton records, but the ability of the existing infrastructure to scale to very large numbers of triples.

Whilst the proactive approach has many benefits when compared to the retroactive linking method, the requirement for manual preparation of skeleton data sets remains an issue. We expect this to be somewhat mitigated by the inevitable increase in the number of items semantically described on the Web, which can then be automatically harvested and imported into Revyu. However, there will always remain a significant proportion of items that are not described in this way, for which alternative methods will be required. Additional issues to be resolved include management of skeleton records related to items that frequently change, or those that are removed from the external data sets from which skeleton records were derived.

5.3. Consuming Linked Data to Enhance the User Experience

Once links are made between items in Revyu and those in external data sets, these are actively exploited to enhance the experience of our users. This manifests itself in a number of ways, such as using geodata from the Open Guides to provide show locations of reviewed items on a map, as described above. In addition, where `owl:sameAs` statements exist linking films on Revyu to their entry in DBpedia, we retrieve additional information about the film, such as the URI of the films promotional poster, and the name of the director. This information is displayed on the Revyu HTML page about the film (as shown in Figure 3 below), thereby enhancing the value of the site for users without requiring this information to be manually entered into Revyu. Similarly we use `owl:sameAs` links between Revyu and the RDF Book Mashup [6] as the basis for retrieving book cover and author information which is also then displayed on the Revyu HTML page about the book.

Figure 3.

6. Reusing Existing Personal Profiles

Reuse of external data is not limited to additional information about reviewed items, but also about reviewers contributing to the site. A common experience with existing Web applications is that, on registration, the user must create a new profile that duplicates profiles they have already created on other sites. This can increase the burden on the user as they must manage multiple redundant sets of personal information stored in different locations. In addition to consuming external data about items that have been reviewed on the site, Revyu aims to address this issue by reusing existing profile information about reviewers, where this is available on the Semantic Web.

Consequently, people registering with the site are not required to provide copious information to populate their user profile. Instead, where they have an existing FOAF description in an external location they may provide its URI to Revyu. In such cases, Revyu dereferences this URI and queries the resulting graph for relevant information (such as a photo, location, home page URI, and interests), which is then displayed in the HTML version of their user profile, as illustrated in Figure 4.

Figure 4.

In contrast to the HTML view, information from external FOAF files is not republished in the RDF version of the reviewer profile. Whilst this does introduce a disparity between the HTML and RDF documents, in terms of what information is exposed, this decision was taken in order to avoid duplicating RDF data that is already available elsewhere on the Semantic Web. To enable Semantic Web applications to locate this related RDF data `rdfs:seeAlso` links are set between the RDF description on Revyu and the URI of the reviewer's external FOAF file. Where the user has assigned herself a URI in her FOAF description, Revyu also sets `owl:sameAs` links asserting that this URI identifies the same resource as her Revyu URI.

One limitation of this approach is that users who are not familiar with the Semantic Web may not know the location of existing FOAF profiles published for them by sites such as MyOpera [22]. In cases where a user has not listed a FOAF file URI in their Revyu profile, we use two alternative methods to try and enrich their profile.

First, to locate additional profile data about a user we query with SPARQL a store of FOAF data crawled from across the Web and stored in the Talis Platform [20]. Where data is found about the user we display this, however at present the provenance of the

information is not communicated to the person viewing the profile, and no reasoning is performed about the trustworthiness or relevance of such information based on its provenance. We expect to address this issue in ongoing work.

If querying the store of crawled FOAF data returns no information about the user, our second approach is to use the Sindice Semantic Web index [29] and attempt to locate additional RDF documents on the Web that may describe the user, using the following procedure: a SHA1 [11] hash of the user's mailbox URI is generated from his or her registered email address; using the FOAF property `mbox_sha1sum`, and the Inverse Functional Property lookup service in Sindice, a list of URIs of documents that contain this predicate and object is returned; the first of these URIs is dereferenced and queried for relevant information about the user.

As the Sindice index is more comprehensive than our current store of FOAF data, this has greater potential to return information as a result. However, the service returns only links to documents containing RDF data, however this data itself cannot be queried without first retrieving the document. Therefore, at present, only the first document in the Sindice results is queried for data about the user, the result being that Revyu is dependent on relevance ranking within Sindice, rather than applying our own relevance ranking algorithms. As with reasoning about trustworthiness based on provenance, we hope to address this issue in future research.

7. Inferring Trust Relationships from Revyu Data

A broader goal of our research is to use knowledge held within trusted social networks to support information-seeking tasks on the Web. Data collected within Revyu forms a basis for this research, as the reviews created by users, and the tags they apply to items, enable us to infer trustworthiness relationships between people, with regard to particular topics. The algorithms used to generate metrics that represent these trust relationships are referred to as the 'Hoonoh algorithms'.

These algorithms generate 'experience', 'expertise' and 'affinity' metrics, which represent respectively the predicted trustworthiness of an individual with regards to a topic, based on his or her experience of and expertise in that topic, and the predicted trustworthiness of an individual based on the 'affinity' relationship between the information seeker and that individual. These factors are derived from previous research into trust and source-selection in information seeking activities [17].

A fundamental aspect of our approach in this area, based on our previous research [17], is the principle that trust can be topical; one person may be highly trusted for recommendations in one domain but trusted very little in others. For example, one may trust a friend who is a banker to give sound financial advice, but never trust her film recommendations. This trust topicality is supported by the experience and expertise algorithms, whilst affinity captures a more universal trust relationship from one individual to another that is not topical in nature. Keyword tags used in Revyu seed the list of topics in which individuals may have experience or expertise, and also provide a basis for computing measures of experience. Each keyword tag is taken to denote one topic.

Broadly speaking the experience algorithm generates a metric of a person's experience regarding a particular topic, by calculating the proportion of all items tagged with a particular tag that person has reviewed. The credibility algorithm computes *person* \rightarrow *topic* credibility metrics by comparing the numerical rating component of each review to the mean rating of that item across all users. A mean is then taken of all a reviewer's review-specific credibility scores for items tagged with a particular tag, to produce a reviewer's credibility score for that topic. Whether or not an affinity exists between two individuals is determined by a combination of the following factors derived from the reviews they have submitted to Revyu: the extent to which both parties have rated the same items (i.e. the overlap in rated objects), and the consistency in the ratings given by each party to items both have reviewed (this is referred to as the 'rating overlap'). A more detailed account of preliminary versions of these algorithms is provided in [18].

Based on data generated with these algorithms, a Web-based system, Hoonoh.com, has been implemented and deployed that uses these metrics to support source-centric information-seeking within an individual's social network. Hoonoh allows users to search for people with knowledge of particular topics and rank these potential information sources according to the *experience*, *expertise* and *affinity* trust factors.

8. Collecting and Exposing Social Network Data

In addition to consuming data from existing, external FOAF profiles, Revyu enables users to state that they know other Revyu reviewers. At this point the relationship is recorded in the triplestore using the `foaf:knows` property, and exposed (privacy settings permitting) in the user's RDF description on the Revyu site. This ensures that social networking data created in Revyu is not automatically rendered inaccessible to other services, and can play a first class role in a broader, Web-wide 'social graph'. This FOAF data from Revyu is one of many potential sources of social network information that can be used by Hoonoh to rank known individuals as information sources.

9. Future Work and Conclusions

In addition to encouraging further user participation in order to increase the value delivered by the site, we plan to integrate Revyu with a number of additional data sets, as discussed above. It should be noted that our aim in linking to external datasets is not to constrain, but merely to seed, users conceptions of what can be reviewed. As we integrate further data sets we hope to develop techniques for automated population of Revyu with data about reviewable items, from the Semantic Web at large.

A medium-term goal for Revyu is to aggregate review and rating data from external sources, in order to provide a single point of access for structured review data across the Web. An essential prerequisite for this development is more explicit licensing of review data published on the Web, using licenses such Creative Commons (for creative aspects of reviews) [10] or the Open Data Commons Public Domain Dedication and License (for more factual data about reviewed items) [21]. To enable the storage of vastly increased amounts of data on which we can provide a wider

range of services, Revyu will also be rebuilt as an application on top of the Talis Platform [20]. It is our hope that aggregation and search services for review data will help encourage more extensive publication of reviews and ratings in RDF.

As we consume increasing amounts of external data from the Semantic Web at large, we expect that algorithms for reasoning about the trustworthiness and relevance of data based on provenance will become increasingly important. Our aim is to investigate these issues in ongoing research.

In order to provide more structured access to increasing numbers of reviews, we intend to migrate from keyword tagging to a form of semantic tagging. For certain types of information, such as the location of an item that has a physical presence, we plan to introduce a tagging interface that allows users to pick specific locations from a pre-assembled gazetteer, in which each location is associated with a URI in data sets such as Geonames [14] or DBpedia. The nature of the relationship between the reviewed item and the tag could then be explicitly recorded in RDF, allowing users to search for items by location.

In conclusion, in this paper we have described Revyu, a human usable reviewing and rating Web site built on Semantic Web technologies, and fundamentally designed to contribute to the realization of a Web of Data. Whilst superficially not unique in functionality, the site is rare in its status as a publicly available service in daily use that is oriented towards human users, yet also embodies current best practices in developing for the Semantic Web. Experience gained from the implementation of Revyu, and reported here, provides valuable insights for others creating Linked Data and Semantic Web applications.

Acknowledgements

This research was partially supported by the Advanced Knowledge Technologies (AKT) and OpenKnowledge (OK) projects. AKT is an Interdisciplinary Research Collaboration (IRC) sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01. OK is sponsored by the European Commission as part of the Information Society Technologies (IST) programme under grant number IST-2001-34038. The Open Guides and DBpedia communities, and the RDF Book Mashup team deserve our special thanks.


References

- [1] Amazon.com: <http://www.amazon.com/> (accessed 14th May 2008).
- [2] Amazon.com Associates Web Service: <http://www.amazon.com/E-Commerce-Service-AWS-home-page/b?node=12738641> (accessed 14th May 2008).
- [3] S. Auer, C. Bizer, J. Lehmann, G. Kobilarov, R. Cyganiak, Z. Ives: DBpedia: A Nucleus for a Web of Open Data, in: Proceedings of the 6th International Semantic Web Conference, 2007.
- [4] D. Ayers, T. Heath: Review Vocabulary, v0.2 <http://purl.org/stuff/rev#> (accessed 14th May 2008).
- [5] T. Berners-Lee: Linked Data <http://www.w3.org/DesignIssues/LinkedData.html> (accessed 14th May 2008).

- [6] C. Bizer, R. Cyganiak, T. Gauss: The RDF Book Mashup: From Web APIs to a Web of Data, in: Proceedings of the 3rd Workshop on Scripting for the Semantic Web, 4th European Semantic Web Conference, 2007.
- [7] C. Bizer, R. Cyganiak, T. Heath: How to Publish Linked Data on the Web <http://sites.wiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/> (accessed 14th May 2008).
- [8] C. Bizer, T. Heath, K. Idehen, T. Berners-Lee: Linked Data on the Web (LDOW2008), in: Proceedings of the 17th International World Wide Web Conference (WWW2008), 2008.
- [9] D. Brickley, L. Miller: FOAF Vocabulary Specification 0.9 <http://xmlns.com/foaf/0.1/> (accessed 14th May 2008).
- [10] Creative Commons: <http://creativecommons.org/> (accessed 14th May 2008).
- [11] D. Eastlake, P. Jones: RFC 3174 - US Secure Hash Algorithm 1 (SHA1) <http://isc.faq.s.org/rfc3174.html> (accessed 14th May 2008).
- [12] Epinions.com: <http://www.epinions.com/> (accessed 14th May 2008).
- [13] M. Gaved, T. Heath, M. Eisenstadt: Wikis of Locality: Insights from the Open Guides, in: Proceedings of the 2006 International Symposium on Wikis (WikiSym2006), 2006.
- [14] Geonames: <http://geonames.org/> (accessed 14th May 2008).
- [15] J. Golbeck, J. Hendler: FilmTrust: Movie Recommendations using Trust in Web-based Social Networks, in: Proceedings of the IEEE Consumer Communications and Networking Conference, 2006.
- [16] R. Guha: Open Rating Systems, in: Proceedings of the 1st Workshop on Friend of a Friend, 2004.
- [17] T. Heath, E. Motta, M. Petre: Person to Person Trust Factors in Word of Mouth Recommendation, in: Proceedings of the CHI2006 Workshop on Reinventing Trust, Collaboration, and Compliance in Social Systems (Reinvent06), 2006.
- [18] T. Heath, E. Motta, M. Petre: Computing Word-of-Mouth Trust Relationships in Social Networks from Semantic Web and Web2.0 Data Sources, in: Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web 2.0, 4th European Semantic Web Conference, 2007.
- [19] G. Klyne, J. J. Carroll: Resource Description Framework (RDF): Concepts and Abstract Syntax <http://www.w3.org/TR/rdf-concepts/> (accessed 14th May 2008).
- [20] J. Leavesley, I. Davis: Talis Platform: Harnessing Sophisticated Mass Collaboration on a Global Scale. Talis, http://www.talis.com/platform/assets/harnessing_sophisticated_mass.pdf (2007).
- [21] P. Miller, R. Styles, T. Heath: Open Data Commons, A License for Open Data, in: Proceedings of the 1st Workshop on Linked Data on the Web (LDOW2008), 17th International World Wide Web Conference (WWW2008), 2008.
- [22] MyOpera: <http://my.opera.com/> (accessed 14th May 2008).
- [23] R. Newman, S. Russell, D. Ayers: Tag Ontology <http://www.holygoat.co.uk/owl/redwood/0.1/tags/> (accessed 14th May 2008).
- [24] R. Oldakowski, C. Bizer, D. Westphal: RAP: RDF API for PHP, in: Proceedings of the 1st Workshop on Scripting for the Semantic Web, 2nd European Semantic Web Conference, 2005.
- [25] T. O'Reilly: What Is Web 2.0 <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (accessed 14th May 2008).

- [26] E. Prud'hommeaux, A. Seaborne: SPARQL Query Language for RDF
<http://www.w3.org/TR/rdf-sparql-query/> (accessed 14th May 2008).
- [27] F. M. Suchanek, G. Kasneci, G. Weikum: Yago: A Core of Semantic Knowledge - Unifying WordNet and Wikipedia, in: Proceedings of the 16th International World Wide Web Conference (WWW2007), 2007.
- [28] TripAdvisor.com: <http://www.tripadvisor.com/> (accessed 14th May 2008).
- [29] G. Tummarello, E. Oren, R. Delbru: Sindice.com: Weaving the Open Linked Data, in: Proceedings of the 6th International Semantic Web Conference, 2007.
- [30] W3C Technical Architecture Group (TAG): httpRange-14: What is the range of the HTTP dereference function?
<http://www.w3.org/2001/tag/issues.html#httpRange-14> (accessed 14th May 2008).
- [31] G. Walsh, S. Swinford: Hotel review websites: a five-star scam. The Sunday Times (2006)
http://travel.timesonline.co.uk/tol/life_and_style/travel/article634136.ece

Figure 1



[Browse Things](#) | [Search Things](#) | [Browse People](#)
[Login/Register](#) | [New Review](#)

What is Revyu.com?

Revyu.com is a web site where you can review and rate things. Unlike many other reviewing sites on the web, Revyu.com lets you review and rate absolutely anything you can name.

Start a New Review

enter the name of the thing you would like to review

Recent Reviews NEW FEED [\(All Reviews\)](#)

1. [Java Language Specification, Third ...](#) by [topclass](#)
2. [Head first Java](#) by [topclass](#)
3. [Crocodyl](#) by [rhinokitty](#)
4. [The Lowndes Arms, Whaddon, Milton K...](#) by [peterH](#)
5. [East Z East Indian Restaurant, Manc...](#) by [cohenira](#)
6. [The Tradition Restaurant](#) by [serina](#)
7. [Mud Dock Cafe and Restaurant, Water...](#) by [sarah](#)
8. [Brunch Paris Alcazar](#) by [Reviu](#)
9. [Hotel Plaza Opera](#) by [teddypolar](#)
10. [Get Your Data Out! - Danny Ayers' D...](#) by [tom](#)

Popular Tags [\(All Tags\)](#)

[accommodation](#) [article](#) [aswc2007](#) [bar](#) [beer](#) [beer-gardens](#)
[book](#) [bristol](#) [cafe](#) [cinema](#) [conveyer-belt-sushi](#) [craft-shops](#)
[eating-out](#) [film](#) [food](#) [food-served-evenings](#)
[food-served-lunchtimes](#) [free-qaodata](#) [fullers](#)
[good-beer-guide](#) [good-beer-guide-2007](#)
[hotel](#) [iswc2007](#) [japanese-food](#) [london](#)
[milton-keynes](#) [movie](#) [music](#) [paper](#) [paris](#) [pie](#) [pub](#)
[pub-food](#) [real-ale](#) [real-cider](#) [research](#) [restaurant](#)
[samuel-smiths](#) [sandwich-shops](#) [semantic-web](#) [soup](#)
[stony-stratford](#) [sushi](#) [take-kake-here](#) [takeaway](#) [wetherspoons](#)
[wolverton](#)

Top Reviewers [\(All People\)](#)

[Absynthminded](#) [AdamRae](#) [AdrianStevenson](#) [al](#) [AlanJ](#) [AlexLittle](#)
[Aneta](#) [AxelPolleres](#) [bluesnow](#) [bouquet](#) [castagna](#) [Chaz6](#) [Crash](#)
[davidseth](#) [Dnyanesh](#) [drewp](#) [esfandia](#) [Fin](#) [Fouad](#) [glittrgirl](#)
[hockeyshooter](#) [iccq](#) [Kake](#) [k|wa](#) [louiedeparis](#)
[magicrebirth](#) [Mark](#) [Marta](#) [martinp](#) [mgaved](#) [mhausenblas](#)
[Paddy](#) [Paul](#) [pip](#) [redwards](#) [revking](#) [ronny](#) [Rui](#) [Sanyukta](#) [smonroe](#)
[sofia](#) [Stefania](#) [stoermer](#) [teddypolar](#) [testing123](#) [tom](#)
[trickybastard](#) [vlatdn](#) [wcrosbie](#) [xcv](#) [zibididu](#)

Figure 2

[Home](#) | [Browse Things](#) | [Search Things](#) | [Browse People](#)

[Login/Register](#) | [New Review](#)

Ye Olde Swan, Woughton On The Green, Milton Keynes

Links

See Also: http://miltonkeynes.openguides.org/?Ye_Olde_Swan%2C_Woughton_On_The_G...

Tags

[bar](#) [beer](#) [food](#) [milton-keynes](#) [pub](#) [restaurant](#) [woughton](#)

Map

Reviews (2)

★★★★☆ [by tom on 20 Nov 2006](#)

The Swan (as it's commonly known) is one of the nicer pubs within the Milton Keynes boundary, being situation in the old village (more a large hamlet really) of Woughton on the Green. It's got log fires, a crooked roof, and low ceilings, which all give it a cosy feeling on a winter evening. Come summer time there are plenty of tables outside in a pleasant garden. Given the other options nearby it's a good choice, but has plenty of shortcomings. The pub is heavily geared towards food, which is generally quite good, and this is responsible for a stupidly restrictive policy about where you can sit during busy periods. In practice you may find more than half the pub reserved for diners, making it hard to get a table if you just want a drink. There are a few decent ales on tap, including Deuchars IPA and Old Speckled Hen, which generally seem well kept. Anywhere else I'd give this pub a rating of 3, but in MK the lack of decent choice makes this a 4.

[RDF Metadata About Ye Olde Swan, Woughton On The Green, Milton Keynes](#)

[ADD META](#)

[Write a Review of Ye Olde Swan, Woughton On The Green, Milton Keynes](#)

[Add to del.icio.us](#)

Figure 3

REVIEW.COM
Review ANYTHING

[Home](#) | [Browse Things](#) | [Search Things](#) | [Browse People](#)
[Login/Register](#) | [New Review](#)

The Prestige

Links

Homepage: <http://theprestige.movies.go.com/>
See Also: <http://imdb.com/title/tt0482571/>

Tags

[christian-bale](#) [christopher-nolan](#) [drama](#) [entertainment](#) [film](#) [hugh-jackman](#) [illusion](#) [magic](#) [michael-caine](#) [movie](#) [murder](#) [period](#) [scarlett-johansson](#) [science-fiction](#) [whodunnit](#)

Reviews (1)

★★★★★ [by martinp on 23 Jan 2007](#)


This is a drama about intense rivalry between stage magicians in the late 19th Century. The evocation of the period, although first rate, is not the main attraction, however. The Prestige has an incredibly clever plot including the most ingenious murder I've ever come across. It also has a deeply moving and sad love story hidden in it, which gradually emerges over the course of the film.

The film requires a strong suspension of disbelief on some key points: there is a science-fiction premise which is introduced using the real historical character of Nikola Tesla (I'd rather they had used a fictional scientist). There are a couple more implausibilities required to hold it together (something odd that goes on that none of the characters pick up on and a dead-end that by a huge coincidence turns out not to be a dead-end: I can't be more specific without spoiling the plot).

However, rather than feeling cheated by these aspects of the film, I'm hugely impressed. The writers have taken an implausible (okay, impossible) premise but created an intricate, involving and visual story that would be impossible without that premise. Scenes join up with each other in many subtle ways, echoing the same writers' earlier film Memento. Even when you've seen the twist coming, the final scene which lays it all out are has a lot of impact and I suspect the final shot will haunt my dreams.


I expected the film to be about nice costumes or impressive magical trickery, but it is actually about deep emotions felt by the main characters as they deal with the situations life has dealt them, and it rather than serving up those emotions on a plate, it requires you to think and piece together what you've seen. That's got to be a good thing, in fact the best of what film a be.

The Prestige



directed by [Christopher Nolan](#)

[RDF Metadata About The Prestige](#)



[Write a Review of The Prestige](#)

[Add to del.icio.us](#)

Figure 4

[Home](#) | [Browse Things](#) | [Search Things](#) | [Browse People](#)

[Login/Register](#) | [New Review](#)

Reviews by *kjwa* (8)

[ARC RDF Classes for PHP](#)

★★★★★ on 26 Feb 2008

A very usable and fairly comprehensive suite of classes for parsing RDF and/or storing it in MySQL and querying with SPARQL.

I use it all the time.

[Smultron](#)

★★★★☆ on 26 Feb 2008

An open-source text-editor for macosx. I don't like it as much as Textmate or Vim (the interface has too many icons for my tastes), but it's quite good, and free and open source, and has been getting more powerful.

[Colloquy](#)

★★★★☆ on 26 Feb 2008

A nice IRC app for Mac OSX, does logging, etc.

[cyber duck](#)

★★★★☆ on 26 Feb 2008

A decent OS FTP client for Mac OSX. Crashes sometimes, but generally does the job.

[planet rdf](#)

★★★★☆ on 26 Feb 2008

Great aggregation of semantic web related blogs

Do you know *kjwa*? [Login](#) or [Register](#) to add *kjwa* to your Network

[Web Feed of *kjwa*'s Latest Reviews](#)

WEB FEED

About *kjwa* (*Keith Alexander*)

[kjwa's Home Page](#)

[RDF Metadata About *kjwa*](#)

META

[Add to del.icio.us](#)