

Creating Semantic Mind Maps from Linked Data with AutoMind Creator

Csaba Veres
The University of Bergen, Norway
csaba.veres@uib.no

ABSTRACT

AutoMind Creator¹ is an iOS application that lets users interact with linked data to produce customized views. These can be exported as graphical visualizations we call Semantic Mind Maps, as well as rich text (RTF), outline (OPML) and Freemind. We present a new technique for linked data visualisation called Semantic Mind Maps which are rich Mind Maps whose nodes are semantically grounded with a defining URI. The maps are essentially a compact knowledge representation format from which users can further explore information of interest. This paper describes the implementation of AutoMind, and highlights some particular pitfalls in programming for a commercial application with linked data, especially on the Apple ecosystem.

Keywords

linked data, mind maps, iOS, Apple, visualization, knowledge discovery, information space

1. INTRODUCTION

Linked open data presents an exciting opportunity for turning the Web of documents into a Web of data [1]. The Linking Open Data Project has been involved in identifying existing open data sets that can be exposed as RDF. Prominent current examples of such datasets are DBpedia, W3C WordNet, and Geonames. Ideally, linked data descriptions should be machine readable and encompass a useful notion of semantics to enable the programming of knowledge rich applications.

In addition, the web of data also provides an opportunity for humans to find clear, unambiguous facts about topics of interest. DBpedia, for example, summarizes the key facts about topics in Wikipedia pages. These facts could be very useful for humans if they had a suitable application that facilitated user friendly interaction with the data. The motivation for AutoMind was to create a tool for humans to

¹<https://goo.gl/OzAstw>

explore linked data and to export interesting summaries of that data in a useful form.

The remainder of this paper explains the design behind the application, including the rationale of semantic mind maps, and discusses some issues with implementation.

2. BACKGROUND

Mind Mapping is a freeform diagramming technique for intuitively capturing key concepts in a domain. It is perhaps the simplest concept diagramming technique, consisting of a single central concept from which sub concepts radiate in independent tree structures. Each branch is labelled with a keyword or image. It is possible to embellish mind maps with additional features, especially if one uses mind mapping software. A typical addition is to highlight concepts with colour, font, shape, or any number of demarcating features. While these additions do not have a formal semantics, they can take on idiosyncratic interpretations to individual mind mappers. In addition, colour and imagery can help with the visual organisation of the concepts in a mind map [3]. The lack of well defined semantics for the model components indicate that mind maps are not so much a formal modelling language, but rather a way to capture “brainstorming sessions” in a concise, structured representation [2]. The visual components are designed for human comprehension rather than formal interpretation.

In spite of the lack of formal rigour, mind maps have proved useful in the software development process. For example Bia et. al. [5] used mind maps to model XML DTDs and Schemas as sets of parallel trees and implemented XSLT transformations to generate FreeMind mind maps. The advantage of mind maps in representing the complex graph structures is that they enable the intuitive navigation of the structures with selective hiding of sub branches. They managed to successfully model, design and modify complex Schemas by constructing manageable, easily comprehensible diagrams.

The goal with semantic mind maps was to enhance the basic mind map notation in a minimal way that would preserve its simplicity and user friendliness, but nevertheless add a level of semantic description to enrich the expressiveness and comprehension of the map. Eppler [4] notes that mind maps can become inconsistent and comprehensibility can suffer as the size of the mind map grows. Since links have no formal interpretation, concepts can become linked in idiosyncratic

ways, and interpretation can suffer. Semantic mind maps are designed to mitigate this problem by semantically grounding the nodes and links of a mind map.

Unfortunately, the creation of semantic mind maps is complicated by the need for grounding each concept; This is where linked data comes in, by providing a supply of grounded, inter related concepts. The goal of the present application was to enable the exploration of domain general linked data with semantic mind maps. In particular we used DBPedia as the primary data hub, and exploited several links to other data sources like Project Gutenberg, New York Times Open Data, and the CIA World Factbook. However, the approach can be extended to any data sources. AutoMind Creator is an exploration and visualization tool for linked data.

3. EXAMPLE USE CASE

A primary use case is for students or office workers who want to produce a quick presentation on a topic, or to create a rich basis from which to further develop the presentation. For example, suppose a student had to make a presentation related to Van Gogh, and searches for the words in the first screen of the application. The search returns a number of results related to Van Gogh, including some unexpected ones like "Van Gogh (1948 film), Theo Van Gogh (Art Dealer), and Van Gogh Museum" (see figure 1). In order to differentiate himself from his fellows the student selects "Van Gogh Museum", which reveals the first property selection screen.

A typical set of properties for the resource *Van Gogh Museum* is shown in figure 2(a). Selecting a row will initiate a segue to a new table which displays all the objects for the selected predicate. Users can select which new resources to include in their mind map with the use of toggles, as shown in 2(b). In addition, when these resources are themselves the subject of another set of triples, then selecting the row will cause a transition to a new table displaying the properties which are relevant to that resource. One of these predicates can then be selected, transitioning to a new table with the objects of that predicate. The user can of course go back in the series of tables at any time. The forward or backward transitions can repeat ad infinitum. For example, the user can select "Van Gogh Museum hasLocation Amsterdam", but then on a further screen they could select "Amsterdam birthPlace (of) Jaap Voigt", and then "Jaap Voigt subject Dutch field hockey players", then go back to select another property of "Van Gogh Museum" and so on.

The user can preview the mind map at any time during the construction process. A typical graph segment is shown in figure 3. The mind map is interactive in that selecting a node will reveal the hyperlinked resource. However, nodes in the mind map can not be moved, deleted or added. In order to add or delete nodes the user must return to the selection mode and toggle switches to chose the desired resources. A future release might include the option to delete and re order nodes. However, at this stage the intention is to leave the creation of the nodes with the system of tables. More elaborate modification of the maps would be possible by exporting to a more general mind mapping tool.

The mind map can be shared at any time in several ways, including Twitter or email. Since the code generating the

mind map is HTML and Javascript, it is not possible to directly share this on the typical social networks. Our solution is to upload the HTML to a private FTP server, and share a public link to that file. The source HTML can be saved locally and edited or reused in any way.

The mind map is a rich medium for presenting information. The nodes can be clicked to expand or collapse, to highlight relevant detail. Each node is defined by a grounding URL. Double clicking the node (or long pressing on a touch device) opens a new window with the reference of defining URL. An example map for the Van Gogh Museum can be retrieved at <http://csabaveres.net/VanGoghMuseum.html>.

The application also makes it possible to export the knowledge graph in a number of useful formats, currently restricted to Outline Processor Markup Language (OPML), Rich Text Format (RTF) and FreeMind mind mapper format. OPML files can be opened with many outliner applications which present the facts in bullet list that could be used to explain the topic in a clear sequential manner (figure 4(a)). RTF is a portable document format which can be edited by most popular word processing and presentation software. The RTF text is in a clear tabbed format and complete with citations in the form of the URL links which makes an ideal basis for an essay (figure 4(b)). Alternatively the points can be easily transformed into a presentation which can also be enhanced with additional points. Finally the mind map can be exported in FreeMind format, which is a popular open source mind mapping application that can be used to edit and supplement the map generated by AutoMind. Our approach to building the tool follows the UNIX philosophy of combining "small, targeted tools" to accomplish bigger tasks, so we do not try to replicate editing capability which is already provided by FreeMind. The intended workflow is that users generate a rich but possibly incomplete mind map with our tool, and then enhance this in FreeMind by adding nodes from sources not available in AutoMind. Since the base map is well structured and provides a coherent semantic framework, there is at least a possibility that additions will themselves be systematic and maintain the semantic integrity of the mind map.

4. RESEARCH ISSUES

AutoMind creator was primarily designed as a practical tool to enable users to navigate, serialize and visualize linked data in a user friendly manner. A widespread adoption of the tool would provide a unique opportunity to study how users create personalized data structures from open linked data.

We have some preliminary evidence that people find the formalism useful in complex information processing tasks. In a currently unpublished masters thesis, a student performed a case study in which she constructed a set of semantic mind maps that captured operations at an insurance company from different operational perspectives. One was a high level map of the company objectives and mission statement. Another captured part of the sales process, and yet another involved data warehouse facts. Thus each map was constructed with concepts that are meaningful to different target audiences who are typically not privy to each other's concerns. However, the nodes in each map were semanti-

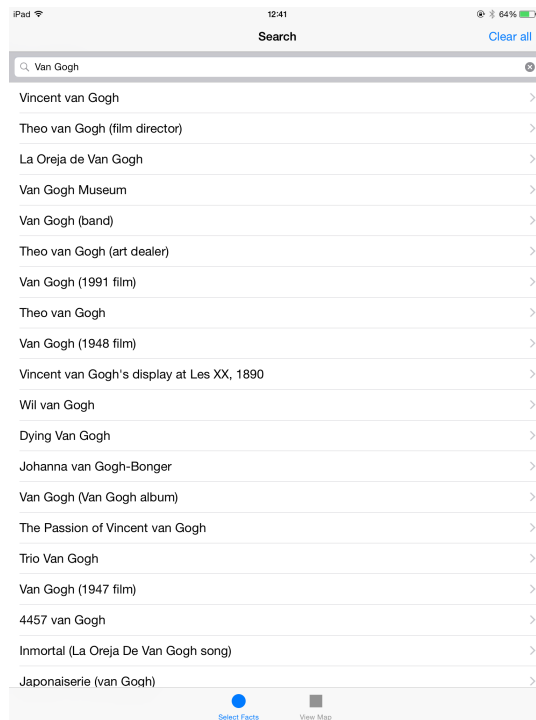
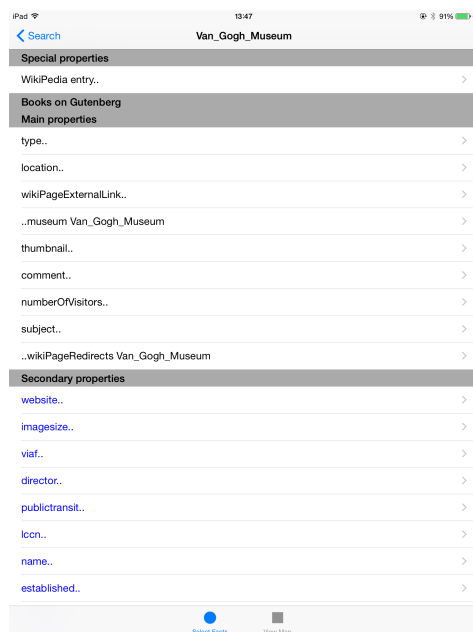
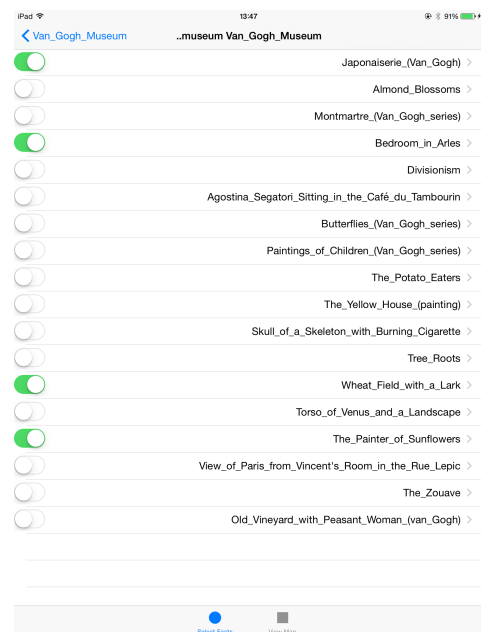


Figure 1: The search screen.



(a) Chose property row



(b) Chose property value with toggles

Figure 2: The information selection screens

cally annotated with concepts from a domain ontology that included concepts from every level of description. Nodes in different mind maps could be related through the ontology. For example, high level strategic goals were related to the sales processes that support those goals, and to data warehouse facts that reported on the success of achieving

the goals. By clicking on the hyperlinks of any mind map, the users were directed to different mind maps in which the related concepts appeared. This way managers could see which sales processes were successful, and data warehouse people could see which strategic goal each fact impacted. The case study was very successful, and users generally

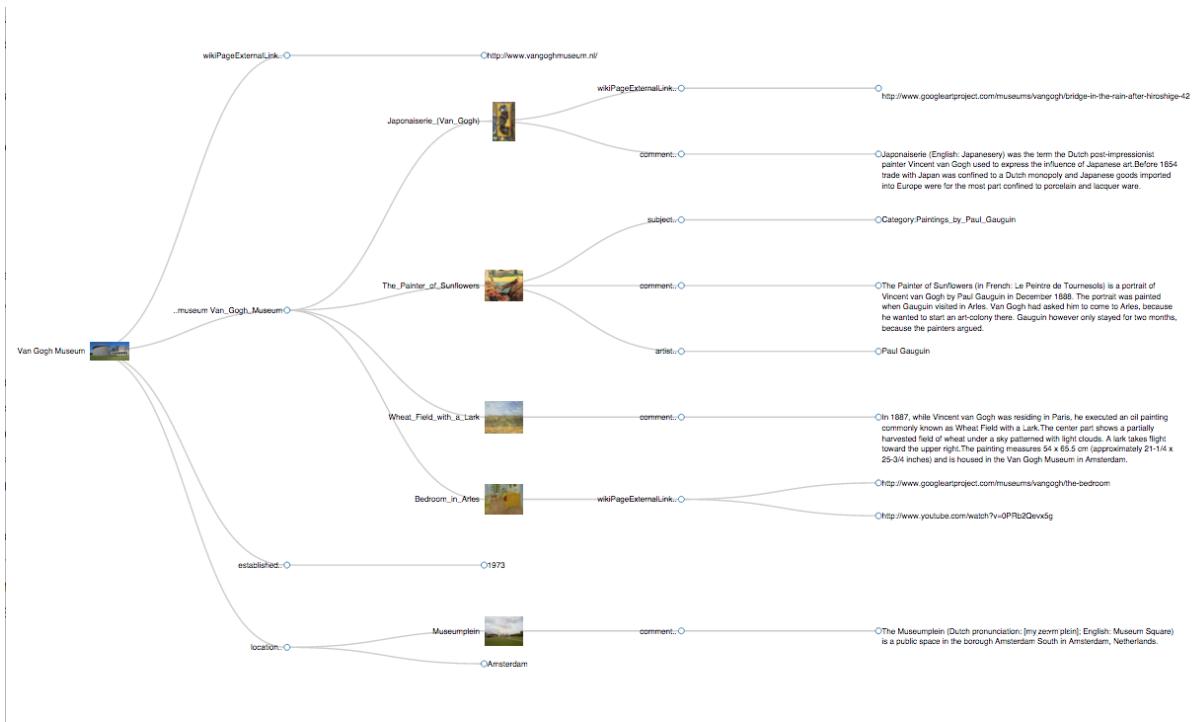


Figure 3: A section of a mind map.



(a) OPML export

(b) RTF export

Figure 4: Two export formats

found the tool to be easy to use and a great help in understanding processes, especially when there were anomalies in the process. For example, the mind maps revealed that

the sales process recorded a new contract at the point an offer was made. However, the accounting process recorded the new contract at the point it was actually accepted and

payment made. This revealed a previously mysterious gap between the reported number of new contracts and the income from a particular division in the company.

We are hopeful that an intuitive way to exploit linked data will facilitate its popular adoption.

5. SYSTEM ARCHITECTURE

This section describes the implementation of AutoMind in Objective-C, which was the main development language for iOS prior to the release of Swift in June 2014. The logic is entirely client side, but no data is stored locally. Data is retrieved on demand by sending a SPARQL query to an endpoint using an `NSURLRequest` object.

5.1 Concept Search and Selection

The initial task is to locate the appropriate concept within DBpedia in response to a user's search request. This could be done in several ways with SPARQL queries filtering on the text string. In the initial implementation we chose to use the DBpedia Lookup Service instead². The service can be used to look up DBpedia URIs by related keywords. For example the resource `http://dbpedia.org/resource/United_States` can be looked up by the string "USA" or "United States". In addition the results are ranked by the number of *inlinks* pointing from other Wikipedia pages at a result page. Therefore the quality of the returned results is very high. However, the index required for the service has not been maintained and is based on DBpedia 3.8 which is four versions out of date. It is possible to build a new index but the software provided does not compile a new version and an extensive discussion on the help forums did not yield any results. In spite of some interest, no one seems to be able to build a new index, and the original developers don't appear to be very helpful. The newest versions therefore use SPARQL queries extended with Virtuoso server's free text indexing capabilities.

The search results are presented in a table view which allows the user to select a row. This sends out a volley of SPARQL queries to retrieve data about the resource. There are three SPARQL endpoints in use at the moment: DBpedia at a private mirror which also includes the NYT open data, Gutenberg at the University of Mannheim and CIA at the University of Mannheim. The DBpedia triples are served from a private mirror for two reasons. First, the public DBpedia endpoint can be unreliable. Second, a curated subset of the triples is more user friendly than the entire set available at the public endpoint.

The triples which describe the resource are pooled and presented to the user for selection. Of course different resources will typically have a different set of triples. For example only countries have entries in the CIA World Factbook. Information about the user selected nodes are stored in a SQLITE database through the Core Data framework offered in iOS.

5.2 Graph Generation

The goal was to automatically generate high quality mind maps that could re adjust as nodes are added or deleted. We decided not to implement our own graphics sub system, and

²<http://wiki.dbpedia.org/lookup/>

opted for a freely available framework. There are a number of interesting Javascript visualization libraries that could be run inside an iOS WebView. After trying a number of them we settled on two of these for production use.

The first framework we used was the ECOTree.js framework³. This produced realistic looking mind maps from a simple textual specification of the nodes, as in the example below. The instructions can be generated from the database in a straightforward manner.

```
var myTree = new ECOTree("myTree","myTreeContainer");
myTree.add(0,-1,"Apex Node");
myTree.add(1,0,"Left Child");
myTree.UpdateTree();
```

A major problem with the ECOTree framework is that it uses the HTML5 Canvas element, whose maximum size is limited by iOS to 3 megapixels for devices with less than 256 MB RAM and 5 megapixels for devices with greater or equal than 256 MB RAM. This meant that maps beyond 30 or so nodes could not be rendered on iOS, and the exported HTML crashed an Android device on testing.

We switched to the popular D3.js library⁴ for data driven documents. D3.js uses SVG to render the image and is not subject to such resource constraints (at least none that we have discovered so far). Fortunately D3.js can generate tree diagrams from "flat" data as well as its more typical JSON representation, so switching to D3.js mainly involved small adjustments in the way the descriptions are generated from the database.

5.3 Export Formats

As previously noted, in addition to exporting the graphs as HTML containing the D3.js code, the application can export to RTF, OPML, and as FreeMind mind mapping format. Fortunately, since we chose to represent the basic description of the nodes and their relationships with the flat textual descriptions, it was relatively easy to generate all of the formats in the same traversal of the nodes.

Figure 5 shows the relationship between a root node and its first child as represented in the D3.js, OPML, and RTF formats.

6. EXPERIENCE

There are a number of lessons to learn from developing a commercial application in which there is at least an implicit promise of speed, reliability, and continuity.

First, obtaining data is not optimal, even though the data is open linked data. For example the `dbpedia.org/sparql` endpoint was not reliable during the development process, so we decided to establish our own mirror with a reliable response time⁵. Unfortunately this did not help with the

³<http://www.codeproject.com/Articles/16192/Graphic-JavaScript-Tree-with-Layout>

⁴<http://d3js.org/>

⁵Thanks to a reviewer for suggesting that the Linked Data

CIA WorldFactbook data which is not available in RDF for running on a local server. Instead they are generated from database dumps using D2R, and provided by the University of Leipzig and the Free University of Berlin, so the application is currently at the mercy of that service. If it ceases to be maintained, that part of the application will stop working. As an added problem, the DBPedia mapping file for Project Gutenberg uses an identifier at the FU-Berlin, rather than at the Gutenberg site. For example, the book "The Motor Girls on a Tour" by Margaret Penrose has the identifier <http://www.gutenberg.org/ebooks/2789> at Gutenberg, whereas it appears as <http://wifo5-04.informatik.uni-mannheim.de/gutendata/page/etext2789> in the DBPedia mapping file. Therefore the developer has the choice of performing some manual URL rewrites, or to use the service at the Uni-Mannheim site. This need for this kind of hacking should not be necessary for linked data.

The data contained in sources, primarily DBPedia, can be patchy and idiosyncratic. Many concepts have very little useful information in DBPedia, while some have too many properties to easily assimilate. Further, the automatically extracted data can include non substantive properties like *image_size*, whose use is further impaired by the unstructured nature of linked data, which makes it impossible to properly apply these predicates. For example there is no way to know which image has the *image_size* property. Thus, the quality of mind maps created from linked data can vary a great deal depending on the root concept. Our partial solution involved curation, where we limited the available DBPedia data sets in our server.

Releasing the application through the Apple distribution channels also proved challenging. During one of the iterations an Apple reviewer noticed the New York Times data and blocked the release until documentation could be produced to prove that we had the rights to use the data. After many iterations in which we explained the status of open data under the creative commons license, the app remained blocked. Finally an appeal to the Appeals Board was able to convince them to release the app, but with the condition that all screen shots of NYT pages were removed from the app page. With this warning about potential future problems in such a controlled ecosystem, we are planning on moving our effort to the Android platform.

AutoMind was initially released as a paid application, but subsequently a free version was made available. The free version uses the public service end points with no guarantee of performance or reliability. In addition, the data is restricted to DBPedia and does not include the New York Times or Project Gutenberg links. The DBPedia data is not curated, and therefore includes significant number of predicates with dubious usefulness.

In terms of further development, we are very keen to extend the social computing aspect of the application. As already mentioned, the mind maps created by our users are uploaded to an FTP server. It is our intention to create a portal around this server through which users can explore

Fragments and the Triple Pattern Fragments interface may be helpful. We will certainly explore how this might be of help.

the mind maps contributed by the community. In addition we aim to provide tools which can integrate mind maps with overlapping concept nodes.

7. CONCLUSION

AutoMind is a new app for iOS which attempts to be a user friendly tool for humans to navigate their way through linked data, and to summarize their findings with the novel new representation of Semantic Mind Maps. A popular adoption of the approach should drive a need for good quality data which will benefit the linked data effort.

The development process taught us that working with linked data does not necessarily make the task of information search and integration easy. It can sometimes be difficult to link different relevant data sets, contrary to the linked data vision. A uniform interface like Triple Pattern Fragments is certainly a step in the right direction, and we are keen to explore its implications for our application. Raw linked data can be confusing for novice users, prompting the need for curation. Finally, the inclusion of open data in applications controlled by large corporations can be challenging.

8. REFERENCES

- [1] Christian Bizer, Tom Heath and Tim Berners-Lee (2009) Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, Vol. 5(3), Pages 1-22. DOI: 10.4018/jswis.2009081901
- [2] Buzan, T., and B. Buzan. (1993). TheMindMap book: How to use radiant thinking to maximise your brain's untapped potential. New York:Plume
- [3] Budd, John. W., Mind Maps as Classroom Exercises. The Journal of Economic Education, Vol. 35, No. 1 (Winter, 2004), pp. 35-46
- [4] Eppler, M. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. Information Visualization 5, 3, 202.
- [5] Bia, A., Munoz, R., and Gomez, J. (2010). Using Mind Maps to Model Semistructured Documents. In Research and Advanced Technology for Digital Libraries, Vol. 6273, pp. 421-424. Berlin, Heidelberg: Lecture Notes in Computer Science, Springer.
- [6] Pink, D.H. (2005). Folksonomy. The New York Times, December 11, 2005.
- [7] Golder, S. and Huberman, B. A. (2006) Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2):198-208.

```

{"name":"Van Gogh Museum", "parent" : "null", "URL":"http://en.wikipedia.org/wiki/Van_Gogh_Museum
", "icon":"http://commons.wikimedia.org/wiki/Special:FilePath/Van_Gogh_Museum_Amsterdam.jpg?width=300" },
{"name":"wikiPageExternalLink..(9)", "parent" : "Van Gogh Museum", "URL":"http://dbpedia.org/ontology/wikiPageExtern
{"name":"http://www.vangoghmuseum.nl/(10)", "parent" : "wikiPageExternalLink..(9)", "URL":"http://www.vangoghmuseum

<?xml version="1.0" encoding="UTF-8"?>
<opml version="2.0">
<head>
<title>VanGoghMuseum</title>
</head>
<body>
<outline text="Van Gogh Museum" type="link" url="http://en.wikipedia.org/wiki/Van_Gogh_Museum">
<outline text="wikiPageExternalLink.." type="link" url="http://dbpedia.org/ontology/wikiPageExternalLink">
<outline text="http://www.vangoghmuseum.nl/" type="link" url="http://www.vangoghmuseum.nl/">
</outline>
</outline>

{\rtf1\ansi\deff0 {\fonttbl {\f0 AppleCasual;}}
{\colortbl;\red0\green0\blue0;\red255\green0\blue0;}
Van Gogh Museum\line(http://en.wikipedia.org/wiki/Van_Gogh_Museum)\line\line
\tab\f0\bullet wikiPageExternalLink..\line
\tab\tab\f1\bullet http://www.vangoghmuseum.nl/ (http://www.vangoghmuseum.nl/)\line

```

Figure 5: The same facts in D3.js, OPML and RTF