Extending Traditional Wiki Systems with Geographical Content

Piotr Gawrysiak
Warsaw University of Technology
Institute of Computer Science
Ul. Nowowiejska 15/19, Warsaw, Poland
pgawrysiak@supermedia.pl
WWW home page: http://www.ii.pw.edu.pl

Abstract. The paper is an overview of the project that is being carried out at the Warsaw University of Technology. The project strives to create a Wikipedia – like system, that will allow people to collaboratively create and edit vector data, especially city maps and plans (instead of text, as in traditional wiki systems, such as Wikipedia) in order to provide communities and businesses an accurate and cheap geospatial information. Contrary to other systems that have been created towards this purpose, and which could be potentially integrated into Wikipedia, the system which has been described assumes that geospatial accuracy is not achievable for ordinary users and thus provides built-in mechanisms for dealing with data uncertainty and inaccuracy. Additionally a collaboration infrastructure essential for Wikipedia-like growth (version control, possibility to discuss and revert changes, reliance on conceptual objects – such as crossroads – instead of absolute geospatial coordinates etc.) is also provided. Finally the system prototype is designed in such a way, that no external software (such as Java VM), apart from ordinary web browser, will be required.

1 Introduction

1.1 The importance of Wiki content creation model

Traditionally, all knowledge workers have been producing new information (in a form of business reports, scientific papers etc.) alone. Of course the work quite often

was shared between several people, but the sharing was done only periodically, so in effect people individually created parts of documents that were later combined (also – usually – by a single person), or worked on subsequent versions of a single document. The computerization did not in fact change much in this model. Obviously, people were using electronic means of communication, and the document combining stage has become much easier, thanks to capabilities of modern word processors.

The real breakthrough in content creation has been achieved only with the advent of Internet and so called Wiki systems.

Traditional wiki [Cunningham05] (or wikiwiki as these were originally named, after word "quick" in Hawaiian language) is a collaborative web site, that allows its modification – be it addition of new information, deletion or edition – by any user. In practice such wiki systems are usually implemented as multiple tier systems, comprising of a database storing contents of web pages and a display engine, that creates individual pages on demand basing on the contents of the database.

While in theory it would be possible to allow users to use HTML to input contents of wiki pages, most systems use simplified formatting languages, that provide only limited formatting functionality (for example allow to mark sections of the text as bold of underlined) and simplified linking capabilities that facilitate including hyperlinks to other sections of a wiki site into text entered by a user.

Due to the fact that every user of the wiki system is able to change any piece of the content stored within, most wiki systems maintain a history of the changes introduced into pages and allow to compare current version of the contents with one of the stored, historical versions. Such functionality is crucial for effective growth of the content database. Various users may have different, and sometimes contradictory knowledge that can be inserted into database, so mistakes or even deliberate acts of vandalism (content deletion or falsification) are virtually unavoidable, thus ability to revert changes is required.

Wiki systems, while conceptually very simple, make possible unprecedented speed of creation of information resources. Due to drastically lowering entry barriers to creating web content – potential content contributors do not have to be IT professionals and do not have to know even the HTML language – wiki systems fulfill the promise of universal publishing rights.

Probably the most famous wiki system is the wikipedia encyclopedia [Wikipedia], that currently comprises over 940 thousand articles in English language (and several thousand in other languages including German, French, and Polish) what means that it is currently the largest available encyclopedia (for comparison Encyclopedia Britannica contains around 120 thousand articles).

Obviously wiki systems are not perfect. Their main weakness is directly related to their main functionality i.e. ability to change contents by virtually anybody. As long as all content submissions are honest, the system maintains its integrity, but as there is no content verification mechanism a malicious user can enter incorrect information into a wiki system, or delete important content. In practice however this problem is circumvented thanks to the so called many-eyes principle – it is assumed that a particular piece of content is observed by several people, so even if it is deleted or wrongly modified, it will be quickly corrected by other users, additionally

abovementioned versioning systems implemented into most popular wiki sites, allow relatively easy reversion of destructive operations – such as massive content deletion [Krotzsch05], [Shah05].

1.2 Demand and supply – a reason for Wikipedia success

The rapid growth of Wikipedia was somewhat unexpected and is regarded sometimes as a one-time phenomenon. It was however only possible as a natural consequence of a typical economic[10] demand and supply lack equilibrium. With the ever increasing importance of knowledge economy, the need for good quality information becomes stronger while traditional information repositories – such as encyclopaedia – become not sufficient for a variety of reasons:

- **Cost factor** – while for most inhabitants of western countries the cost of i.e. Encyclopedia Britannica seems to be relatively small, it might be prohibitive for citizens of developing countries. Moreover, some more detailed sources of information (such as special – purpose information repositories, market reports, constantly updated news services etc.) may have even higher cost.
- **Reliability** – information repositories that are not free can be manipulated or censored and thus the information quality might be poor while the repository created by its users in wikiwiki style is basically a peer-reviewed content database and as such, potentially very reliable.
- **Copyright restrictions** – copyright restrictions imposed on content stored in traditional sources limit usually the applications of this content, while the free (in GNU/FDL sense) information can be put to more uses. Metaphorically one might say that a free content can live independently while traditional content is just passive.

On the other hand the technical infrastructure of the internet became sufficient to allow construction of a collaborative editing system, that would allow to exploit the "wisdom of crowds" effect, observed for example by Sir Francis Galton[11]. Towards this end the system must exhibit at least these two properties:

- **information aggregation mechanism –** obviously, the pieces of knowledge that are possessed by individual users of the system must be somehow integrated. Textual, encyclopedic content is especially well suited to such integration, as even traditional reference books have many authors.
- **low entry barrier** – the success of the system is dependent on its ability to attract users. Therefore the mechanism used to input knowledge into a database must be user-friendly and not intimidating. Wikipedia, via its of web page editing (no HTML knowledge is required, users do not even

[10] Obviously, we deal with knowledge economy in his context.

[11] Who observed that a group of people gather on a trade fair can collaboratively (by expressing their guesses and averaging them) approximate quite well a weight of a ox which was exhibited there. For description and a detailed discussion see for example [Surowiecki04].

have to create accounts in a system in order to edit contents) makes it possible.

1.3 Non textual content types

Currently available wiki systems are text oriented – they allow entering and manipulation of text (in some cases illustrated with simple images). While text is certainly the most prevalent means of human expression, there are many other ways of storing concepts and ideas. These include music (in notation form but also as recordings), sculpture, painting, diagramming etc. While most of these means of conveying ideas are not easily susceptible to computerization, some seem to be quite for systems similar to wikiwiki. The content that could be potentially edited in a wiki system must of course be in some way dispersed or decentralized – so that everyone can add his piece increasing overall scope of the entire database. It must be also easily editable in a collaborative way and finally it must be relevant to wider audience since the ability to attract people contributions is essential for system growth. Lets consider some examples:

- **Music** (or audio in general) – there is currently no easy way to edit this content collaboratively, however polyphonic music could be potentially created this way;
- **Sculptures** – it potentially could be decentralized, but current technology does not allow for collaborative, online editing; someday perhaps, with robotic tools or 3D virtual environments, it should be achievable;
- **Scents** – definitely this is not a content easily editable, but it could be very useful. One might imagine for a example a "wikiparfum" database;
- **Source code** – this sounds interesting. Editing source code „wiki style" is a wild idea (such software would not even compile most of the time), obviously not doable with all kinds of programming languages – with scripting languages it should be feasible;
- **Photographs** – difficult to imagine the collaborative editing process, perhaps retouching could be collaborative. There is however no „crowd wisdom" to be tapped, as long as one considers only single photograph at a time. On the other hand the collaborative photographic databases are possible, such as Flickr! [Flickr];
- **Drawings** – this seems to be the most suitable content type for wiki editing. Drawings are – similarly to text – a content that where smaller pieces can be combined in order to create bigger drawings; it is possible to create a collaborative editing tool (see next chapters) and finally – there is quite a lot important information contained in drawings – such as, for example, city maps and plans.

2 Drawings as wikicontent

2.1 Problem Description

There are practically no free (as in GNU/FDL) repositories of map data, while (as for example Tim Berners Lee recently noted during a speech at the Oxford University and as various Google Local mashups [Gibson06] demonstrate) there is a lot of potential uses of this data. In short, this seems to be an ideal situation to be remedied by a community driven, collaborative effort, similar to Wikipedia.

Some work has been already done in this field (see for example [Haraguchi03] or [Openstreetmap]) but the results are highly unsatisfactory. Available tools are either too complex or too primitive to allow us to easily create together such a simple thing as a tube map and, while some efforts resulted in a creation of a public map wiki systems, these remain unpopular and contain practically no data yet.

Above statement might seem a bit surprising, as diagrams, and especially geographical diagrams, such as maps and plans, are especially good candidates for collaborative editing. Their complexity – as far as potential processing algorithms are concerned – is relatively small as most such diagrams can be represented as graphs. At the same time the information represented in this form is immediately understandable, highly useful and usually not easily available to individual users. Most GIS data is protected by quite restrictive licenses, so while there are in theory many sources of good vector maps of – for example – major cities, and quite a large number of consumer products incorporating this data (map booklets, mapping software etc.), it is very difficult and expensive to obtain royalty free geographical data, for example for small, non commercial projects.

In most cases exact geospatial data – i.e. exact geographical coordinates is not crucial for typical applications, as long as logical map structure, that is the street and crossroads layout is complete. In other words a city plan is in most cases sufficient and detailed map is not required.

Above observation is important, as most people are able to draw plans and not detailed maps. This means that it should be possible to create a wiki system, where different people could be able to submit information about geographical features – such as street layout of the city or location of points of interests. Of course such data would be in many cases incorrect or incomplete, but as it will be reviewed by many people, the quality of such map would constantly improve, provided that the system would be equipped with algorithms able to automatically resolve conflicts in submitted data such as different locations of i.e. streets.

An OpenStreetmap system, mentioned above, is a good example illustrating a failed attempt of such system, where above observation was not considered important. This system, enforces accuracy on users, do effectively in order to add information about a geographical feature, such as street, one must know its exact location, which is usually impossible to obtain without a GPS equipment. Additionally, the system is based on Java applets, so its relatively slow thus further increasing entry barrier. In effect the number of contributors and amount of data stored in OpenStreetmap remain miniscule.

A counter example might be a Placeopedia system [Placeopedia], which is a rough geotagging tool for Wikipedia entries. It allows quick adding of approximate

geographical location (using Google Maps API) relevant to selected Wikipedia articles. As the entry barrier is very low (to add content one must just click on Google map or a Google satellite image around a place which he thinks most appropriate) the system is growing quickly and is already quite useful for Wikipedia users.

2.2 Demand and supply analysis

Lets consider demand and supply factors mentioned in the introduction, concerning geographical diagrams and their collaborative editing.

Demand:

- Usefulness of such data is unquestionable, as numerous Google Maps mashups demonstrate.
- There exist many repositories of GIS data, in most cases created by public government institutions. However usually the information contained in these repositories is available only at a fee and is heavily copyrighted. Some efforts are underway to open (or "liberate") these repositories (Tim Berners Lee for example is campaigning for opening the databases of Ordnance Survey of United Kingdom), but are not yet successful.
- Usefulness of data obtained from commercial maps is usually very limited. Even such natural operation as xeroxing a section of a map in order to show friends how to drive to ones house in most cases represents a copyright violation!
- Commercial maps may contain errors, even deliberate:
  - Soviet maps were known to contain non-existent streets – to confuse enemy soldiers
  - Commercial maps contain so called „EasterEggs" (non existent streets, misspellings in street names etc.) to track illegal copying of map data

Supply:

- The ability to draw diagrams and plans is quite pervasive in literate societies (however this do not apply to drawing maps which would be accurate in geographical sense, in other words – not many people are natural born cartographers) .
- Just as it is relatively easy to continue writing a text commenced by other person, it is also easy to complete a drawing sketched by someone else, so the content can be created collaboratively.
- Drawings capture human knowledge (maps, blueprints, activity diagrams, organizational charts etc.) in a way not dissimilar to text. After all a „professional" map of the world is also a sum of work of many cartographers.

- It is very popular activity (you probably drawn your first drawing several years before writing your first letter) so the amount of potential, useful content is high.

Some problems can be also identified. First of all the plans, that most people are able to create are often inaccurate (in GIS sense). However, as popular experience shows in most cases such plans are sufficient for orientation and route planning, as long as intersections, and relative road lengths are sensible. The plans can be also very different, as our mental representations of geographical surroundings might be also different. However this might be also seen also as an advantage as it creates a diversity of opinion that is required for the emergence of "wisdom of crowd" effect [Goldstone05]; perhaps also such various mapping perspectives might turn out as a useful feature. One can supply a lot of useful, but different between two people, information on a map – examples may include favorite pubs, tram stops, wifi hotspots etc. It is also possible to be able to represent even different versions of the street layout, for example for historians, who want to create maps of eg. Victorian London and compare them with current maps.

It seems therefore the, as the supply and demand are out there, the missing piece remains a collaboration tool, that would guarantee a low entry barrier for all people willing to share their geographical knowledge. Such that would deal with inaccuracy, exploit diversity of „casual" map data that ordinary people (not GPS fanatics) can provide and which will allow to store & retrieve a variety of map related data. Our team at the Institute of Computer Science of the Warsaw University of Technology started to design and implement a prototype of such system, calling it with a temporary name *Wikiplan*.

3 A Wikiplan system

3.1 Technology issues

As the most important property of a editing system seems to be its low entry barrier to new users, it should ideally be a web browser based tool, not requiring any download. This is a difficult task, as web browsers are tools for displaying (and sometimes editing) mostly textual contents and not vector data.

However, with the recent advent of AJAX technology (Asynchronous Javascript and XML) it became possible to create browser based applications that behave in a way not dissimilar to desktop programs. Examples of such software include primarily Google offerings such as GMail or Google Maps, however AJAX is being used more and more frequently, sometimes credited as a technology that might potentially lead to a whole new category of Internet experience for end users, dubbed Web 2.0.

Using AJAX it is possible to create a web page that would dynamically react to user input in real time. In order to implement a map drawing tool, an ability to display vector graphics is also required. Such requirement is much more difficult to

fulfill, as the most popular web browser – Microsoft Internet Explorer version 6.0, does not support vector image display. However other popular browsers (beta version of Internet Explorer 7, Firefox, Safari) support the SVG vector image standard, and it is possible to install a plugin in IE6 giving it the same functionality.

Summing up – using AJAX and SVG support for browsers it is possible to create a diagramming tool that does not need external applications. A proof of concept of such design may be for example Ajaxsketch [Ajaxsketch]– an online vector editor, compatible with popular Firefox web browser.

Of course apart from a good user interface, a system backend is also necessary for data storage, user management, versioning & conflict resolution etc., but such being a server side application, it can be implemented in a well established technology such as Java.

3.2 Prototype description

The Wikiplan prototype is being developed with several assumptions, such as:

- **portability** – the system backend is being implemented in Java (with some additional tools such as Hibernate for object persistency) and thus is immune to incompatibilities between operating systems;
- **graphical user interaction** – the main principle is simplicity here, as it is crucial to create a very low "learning barrier" for new users in a manner not dissimilar to traditional wiki system; to this end a graphical user interface is being implemented as Ajax+SVG system mentioned above with a minimalist approach to interface design; currently the system is being tested only on Firefox, with planned support for other web browsers at later stage of development;
- **openness** – the system is being developed in open source model from the start and will remain free software;
- **ability to incorporate into existing textual wiki systems** – the Wikiplan system is being implemented in such as way, as to facilitate its integration with other wiki systems, thus allowing creation of the web site where people can edit both text and maps (see next chapter for example applications of such combination);
- **automation** – the system is able, to certain extent, to automatically resolve conflicts in data submitted by users (for example when two people think that a given street is in two different places) and compute approximate properties of geographical object when no sufficient data is provided by a user;
- **versioning** – a system is storing detailed information about all changes introduced into a given map, allowing to reverse ay editing action and display map history – analogously to history mode of traditional wiki systems.

Considering the internal processing engine, the main assumption is that geospatial accuracy is not required (however it is possible, when sufficient data is provided). A

coordinate system that is being used is flat, so no geographical projection is applied. Such system is perfect for city plans, but of course at the expense of prohibiting the creation of larger maps, where curvature of the Earth would become relevant.
Each map object has associated a certainty level, which is highest for GPS data and lowest for objects that can be dubbed as "prototypes" – i.e. users know that such objects (say point of interest) should be placed on a map, but have no idea about their exact location. Consequently every user can position objects in a place, that seems right to him – or „confirm" a proposal of other users, and the system is to calculate average properties of objects based on above inputs. Obviously, the more users confirm the location of the object, the higher their certainty level.
It is possible to create objects with undefined properties. A good example would be a street with unknown length, but which is known to join two other streets. The system stores information about such object using the mechanism of "anchors" – metadata that captures the allowed spans of other properties for an object and marks these properties which are certain. In a situation presented on a Figure 1, a Parkingowa street object will be associated with anchor objects, defining possible placements of both ends of a street (and joining it to these streets at the same time). This way we store *certain* information (that Parkingowa joins Nowogrodzka with Jerozolimskie between two particular crossroads) and inaccurate information (such that we do not exactly know where Parkingowa is located).

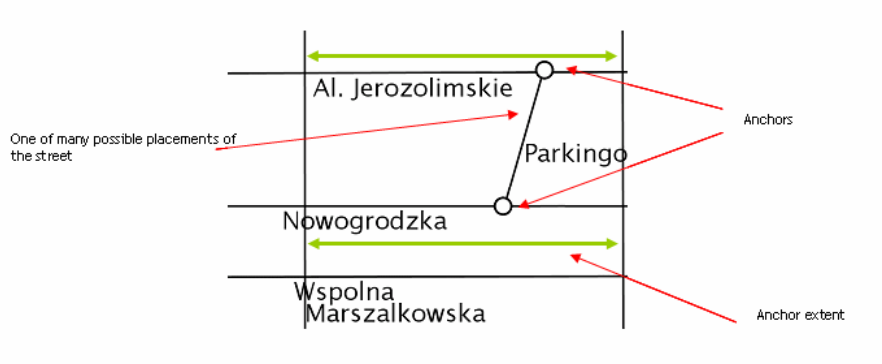

Fig. 1. Illustration of anchor objects

The system is also equipped with a variety of wiki-like tools. There is possibility to create comments and discuss changed made to the map, every user can create his own layer (both for streets and points of interest – this allow creation of several versions of a single map). For professional use (or for populating the system with readily available data) an additional Java editor is provided, which can also import and export map data in several formats.

4 Applications and conclusion

There are many possible applications for a system such as Wikiplan. First obvious area of application is of course acquisition of base geographical data in situations where such data (i.e. information about streets etc.) is not readily available. However, even if such data is available (as is often the case in eGovernment applications, when the basic map data pertaining to street layout is free for use for government institutions), the Wikiplan system can be used to augment it with community created data about points of interest – such as most interesting restaurants, dangerous places, best shops, best views etc.
Thanks to system capabilities people can collaboratively create information resources containing geographical metadata – such as travel guides or "best in the city" listings. Additionally it is also possible to share information of a spatial nature (such as – the favorite bicycle trips, locations of favorite shops etc.) with others, in a way similar to sharing own though on a blog site, or personal photos in services such as Flickr! [Flickr].
Other possible applications might include:

- Street plans, country borders, highway maps editing on Wikipedia and similar systems based on Media Wiki software;
- Augmenting Google / NASA satellite imagery with street data;
- Augmenting „professional" geospatial wiki system (such as Openstreetmap) with a „rapid prototyping tool";
- creating diagrams of machinery, circuits etc. (however for this task a much simpler editor is probably more appropriate – not necessarily dealing with uncertainty, but still Ajax+SVG based)
- creating databases of POI, for example linking all wikipedia entries dealing with a given POI
- allowing people to create their own sections of map data – for example „best pubs in London and how to get from to another"
- creating standalone open source applications (route planning software, postal code databases etc.) using free map data.

The system is still far from completion, but the functionality that is already implemented allowed us to create a prototype wiki system that allows to create a database of the most respected pubs in vicinity of the University with information about commuting between them. When completed, we hope that it will enhance popular wiki systems with ability to share "graphical" knowledge and thus further increase the amount of free (as in freedom) information available to humanity.

References

1. W. Cunningham, Wikis Then and Now, Intl. *Wikimania Conference, Frankfurt*, 2005.

2. Flickr!, http://www.flickr.com.

3. R. Gibson et al., Google Maps Hacks: Tips & Tools for Geographic Searching and Remixing, O'Reilly, 2006.

4. R. Goldstone et al., Computational models of collective behavior, TRENDS in Cognitive Sciences Vol.9 No.9, 2005

5. Google Maps, http://maps.google.com.

6. Y. Haraguchi, et al, The Living Map-A communication tool that connects real world and online community by using a map, Journal of the Asian Design International Conference, 2003.

7. M. Krötzsch et. al. Wikipedia and the Semantic Web – The Missing Links, Intl. Wikimania Conference, 2005.

8. Open Stret Map, http://www.openstreetmap.org.

9. Placeopedia, http://www.placeopedia.org.

10. S. Shah, Controversy *and Stability: How Wikis Have Productive Conflict*, Intl. Wikimania Conference, 2005.

11. J. Surowiecki, *The Wisdom of Crowds - Why Many Are Smarter Than the Few, and How Wisdom Shapes Business, Economics, Societies, and Nations*, (Doubleday, 2004).

12. Wikipedia, http://www.wikipedia.org.