

Word cloud visualisation of locative information

Bjørnar Tessem, Solveig Bjørnstad, Weiqin Chen & Lars Nyre

To cite this article: Bjørnar Tessem, Solveig Bjørnstad, Weiqin Chen & Lars Nyre (2015) Word cloud visualisation of locative information, Journal of Location Based Services, 9:4, 254-272, DOI: [10.1080/17489725.2015.1118566](https://doi.org/10.1080/17489725.2015.1118566)

To link to this article: <http://dx.doi.org/10.1080/17489725.2015.1118566>



© 2015 The Author(s). Published by Taylor & Francis



Published online: 22 Dec 2015.



Submit your article to this journal [↗](#)



Article views: 123



View related articles [↗](#)



View Crossmark data [↗](#)



Word cloud visualisation of locative information

Bjørnar Tessem^a, Solveig Bjørnestad^a, Weiqin Chen^b and Lars Nyre^a

^aDepartment of Information Science and Media Studies, University of Bergen, Bergen, Norway; ^bDepartment of Computer Science, Oslo and Akershus University College of Applied Sciences, Oslo, Norway

ABSTRACT

This article describes an explorative study of a prototype for visualisation of locative information from Wikipedia. PediaCloud is a smartphone app that uses *word clouds* for graphical display of links to text and photos relating to a particular place. This approach to accessing location-based information is different from more common approach of using interactive maps to visualise the information. PediaCloud gives you a word cloud representing your location, and the words are links to Wikipedia articles. In addition to reading the Wikipedia entries, you can re-centre your information probe by getting a topical word cloud weighting the located information content in relation to that particular word. This results in word cloud visualisations that could bring the user into a variety of topics. In this article, we explore the potential use value of a service like this and identify opportunities for further design and development. PediaCloud was tested in 2014 in a field trial in London, a big city with large amounts of located information, with eight participants from the city. The data collected indicate that PediaCloud will be experienced as an explorative locative service, where you discover nearby information by coincidence rather than after searching for it instrumentally. The analysis has explored the nuances and potentials of this 'sideways' search that is embedded in PediaCloud, and shows how this can be considered a serendipitous search approach, giving the users information in a manner that can be considered fun, surprising and interesting.

ARTICLE HISTORY

Received 27 March 2015
Accepted 4 November 2015

KEYWORDS

Semantic web; locative information; word cloud; field trial; user experience; serendipitous search

1. Introduction

The smartphone user of today is to a large extent using location-based services to navigate in and gather information about the nearest geographical surroundings. This is very often done through the use of interactive maps and information found on those maps. This is so because maps are a low threshold tool for getting an understanding of your surroundings, providing a visual presentation of the geographical aspects of the place you are at. Annotations (typically commercial services like restaurants and shops) on the map provide additional information that has potential value for the user. Street view services increase the usefulness of maps by giving systematic access to pictures taken at locations in a city or on a street.

CONTACT Bjørnar Tessem  bjornar.tessem@uib.no

© 2015 The Author(s). Published by Taylor & Francis

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Word clouds is another visualisation technique that has a completely different goal as it provides the user with an overview of the content of a collection of texts. The word size (as in font size) indicates how often the word occurs in the text collection, normally after having removed so-called stop words. Potentially, with an increasing amount of texts annotated with locations attributed to a text, word clouds of those texts could possibly add to the information found in maps, and even provide a different perspective on a location than the one found in a map.

The PediaCloud smartphone app is intended as a tool that does exactly that. By using the position of the smartphone to collect position annotated texts from DBpedia (a project where semantic tagging is used to give more meaning to Wikipedia articles), we provide users with an entrance to location-based information which is different from a map, and possibly gives the user a different experience in searching for that information. However, PediaCloud also gives access to information related to a particular located DBpedia entry by using DBpedia's semantic web links. This has the potential of giving the user a richer understanding of the location, by providing access to information that is location relevant but not necessarily located itself.

To understand how such a tool is perceived by users we have conducted an empirical study, aiming at how users are able to access location-based information through word clouds, what information they find interesting, whether they find the information in some sense useful and whether they have feedback regarding how word clouds provide access to information.

In the next section, we provide background about locative media services, the semantic web, DBpedia and word clouds. We continue with a description of the information model and the computational approach chosen in PediaCloud. Screenshots are used to show the information layout on the mobile phone display. We then present how we conducted the field trial and present analysis results relating to how the technology was used, the impression of the users and some suggestions for improvements. In total, our analysis suggests that PediaCloud is a little buggy, but fun and whimsical app compared with utilitarian apps like Google Maps™. Further, we discuss the analysis and suggest that PediaCloud can be coined by the serendipitous, before we conclude.

2. Background

2.1. Location-based information

Mobile applications have perhaps become the most pervasive kind of software, as we now communicate through our mobiles supporting text messages, emails, chats and video; we engage in social networks through our mobile; the mobile is used to get access to commercial and public services; and we use it to play games, listen to music and take photos. Arguably, the mobile has taken over the role of the traditional desktop or laptop computer for day-to-day use of software, and it has a significant supporting role when we search for information both for work and leisure.

Location is a main focus in many applications we use on the mobile phone, as they have the potential to support your information needs at your current location, also termed geolocation. Siri™¹ and Google Now™² are well-known tools that include and use geolocation

to provide you with tailored information. These tools are to a large extent based on user modelling to give you what is considered most relevant for you at the moment.

Tools that have an even stronger focus on location are map services, typically realised in a tool like Google Maps™, a tool that delivers maps of your current location and information about near-by services, buildings and others. Google Maps™ is also providing APIs that make it possible to integrate these map services into other mobile applications. And if you look into the abundance of location-based software available for smart phones the use of maps as a tool for visualising information and providing access is the most common, although approaches like mobile augmented reality are also studied in research and made available on the App market (Höllner and Feiner 2004).

Relevant in this context is exploratory search (Marchionini 2006) as a user-controlled search for information in a particular context, but also when you are not exactly aware of what you are looking for. It includes a freer, leisurely search for fun facts or search for interesting and useful knowledge related to a particular task. However, the interestingness of the information resources found is not necessarily given on basis of specified user preferences or previous behaviour. With the restricted user interface of the mobile, where the possibility to write queries to a search engine is hampered by small screen keyboards, visual tools to aid in exploration are helpful. Visual tools for exploratory search have been studied in the context of traditional computers with large screens, e.g. (Dörk, Williamson, and Cappendale 2012), who present a browser-based explorative search tool supplementing the user with visualisation tools like maps and word clouds. Neumann and Schmeier (2013) have developed tools for mobile exploratory search, but are mostly concerned with the limited user interface and not location-based information in particular.

2.2. Word clouds

A common tool used to visualise text data is a word cloud. Word clouds give some structure to relevant information, by providing the user with terms that indicate what is important. The idea is that the words with the largest fonts are the most important, e.g. as a result of word counts, something that may be modified by some additional criteria. Research on the use of word clouds as a tool to support information search has shown various effects; e.g. (Rivadeneira et al. 2007; Hearst and Rosner 2008; Gwizdka 2009; Wilson, Hurlock, and Wilson 2012). Schrammel, Leitner, and Tscheligi (2009) placed semantically related terms close to each other in the tag cloud. Interestingly, their results suggest that this structuring of topic areas gives a positive effect on the usability of word clouds. Zhang, Qu, and Wang (2011) used eye tracking to study the impact of font size and location on the affinity of words. Larger and more centralised words got more attraction from users. Word clouds also seem to be better when having search tasks of a rather open type compared to specific information searches (Sinclair and Cardew-Hall 2008). This suggests that word clouds could be more useful in exploratory search tasks, which often are formulated in a rather general form.

2.3. Semantic web and DBpedia

Semantic web technologies are relevant in this context as they contribute to connect information sources to give a potentially richer picture. For example, while Wikipedia³ is a general source of knowledge that often is a site for our first searches, DBpedia⁴ uses semantic

technologies to tag and structure information from Wikipedia articles, making it possible to apply semantic technologies to information search. DBpedia, and its use of semantic web technology, provides us with a fundamental technology for the development of software that organises the material and guides a user through the amount of knowledge to be found in Wikipedia.

Mirizzi et al. (2010) exploit a collection of semantic and information retrieval technologies to give users a visual presentation of articles in Wikipedia and their relations to other articles. The DBpedia RDF graph structure, textual analysis of DBpedia abstracts and external information sources are used to establish connections between articles in Wikipedia. The visual tools allow the user to navigate through the Wikipedia content with the aid of the computed links. Becker and Bizer (2009) connect DBpedia resources with additional data from the semantic web to give the user geospatially relevant information in a browser. Veres (2012) improves on this providing an iPad application that presents the user with a more usable interface, and links to semantically tagged information sources relevant for the location. Veres uses a map with Wikipedia articles positioned on it as the main visualisation tool. Church et al. (2010) show an example of exploratory social search allowing mobile users to look at and modify other mobile users' searches to get timely and locative information, focusing on maps as the visualisation tool. Kleinen, Scherp, and Staab (2013) provide a mobile application, 'Mobile Facets', that allows real-time and interactive search and exploration of resources such as places, persons, organisations and events originating from different, integrated social media sources like DBpedia, Eventful™, Upcoming, Flickr™ and GeoNames. Their research applied a participatory design methodology and focused on how the information is organised and displayed on the small display of a mobile phone.

3. The PediaCloud tool

The aim of PediaCloud is to establish knowledge on how we can provide a user with access to information relevant for the user's location. The goal is not necessarily to provide this information in a very structured manner, but to instead use word clouds as a visualisation technique that makes the search more interesting and enables a richer understanding of the location. The procedures of PediaCloud are to some extent inspired by the sociologist Gieryn (2000) who defines the concept of place as something that (1) has a geographical location (2) is filled with material form and (3) is invested with meaning and value. A word cloud would certainly not support the first of Gieryn's levels of place, might support the second level to some extent, and may have the potential to give smart phone users richer understanding of what a place is. Similar ideas as those of Gieryn are described by Bilandzic and Foth (2012) who use the concept translucent hybrid space for how physical locations are now enriched with digital information and social tools that have transformed our interaction with the locations.

A natural candidate for located information in the form of texts is Wikipedia articles, and further their representation in DBpedia. The richness of DBpedia, with more than 1.000.000 resources referring to geolocated English Wikipedia articles, ensures us that we have a large collection of documents to test out our ideas on. The localised resources of DBpedia have an abstract, which essentially is an extract of the first few paragraphs in the corresponding Wikipedia article. A typical abstract looks like

Piccadilly Circus is a road junction and public space of London's West End in the City of Westminster, built in 1819 to connect Regent Street with Piccadilly. In this context, a circus, from the Latin word meaning 'circle', is a round open space at a street junction. (DBpedia's Piccadilly Circus abstract)

Wikipedia articles themselves may be lengthy, and to obtain the texts and extract words, significant communication time on mobile broadband and significant computational overhead may reduce usefulness of a tool. Therefore, we chose to use the DBpedia abstracts as the source of located texts.

First of all PediaCloud should identify located information sources close to the location of the user. But to enrich the understanding of a location, a tool like PediaCloud should not only identify nearby located information, but also information that can be said to be related to that location. Although the meaning a person assigns to a place is connected to that person's experiences and attitudes, it is also constructed from the cultural and natural non-located phenomena related to the location. For instance, in London, we have the street Abbey Road, which contains the 'Abbey Road Studio', a recording studio where The Beatles made several of their popular records. The knowledge about Beatles is not really about any particular location, but adds meaning to the material form we find around us as we walk along in Abbey Road and observe a street crossing depicted on one of Beatles' record covers and the studio. PediaCloud thus collects from DBpedia articles that have a location (primary resources), and articles that have some relation to the primary resources (secondary resources). We have implemented this secondary relation by using semantic web relations found in DBpedia. Such relations are represented as RDF triples expressed as ' $a R b$ ' where a and b are objects that are somehow related to each other through the relation R . Figure 1 illustrates the concept of located and secondary resources.

As shown in Figure 1, the centre indicates the user u 's position, and the located articles are illustrated with points at their locations a_1, a_2, a_3, a_4 . In addition, b_1 and b_2 illustrate secondary resources. Located resources are found in DBpedia through a SPARQL query where we search for RDF triples of the type ' $a \text{ lat } y_a$ ' and ' $a \text{ long } x_a$ ' where a is a primary resource and lat and long are the relations for latitude and longitude of spatial things as defined in the wgs84_pos namespace for the semantic web (http://www.w3.org/2003/01/geo/wgs84_pos#). For the sake of efficiency, in the query, we actually ask for those located resources a that lie within a rectangle enveloping a circle around u . Those resources lying outside of the circle with a radius d_{\max} are filtered away at the smart phone in a later step. Secondary resources are found in DBpedia by querying for secondary resources b that satisfy ' $a_i \text{ rel } b$ ' or ' $b \text{ rel } a_i$ ' where rel is

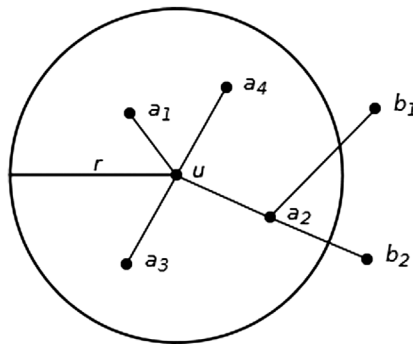


Figure 1. PediaCloud's model for located (primary) and secondary resources.

any arbitrary RDF relation used by DPpedia, and a_i a primary resource. The construction of the word cloud essentially goes through the following steps:

- (1) Get user location u . Set d_{max} to 250 m.
- (2) Get all articles (primary) a_i within radius d_{max} of location.
- (3) Get all articles (secondary) b_j that some primary article link to.
- (4) If collection D , comprised of identified located and secondary resources, has less than 30 resources, double d_{max} and repeat from step 2.
- (5) Find frequency $c(t_k, a_i)$ of each word t_k in each resource's abstract where the abstract abs is found through DBpedia's RDF relation ' a_i abstract abs '.
- (6) For each located resource a_i find its weight $w(a_i) = \max(0.0, (d_{max} - d(u, a_i))/d_{max})$.
- (7) For each secondary resource b_j find its weight $w(b_j) = \prod_i w(a_i) \cdot \text{sim}(a_i, b_j)$ where there is a relation rel from a_i to b_j , $\text{sim}(a_i, b_j) = \cos(T(a_i), T(b_j))$, and $T(a)$ is the word count vector of document a . If a document b_j is linked to several primary sources, it adds to the weight for each of those, as many parent resources indicate a higher likelihood of relevance. An alternative could be to use the maximal contribution for b_j given the parent resources.
- (8) For each word t_k let $w(t_k) = \sum_i w(a_i) \cdot c(t_k, a_i) + \sum_j w(b_j) \cdot c(t_k, b_j)$.
- (9) Modify the $w(t_k)$ for words by multiplying with an *IDF* (inverse document frequency) for each word based on the collected texts, i.e. $IDF(D, t_k) = 1/\log(|\{a_i \text{ in } D: t_k \text{ occurs in } a_i\}|+1)$ and $w'(t_k) = w(t_k) \cdot IDF(D, t_k)$ is the modified word weight. *IDF* is used for the same purpose as in traditional document retrieval; words that are found in all resources in a collection do not contribute much information, and should be given less score.
- (10) Select the highest n (around 50) scoring words. 50 words seem to be the maximum of what we can show on a smart phone screen, without making the smallest words unreadable.
- (11) Use the computed word scores to create a word cloud for current area.

When the word cloud is presented to the user, he or she may click on one of the words (Figure 2(a)). The app will then show a list of possible Wikipedia articles to be read (Figure 2(b)). They are ordered according to how much they contribute to the total score of that particular word, i.e. $w(a_i) \cdot c(t_k, a_i)$. To the right of each article, there is an icon representing a *re-centring* function. Essentially clicking the icon makes the topic of the corresponding article to a centre, and a new word cloud is made from the collected DBpedia abstracts, but with a weighting based on closeness to the new centred article e , i.e. we compute new weights $w(a_i, t_k) = c(t_k, a_i) \cdot \text{sim}(e, a_i)$, and redraw the word cloud. Figure 2(a) shows the word cloud found at location by Piccadilly Circus in London. The nearest located article was about 'Piccadilly Circus tube station'.

Note that the algorithm always will end up with a document count larger than 30 articles. This is ensured by doubling the d_{max} if the collection is less than 30. In rural areas with few localised resources this will lead to few changes in the word cloud as we walk around. The collection of collected articles will very often be identical, and the relative distance to a located article will not change much due to the larger d_{max} . In urban areas, where the number of nearby articles is high, the effect of moving is much bigger. It may remove a significant amount of articles, add some new and the effect of d_{max} is much bigger. Figure 3 shows a new word cloud found when moving 500 m to the north-west from Piccadilly Circus.

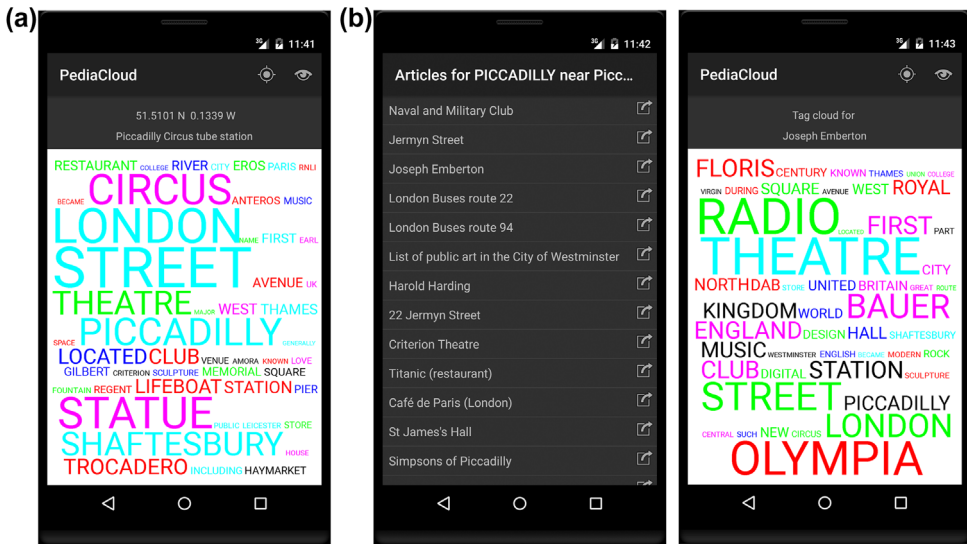


Figure 2. Word cloud visualisations generated at Piccadilly Circus, London. (a) A word cloud for Piccadilly Circus, London, (b) the articles relevant for the word 'PICCADILLY', and the re-centring of content around the topic "Joseph Emberton".



Figure 3. A word cloud at Cambridge Circus, London – 500 m NE from Piccadilly Circus.

If we compare with 2a, we observe that the word 'LONDON' remains, whereas 'PICCADILLY' and 'THEATRE' disappears or gets smaller. Instead, the words 'SOHO' and 'RESTAURANT' gets significant font sizes.

It is also worthwhile mentioning that our aim when constructing PediaCloud was that the tool should run its own computations when creating the word cloud. This has the consequence that the only server that is accessed for data is the DBpedia RDF triple end point. This will normally lead to the downloading of large data-sets, with resulting slow response times when the network connection is not good. At the same time, this also means that no logging of use in servers occurs, and the use is completely private for the user; a model for mobile computing services rarely seen.

4. Method: field trial in London

A field trial is a research method where examines an intervention in the real world, e.g. using a particular service out of the laboratory. Field trials have become particularly relevant in studies of mobile systems (Jensen 2011), as outcomes are observed in a natural setting rather than in a contrived laboratory environment. The field trial that we conducted had two main goals: first; to describe how PediaCloud is used and experienced by the users and second; to ask about ideas for improvements in further iterations, making this an explorative, formative study.

In order to achieve these goals, we collected both qualitative and quantitative data from the users. London, a dynamic metropolitan area, was chosen as the location for conducting the evaluation because it has high density of locative information in Wikipedia, which is suitable for such a study. Through careful screening, we recruited eight participants, aged between 29 and 49 years old ($M = 36.38$, $Std = 5.89$), four males and four females. All participants had Android mobile phones (version 4.1.2 or newer), and were interested in and had good knowledge about new technology, devices and software solutions.

We divided the field trial into two parts. The first part involved exploration of PediaCloud. The participants met up at the university for a 45-min briefing session. They were given a short introduction of PediaCloud and of the field trial, and they signed a consensus form. PediaCloud with a log function, installed for this experiment only, which could record the location and interaction between the user and PediaCloud were installed in their mobile phones. The participants were then asked to use PediaCloud for up to a week exploring the different functions and features with emphasis on four special types of locations: famous/well-known locations or tourist attractions, familiar locations (e.g. locations close to home or workplace), unfamiliar locations (e.g. locations never visited before) and self-chosen locations (e.g. locations connected to sports, entertainment, shopping or other favourite activities). The usage of PediaCloud during the exploration week was logged on their mobile phones.

After the exploration week, they were individually invited to a debriefing session at the university. Each session lasted for 45–60 min and consisted of three sub-sessions. The first sub-sessions were semi-structured interviews focusing on the participant's experiences with locative services and word clouds in general.

The second sub-session focused on how the participants interacted with PediaCloud and usability issues. We provided a standard mobile phone with PediaCloud with a pre-set location. The participants were asked to perform some actions in PediaCloud step-by-step. After each step, they were encouraged to explain their decisions and express their opinions

on the functions provided by PediaCloud. At the end of this sub-session, the participants were given opportunities to talk about the usability issues in PediaCloud that they have encountered during the exploration week and in this sub-session.

The last sub-session in the debriefing is another semi-structured interview where participants were asked to discuss their experience with PediaCloud in relation to other location-based applications such as Google Maps™ and Foursquare™. In addition, privacy issues related to personalised location-based services were also discussed.

In this debriefing session, with the participants' consensus, audio data were collected for the interviews (sub-session 1 and 3) and video data were collected for the interaction (sub-session 2). Each sub-session was conducted by two researchers, one was the main interviewer and the other was an assistant. It was the debriefing session that provided us with most information about the value of the app. In explorative studies of this kind, normally the qualitative data will be more relevant in order to inform about the potential about the technology.

Finally, log files were retrieved from the participants' mobile phones for later analysis. The participants could choose whether they wished to have PediaCloud removed. We uninstalled PediaCloud from their mobile phones in the latter case.

When the respondents were trying out PediaCloud, we logged when they started the app and/or got a new position, whenever they clicked a word, a web page or the re-centring function, as well as other less frequent activities. The log files were collected after the qualitative interview, and were processed to be used for statistical analyses. A total of 253 events where the user clicked a word were recorded, in addition to 171 choices of Wikipedia pages. Figure 4 shows all the sites where we logged use of PediaCloud. As expected most of the uses are found in Central London.

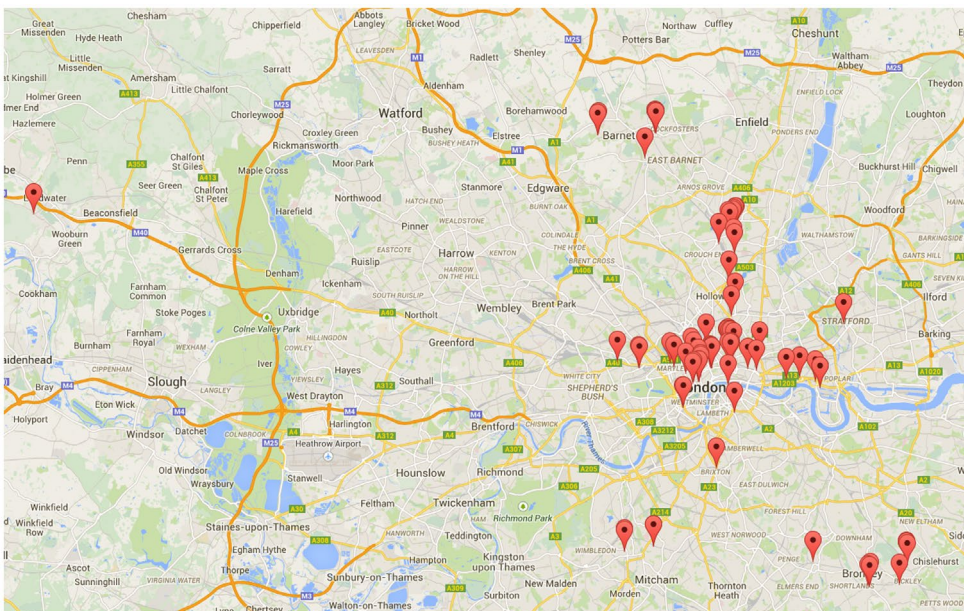


Figure 4. Sites where the respondents used PediaCloud.

5. Analysis of informant responses

In this section we present an analysis of the informants' descriptions, statements and judgments about PediaCloud. We are particularly concerned with the informants' *reasoning about functionality*. We start with a focus on how they approached the word cloud and chose their words, and then we continue with the choice of article, and how they used the re-centring functionality. The analysis concludes with reflections about the use value of PediaCloud, and suggestions for improvements.

5.1. Using the word clouds

When facing the word cloud in PediaCloud, all the participants based their choice on the actual meaning of the words in the cloud. Although they noticed the size and the colour, only two participants responded that they made decisions based on the size of the words at some locations, particularly unknown or unfamiliar locations, or when they were curious about what the small and unexpected words would bring in familiar locations. In both those cases, the chosen words were rather the words with small fonts than those with a large font. The participants mainly chose words according to their interests or words that they were curious about and wished to investigate, for example, words whose meaning is unknown to them. For example, one participant responded

[...] if I was in a place that I already knew, there would be a lot of words that made sense to me, say I was round the corner from a Tube station, the Tube station name would be there. But if I saw the name of a person or a word and I couldn't see any obvious association to where I was, I would pick that word to try and find out what it was. (Cathy)

It seems that for familiar and well-known locations, unknown, unexpected, unusual words and words that they would not normally associate with the locations were often chosen to explore further.

Still there were examples of the opposite type of choice. One chose the word 'JUSTICE' and responded to a question about why: 'Just because to me that's what Temple is all about Law Courts. It's kind of the hub of that industry in London, in the UK, probably. And it's interesting to see what comes through from that.' (John)

One of the participants found the verbs in the cloud intriguing. She was curious about what the verbs would bring for her.

A lot of the time, I found that there were verbs. So it wasn't necessarily the name of a place, or the name of someone, or a noun. It was a verb. And that was intriguing, because I thought, What are they going to find for me? What am I going to find? (Anna)

After trying several verbs, she decided that the nouns, such as name of a place, a person or an event, would bring more interesting information.

At each click on a word, we logged information about the word and the rest of the word cloud, i.e. the colour of the words, the font sizes and the locations on the screen. To study how word sizes influenced choice, we chose to use the clicked word's font size rank in the word cloud. By font size rank, we mean the number we get from ranking the words in the word cloud according to their font size, and use the position of the chosen word in this ranking. This because screen sizes were different for the users, and therefore also the size of the fonts. The data (see Figure 5) shows how the rank of the chosen word is distributed (ChosenRank). The median was at the ninth largest word. The cumulative diagram in Figure

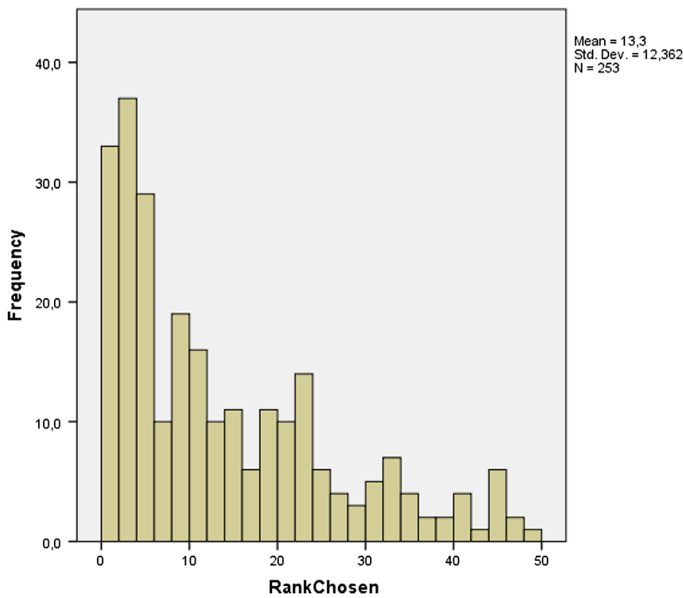


Figure 5. Frequency of words chosen according to their font size rank in the word cloud.

6 shows the count of words with rank less than or equal to the particular rank value. This diagram shows that the respondents in more than one out of four clicks chose one of the three largest words. On the other end of the scale, a word among the 40 % smallest were chosen in 13 % of the clicks.

The words were coloured using six colours, red, magenta, green, blue, cyan and black. There was no significant difference regarding colour of the clicked word, although blue was the most chosen colour. We also logged the x- and y-coordinates of the words. We tested

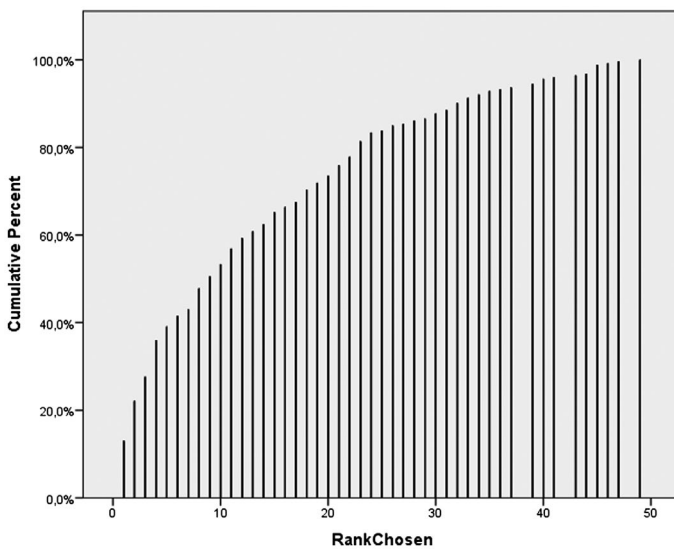


Figure 6. Cumulative percentage for rank of clicked word's font size.

statistically whether there was any significant variation in the placement of the selected words on the screen, and found that there was a small, but significant ($p = 0.002$) tendency to select words at the top half of the screen.

It is hard to interpret the quantitative data as we have no control group data, but it is evident that word size influence a lot, colour does not mean much and placement on the screen has some, but little effect on choice of word. The qualitative data told us that users had a preference for choosing smaller words, as many of the largest words was seen as not surprising and therefore not as interesting to look into. The observation that one of among the 40 % smallest words were chosen in as much as 13 % of the clicks may support this.

5.2. Selecting Wikipedia articles

After choosing an interesting word and clicking on it, the user is shown a screen with possible Wikipedia pages that could be read. They are ranked according to their contribution to the particular chosen word in the word cloud. During the interview, several of the respondents were shown a list of articles containing one article about the 2nd Earl of Essex. They didn't understand his relation to the place at the Strand, and chose this as their example article. This article was both high in the lists presented by PediaCloud, but also it was something that made them curious. One said about selecting articles:

So I'd look at the words that were not obviously relevant to the area. But they all – even if it wasn't immediately relevant – they all seemed to be relevant to the location albeit not in a direct way that I would assume they would be. (Thomas)

But the Wikipedia articles could also serve as a reminder of locations and topics they knew about, but now wanted to check out because of the reminding. One explained: 'It says 'Clement Danes'. And yeah, the only reason I clicked on that is because it's a school near where I grew up as well. So it was a word that jumped out.' (John)

Parts of London are so strongly characterised by the same type of information that they become 'topical locations' like around railway stations. The very abundance of information relating to a particular topic at a place can be somewhat frustrating for the users. One respondent was standing close to Brixton Academy, a music venue where many famous artists have been playing. Many of the rock bands that had played at the venue became secondary DBpedia resources picked by PediaCloud. This contributed to create a collection of articles that were very much about music and little else. This is an effect we also saw close to railway stations where there will often be a lot of material about railway related topics. To the user, the information collected and Wikipedia pages presented seems to be skewed towards a particular, perhaps esoteric topic.

One of the respondents also mentioned the consequence of using the tool in more rural areas, commenting on how random the located content seems to be in such areas:

I suppose just because of the underlying Wikipedia, what's there, when you get to those areas, it's fairly thin. There's almost no references to things. The bigger words, the smaller words. It's all a bit random: there's no great logic it seems behind it. (John)

To summarise the respondents would be inclined to choose the most surprising Wikipedia articles. Most of those would come from the secondary resources picked by PediaCloud, and those were found interesting. Still, this could be seen as problematic as it also lead to a too strong focus on very narrow topics at some locations.

5.3. Re-centring the word cloud

The use of word clouds to find Wikipedia pages seemed to be quite intuitive for the respondents. However, using the re-centring function was somehow confusing for them.

However, many of them had tried it a lot, and had some interesting experiences. One respondent was at the Kings Cross railway station and chose to re-centre based on the word Paddington, which is also a railway station. 'At Paddington when it showed the word WAR, I thought, like, "Really?". Cause that's interesting. And I clicked on that' (Emma). Here, the use of the word clouds diverted the respondent away from the location information and the information sources related to the word WAR was becoming more prominent.

Most of the respondents thought of the function as a way of narrowing down search, although it could more be seen as a way of moving around in a collection of information sources. One said:

It makes sense in terms of going deeper into the information and navigating through it once you've found something that you're interested in. Re-centre the word cloud on the thing that you're actually looking at so that it kind of moves with you. (Rudy)

In the same line we have one respondent who found it to be a somewhat different experience to use a word cloud related to the Wikipedia entry 'Inn' than just restarting a search through the use of the 'back button': 'I focused on Outer Temple and Inns, going back to Inns. Just to explore what inns were about, is something that other than pressing the 'back' button. Just interested in more associated with "Inns"' (Ben)

As a summary, we are safe to say that the respondents had some interesting experiences with this function; they thought it was something for narrowing search, but experienced instead that they were moving around in the collection of information about this place, but that the re-centring changed the focus. In a sense they were moving sideways in the information space. In many situations, instead of narrowing down search, the function was diverting the user away from the location information.

5.4. Views on the use value of PediaCloud

When comparing with Google Maps™ and other popular location-based services, PediaCloud was considered by the participants as a different type of application. It is more 'interesting' and more relevant to 'learning and knowledge' rather than 'practical' and 'functional', as in Google Maps™ and other services. Most of the participants thought that PediaCloud was more informative and relevant 'in a broader sense' than 'just a map' since it links the word cloud to entries in Wikipedia and provides access to more knowledge related to a certain location. Several participants indicated that PediaCloud had the potential to be an e-learning tool. One participant summarised the difference between PediaCloud and Google Maps™:

Because Google Maps is a very tangible I'm-going-to-give-you-directions. [...] I go to it to look for specifically one thing. Whereas PediaCloud is much more creative, so it's more fun than functional. So I wouldn't go to PediaCloud to find out when the next bus was going to come, or whatever. I'd probably go to it because I was standing somewhere and thought, 'Let's see what I don't know about this local area.' I see it more as a kind of learning and knowledge thing, than a functional thing. And I'd say that Google Maps is very much functional. So the PediaCloud is more a kind of leisure thing. (Cathy)

We have earlier mentioned how the secondary resources enabled a 'sideways' movement through the information. An English term that to some extent pinpoints an experience of

random luck is *serendipity*. There were several statements that touch on serendipity. Emma said:

This app doesn't actually bring up the expected. Because if it did, that would be boring. It would be like a Google search, then. But here it took me a while to understand why this or that is coming up. The app is good in that it will only show you the unexpected. At least, I enjoyed that part of it. (Emma)

The informants actually found this serendipity to be the core function of the app. Emma again:

It shouldn't be more focused? – No, I think its unique selling point is the randomness about it. It's just stuff that you wouldn't really be thinking about. I suppose someone else could say 'Why would I care?', but that's what PediaCloud is [...] That's a good thing because you don't want it to be like any other app. (Emma)

Actually the respondents seemed to like the lack of (or different) instrumentality built into the tool. Cathy said: 'PediaCloud is more fun than functional. [...] I see it more as a learning and leisure thing, than a functional thing.'

5.5. Suggestions for improvement

The informants were slightly frustrated with the usability of PediaCloud. Some problems such as fault in a function and the response time of the app were identified by the users. However, even with the noticed problems the core functionality was not threatened or misunderstood. As Thomas said: 'For me using it daily for the tasks I'd need it for, it would have to be a bit more intuitive and smoother running, but the principle and idea are very useful!'

There were also some problems beyond our powers. These include loosing contact with GPS or DBpedia, capacity of smart phones used and the low number of Wikipedia posts in a given area. Nevertheless, a service like this must be made more stable and robust in its delivery of the core functionality. Beyond these issues, we asked our informants about the potential of the service and the app, and for ideas for how it could be improved.

One suggestion was better location-sensitivity, indicating for instance the distance to the varying located resources in the set of articles. John said: 'Is it 2 miles that way – OK I'll do that. Oh, it's 20 miles that way – I won't necessarily.'

This could be extended to a combination of maps and words, where words somehow are placed on the map at places where they are particularly relevant. For instance, Cathy suggested: 'When you're doing the PediaCloud you don't necessarily know what the radius is for that information. So it would be quite interesting if you could see points on the map around where stuff was, or had happened or whatever.'

Also using direction when showing the word cloud was suggested. For instance, John suggested that we should be able to point at something, and then a word cloud would show up. Finally we could possibly use a combination of direction and maps with words and other tags in them. Emma suggested that we had 'links to normal maps, pinpointing "You are here" with an arrow pointing in the direction you are looking'.

The other category of suggestions that we got many of was about filtering and personalisation. Thomas felt that there could be too much information at the first screen level, too much to be presented with before drilling down. He would like there to be categories or ways of filtering for history, etc.

Personally, I would prefer to have two options when I fire up the app: to have it as it is now as a general overview of the area, or have a drop down menu where you could filter on Wikipedia entries for history, leisure facilities, etc. (Thomas)

This perspective was supported by Anna:

'The idea of filtering down is there, isn't it? By moving on to the next word cloud. Or choosing your article from the list that appears when you click on a word. Maybe there should still be a filter? (Anna)

More elaborate personalisation of the information was suggested by Ben, who wants to be able to collect useful information found in the tool and potentially reuse it. This could for instance be used for pedagogical purposes:

I want to save dynamic links to specific word references. I would like to build a small amount of information that I could then re-use, and build a portfolio of what I want to see, where I want to go. And save that information by categories and locations. (Ben)

John suggested a more intelligent tool that actually would take care of the personalisation:

If it's telling me everything that I already know, then what's the point of using it? I want it to be a bit broader than that. So it kind of knows enough and then suggests things that would then be of interest to me. Without me saying 'I'm this age, I like Thai Food, etc.' (John)

6. Discussion

In this article, we explore the use value of the service provided by PediaCloud. PediaCloud gives an entry to locative information that differs significantly from maps, and this empirical study has given us some insights into how word clouds can work as an entrance to location information.

From the data we can imagine the typical use scenario for PediaCloud: The user picks up the mobile phone at a location, opens PediaCloud and gets the located word cloud. The words give the user an impression about 'stuff' that is relevant to the place, and picks a word that catches the immediate interest, more often a big word than not, but small words may be chosen also. The reason for using PediaCloud is the wish to get some interesting information about the place, less instrumental for immediate need, and more for knowledge, leisure and entertainment. So among the Wikipedia articles presented, the user selects one that is a little bit surprising and might indicate a connection to the place that was previously unknown. The user reads the article. An alternative end is that the user uses the topic of the article as a new centre to focus the word cloud. New words become more prominent, and the user may again look into what articles are connected to another interesting word.

The main observation of this story is that the use of PediaCloud is almost purely explorative. The user is open for anything that might show up, and discovers interesting 'matters' where it was not expected. The concept of serendipity has been mentioned in the analysis, and PediaCloud is related to the concept of serendipitous search (Foster and Ford 2003). The lucky identification of important information is to a large extent what makes PediaCloud an interesting approach to locative information search. The users actually made an effort to be surprised, and were often satisfied with what they got.

In contrast to the use of maps for organising and locating the information, the search is more random, but also includes more surprising content. Something which is enabled through the use of the secondary resources. The missing map and the secondary resources, as well as the re-centring function, have essentially increased the serendipitousness of

PediaCloud. Moving sideways in the available information is perhaps a phrase that would describe their way of interacting with PediaCloud.

The serendipitous user is our main observation in the empirical investigation of the PediaCloud concept. However, we also got users who would use it more for focussed information search, for instance for reminding when visiting known places, or just for pedagogical purposes. Some expressed frustration about the strong focus on a very limited topic at particular places, which would disrupt the serendipitous experience.

Several suggestions for improvements were made, for instance, an even stronger connection to the location using distance indicators on content and maps. Filtering and personalisation was also suggested. However, if you want something to be really surprising, one would expect that personalisation actually would contribute to reducing that effect. The information you get might be more useful in a short term with personalisation. However, the value of PediaCloud seems to be in its serendipity, which perhaps will disappear with personalisation.

6.1. Assessment of study

The participants were interested in and had good experience with mobile technology. They also showed willingness and spent time testing out this tool at several locations in London. Their observations and reflections are therefore of high validity and value for our understanding of PediaCloud as a tool. The less technologically interested part of the population may perhaps have different view, but at the same time, they would perhaps be less inclined to use such a tool from the start. The quantitative part of the study was from only eight respondents and hence is not in itself of general value, but still confirmed interesting results from the qualitative interviews.

At many locations, the user would find content from many domains. But, we noticed that at several places there was a tendency to get a lot of material from a particular domain like rock music or railways. It shows that PediaCloud is sensitive to the amount of information regarding a topic, and this may therefore influence the experience a user has. A filtering function might have improved this, but could also influence the level of serendipity felt by the user.

From the tool quality section, we know that the users had several problems with the robustness of the technology. This had several causes, including the availability of the DBpedia server, faulty software design, and also the quality of the smart phones of our users. However, from the interviews, it seems like this had only a minor effect on how they perceived the concept itself. We are therefore willing to claim that the analysis we have done is not weakened by this.

6.2. Improvement of the PediaCloud design

After careful consideration of the comments and suggestions from the users, we have identified a series of aspect to be included in an improved design of PediaCloud in order to make it more intuitive and informative.

- *Removing verbs from word clouds.* The results indicate that verbs in the word clouds do not provide interesting information as nouns representing names of places, persons, or events.
- *Combining word clouds with maps and directions.* Currently PediaCloud does not take advantages of maps and directions. As suggested by some participants in the study, putting words on maps and showing directions will provide users with clearer link between words and locations. Such a design will take advantages of both maps and word clouds.
- *Filtering and personalisation.* Several participants thought some kind of intelligent filtering or personalisation would make PediaCloud more useful. The words and links can be filtered based on some criteria or preferences either provided explicitly by the users, or learned through interaction with the users by PediaCloud. It is known that centralising words, and placing related terms nearer to each other have an effect in addition to word sizes (Schrammel, Leitner, and Tscheligi 2009; Zhang, Qu, and Wang 2011). Combined with a thought through use of colours, which have no impact in the random manner they have been used so far, this allows for building in some personalisation already in the word cloud itself.
- *Saving dynamic links.* Providing a saving function where users can store the words and links related to certain locations will allow users to make a portfolio and revisit the information at a later stage. Such an extension of the PediaCloud design will allow further reuse of PediaCloud information for different purposes.
- *Focus more on smaller words.* The users suggested interest in the smaller words, indicating a search for surprising content. Removing a number of the words with highest weight, and placing smaller words in the centre may create a different experience, focusing even more on the surprising content.

7. Conclusion: serendipitous search

Through analysis of user data collected on the use of PediaCloud in the explorative study we can conclude that the use of word clouds as an entry to location-based information has value for the users. The combination of located and secondary non-located information resources from Wikipedia creates an experience for the user which is different from that of map based serviced. The experience of the place supported by the provided information is surprising, random and stimulates curiosity. We have used the terms serendipity and sideways exploration of information to name this experience. For the future, we may envisage that tools like PediaCloud can be used to provide access also to other sources of information than what is found in Wikipedia, for example, journalism or advertisements. The only requirement should be that this information is geolocated at a high level of precision. Combining the retrieval of content with maps and richer visualisations may improve the serendipitous search functionality, allowing for more creative associations for the user. In this context, it is should also be worthwhile investigating the possibility for personalisation, filtering and other technologies that could enable users to have more control of the information provided; and in this way steer the topical orientation of the word clouds. This should be balanced against the value of the surprising, serendipitous information, as this is something that may disappear with solutions that are too strongly guided by user preferences.

Notes

1. <http://www.apple.com/ios/siri/>
2. <http://www.google.com/landing/now/>
3. <http://www.wikipedia.org/>
4. <http://wiki.dbpedia.org/>

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the UNIFOR/Nansen funds.

References

- Becker, Christian, and Christian Bizer. 2009. "Exploring the Geospatial Semantic Web with DBpedia Mobile." *Web Semantics: Science, Services and Agents on the World Wide Web*. 7 (4): 278–286. doi:10.1016/j.websem.2009.09.004.
- Bilandzic, M., and M. Foth. 2012. "A review of locative media, mobile and embodied spatial interaction." *International Journal of Human-Computer Studies* 70 (1): 66–71.
- Church, K., J. Neumann, M. Cherubini, and N. Oliver. 2010. "SocialSearchBrowser: A Novel Mobile Search and Information Discovery Tool." Proceedings of the 15th International Conference on Intelligent User Interfaces, Hong Kong, China.
- Dörk, M., C. Williamson, and S. Carpendale. 2012. "Navigating tomorrow's web." *ACM Transactions on the Web* 6 (3): 1–28.
- Foster, A., and N. Ford. 2003. "Serendipity and information seeking: an empirical study." *Journal of Documentation* 59 (3): 321–340. doi:10.1108/00220410310472518.
- Gieryn, T. F. 2000. "A Space for Place in Sociology." *Annual Review of Sociology* 26: 463–496.
- Gwizdzka, J. 2009. "What a Difference a Tag Cloud Makes: Effects of Tasks and Cognitive Abilities on Search Results Interface Use." *Information Research* 14 (4). <http://InformationR.net/ir/14-4/paper414.html>.
- Hearst, M. A., and D. Rosner. 2008. "Tag Clouds: Data Analysis Tool or Social Signaller?" 41st Annual Hawaii International Conference on System Sciences. HICSS'08, Hawaii.
- Höllerer, T., and S. Feiner. 2004. "Mobile Augmented Reality." In *Telegeoinformatics. Location-Based Computing and Services*, edited by H. Karimi and A. Hammas, 221–260. Boca Raton, FL: CRC Press.
- Jensen, K. L. 2011. "Remote and Autonomous Studies of Mobile and Ubiquitous Applications in Real Contexts." *International Journal of Mobile Human Computer Interaction* 3 (2): 1–19.
- Kleinen, A., A. Scherp, and S. Staab. 2013. "Interactive Faceted Search and Exploration of Open Social Media Data on a Touchscreen Mobile Phone." *Multimedia Tools and Applications* 71 (1): 39–60.
- Marchionini, G. 2006. "Exploratory Search: From Finding to Understanding." *Communications of the ACM*. 49 (4): 41–46.
- Mirizzi, R., A. Ragone, T. Di Noia, and E. Di Sciascio. 2010. "Semantic Wonder Cloud: Exploratory Search in DBpedia." *The 10th International Conference on Current Trends in Web Engineering (ICWE'10)*, 138–149. Berlin: Springer-Verlag.
- Neumann, G., and S. Schmeier. 2013. "MobEx – A System for Exploratory Search on the Mobile Web." In *Agents and Artificial Intelligence, 4th International Conference, ICAART 2012, CCIS 358*, edited by Joaquim Filipe and Ana Fred, 116–130. Berlin: Springer.
- Rivadeneira, A. W., D. M. Gruen, M. J. Muller, and D. R. Millen. 2007. "Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA.

- Schrammel, J., M. Leitner, and M. Tscheligi. 2009. "Semantically Structured Tag Clouds: An Empirical Evaluation of Clustered Presentation Approaches." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09), 2037–2040. New York, USA.
- Sinclair, J., and M. Cardew-Hall. 2008. "The Folksonomy Tag Cloud: When is It Useful?" *Journal of Information Science* 34 (1): 15–29.
- Veres, C. 2012. "MapXplore: Linked Data in the App Store. In Sequeda." In *COLD, Volume 905 of CEUR Workshop Proceedings*, edited by J. A. Harth and O. Hartig, CEURWS.org. http://ceur-ws.org/Vol-905/Veres_COLD2012.pdf.
- Wilson, M., J. Hurlock, and M. Wilson. 2012. "Keyword Clouds: Having Very Little Effect on Sensemaking in Web Search Engines." CHI '12 Extended Abstracts on Human Factors in Computing Systems, CHI EA '12, 2069–2074. Austin, Texas.
- Zhang, Q., W. Qu, and L. Wang. 2011. "How Font Size and Tag Location Influence Chinese Perception of Tag Cloud?" In *Engineering Psychology and Cognitive Ergonomics, HCII 2011, LNAI 6781*, edited by D. Harris, 273–282. Berlin: Springer-Verlag.