Talking to Linked Data: Comparing voice interfaces for general-purpose data

Master thesis of Information Sciences

Rianne Nieland
Vrije Universiteit Amsterdam
ra.nieland@student.vu.nl

ABSTRACT

People in developing countries cannot access information on the Web, because they have no Internet access and are often low literate. A solution could be to provide voice-based access to data on the Web by using the GSM network. Related work states that Linked Data could be a useful input source for such voice interfaces.

The aim of this paper is to find an efficient way to make general-purpose data, like Wikipedia information, available using voice interfaces for GSM. To achieve this, we developed two voice interfaces, one for Wikipedia and one for DBpedia, by doing requirements elicitation from literature and developing a voice user interface and conversion algorithms for Wikipedia and DBpedia concepts. With user tests the users evaluated the two voice interfaces, to be able to compare them.

Results show that there is no difference in speed, error rate and usability between the two voice interfaces, which means it does not matter if normal web data or Linked Data is used, both voice interfaces are functioning efficiently.

1. INTRODUCTION

The World Wide Web is a big information space (Jacobs & Walsh, 2004), which includes diverse information. Within this information, there is useful information for people in developing countries, which are poor or primitive countries. An example is Wikipedia. Wikipedia contains general-purpose information, like medical, geographical and governmental information and information about plants and trees (Foy, 2013).

People in developing countries cannot access this information on the Web, because they have no Internet access. The cause of this is that there is a lack of high-bandwidth Internet connections and reliable power supplies, also buying a Personal Computer is quite expensive (De Boer et al., 2013). Even if they could access the Web, these people are often low literate, which means they cannot read the information on the Web. For example in Mali, where the case study in De Boer et al. (2013) took place, only 2.7% of the population has Internet access (Miniwatts Marketing Group, 2012) and 31.1% is literate (UNESCO, 2010).

A solution to the literacy and Internet access problem is to provide voice-based access to the Internet by using the GSM network (De Boer et al., 2013). 2G mobile phones are digital mobile phones that use the GSM network (Fendelman, 2014). In Africa the primary mode of telecommunication is mobile telephony (UNCTAD, 2007).

This paper is about how information on the Web efficiently can be made available using voice interfaces for GSM. We developed two voice interfaces, one using Wikipedia and the other using DBpedia. In this paper we see Wikipedia and DBpedia as two different ways to make Wikipedia information available. DBpedia extracts structured information from Wikipedia and makes this available on the Web (Lehmann et al., 2013). We compared the two voice interfaces with a limited experiment to determine which voice interface performs better.

With voice interfaces for the Internet we like to provide farmers in developing countries access to general-purpose information on the Internet, like the information on Wikipedia and DBpedia. Related literature states that Linked Data, like DBpedia, is suitable for sharing information in developing countries, because it is a very lightweight way to share, re-use and integrate various datasets using Web standards such as URIs and RDF (De Boer et al., 2013). The Linked Data version of Wikipedia is DBpedia (Lehmann et al., 2013).

The goal of this research is to investigate whether it is possible to make Web data efficiently available on voice interfaces for GSM. It is not the goal to make such voice interfaces for developing countries. The situation in developing countries we used as motivation for doing our research.

2. RELATED WORK

In this section different systems will be explained. The systems explained in this section are whole systems which

are already implemented to solve a particular problem. Our system instead will only be used for research purposes, it will not be implemented somewhere. We used the principles of these different systems for developing our voice interfaces.

2.1 Wikipedia Zero

Wikipedia Zero provides free access to Wikipedia on mobile phones (Foy, 2013). It provides access to Wikipedia to people in sub-Saharan Africa, who do not have access to Internet, but do have a mobile phone. People first dial a number on their mobile phone. After dialing the number, they receive a text message inviting them to enter a search term. The user enters a search term, after which a menu appears where the user has to clarify which Wikipedia article he is interested in. When the user chooses one of these articles, a menu appears with the section headers of that article. The user then chooses which section he wants to receive. After that the user receives a text message with one paragraph of the article. If the user wants to continue reading he can reply to the text message.

Like Wikipedia Zero, our master project also provides access to Wikipedia on mobile phones without Internet access. The only difference is the way they provide the information to the user. Wikipedia Zero uses text messaging and our project uses voice interfaces.

2.2 RadioMarché

RadioMarché is a web-based market information system that stimulates agricultural trade in the Sahel region for local farmers of non-timber forest products, like honey and tamarind (De Boer et al., 2012). It distributes up-to-date market information via community radio in the area. The offerings are sent using text messages. Someone enters this information in a web form, after which it is accessible for employees of local community radio stations through the Web or through a voice interface that is accessible on first generation mobile phones. In De Boer et al. (2013) the locally produced market data links to external sources on the Web of Data. With this linked market information, it is possible to plot local product offerings on an interactive map or present external product information to the user.

This project made Linked Data available via voice interfaces on mobile phones without Internet. Our project also uses Linked Data, except not to link market information, but to make general-purpose datasets available to the user. Another difference is that RadioMarché uses local languages and is used in development countries and the voice interfaces of our project are in English and are only used for research purposes.

2.3 Voice assistants

Voice assistants are used on devices, like mobile phones and tablets, to enable the user to control their device by means of speech. An example is Siri, which is a voice assistant of Apple, which can be used on Apple products, like iPhone and iPad (Aron, 2011). It uses voice recognition and natural language processing, which enables Siri to understand natural language. Users can for example check the weather and make an appointment by asking aloud. Also it can make reservations on websites, because most web services now offer applications programming interfaces that let apps feed information to them. Siri is only for helping in particular domains.

A similarity between voice assistants and our project is that they both provide information on the Web to the user using speech. A difference is that our project does not make use of natural language input and is available to users with second generation mobile phones.

2.4 Screen readers

If a person is visually impaired, it is hard to travel on the Web, because Web pages are made for visual interaction (Goble, Harper, & Stevens, 2000). Therefore visually impaired people use screen readers to read the Web pages. However, screen readers cannot see the implicit structural and navigational knowledge encoded within the visual representation of a Web page (Yesilada, Harper, Goble, & Stevens, 2004). In Yesilada et al. (2004) the Web pages are annotated with a Travel ontology, which is aimed to encapsulate rich structural and navigational knowledge about these pages. The Travel ontology annotates concepts, like header, menu links, chunk, footer and copyright. By annotating these concepts the screen reader can provide an overview of the page; movement to the focus of the page; a structuring list of links, for example a list of all the links on the home page; and elimination of repetitions, like the copyright and footer, which are the same on every page.

A similarity between screen readers and our project is that they both provide information on the Web to users using audio. The way screen readers extract and provide information of the Web to the user is interesting for our project. A difference with our project is that it does not provide the information on mobile phones without Internet.

2.5 Spoken Wikipedia & Pediaphon

Spoken Wikipedia is an audio interface for and developed by Wikipedia (Wikipedia, 2013). Contributors can record audio files and upload them to the articles. This is a time-consuming process, and lacks some features of the text based Wikipedia articles (Bischoff, 2006). One feature is actuality, it is impossible to change an audio file directly after the text-based article is changed. Another feature is completeness, it takes quite some time to record all articles of Wikipedia. A third feature is objectivity, Wikipedia articles establish an objective view, an emotional interpretation of a contributor may break this objectivity. The last feature is audio quality.

Pediaphon instead is an audio interface for Wikipedia, which uses text to speech (TTS) audio generation. That means it generates audio of Wikipedia articles dynamically (Bischoff, 2006).

Spoken Wikipedia and Pediaphon are both audio interfaces for Wikipedia. In our project, Wikipedia is made available using TTS audio generation, like Pediaphon. A difference is that the interfaces of our project are accessible on mobile phones without Internet.

3. TECHNOLOGICAL BACKGROUND

3.1 Wikipedia & DBpedia

Wikipedia is one big central knowledge source, which thousands of contributors maintain (Mendes, Jakob, & Bizer, 2012). Wikipedia articles have a wide variety of subjects. The articles consist mostly of natural language text (Mendes et al., 2012). Except natural language text, they also contain different types of structured information, like the infobox, images and links to external Web pages (Mendes et al., 2012).

DBpedia extracts this structured information of Wikipedia as RDF triples (Mendes et al., 2012). For every Wikipedia article DBpedia creates a Uniform Resource Identifier (URI). The RDF triples are properties of the corresponding URI. The DBpedia ontology is an ontology made, to avoid a large amount of different kinds of properties on DBpedia (Mendes et al., 2012). This ontology organizes the knowledge on Wikipedia in 320 classes, which form a hierarchy, and 1650 different properties, which can describe the classes (Mendes et al., 2012). The data on DBpedia is interlinked with other data sources, therefore it is called Linked Data.

In short, Wikipedia mostly contains natural language text and DBpedia contains structured information, which are classes described by properties. For each input source we developed a voice interface.

3.2 Voice interfaces

The standard language used to develop voice interfaces is VoiceXML (VXML), which is an open standard markup language developed by W3C (Lucas, 2000). Voice interfaces can have speech input or Dual-Tone Multi-Frequency (DTMF) input, which is dialed input(Patel et al., 2009). The output of a voice interface can be pre-recorded or can be automatically generated, which is called text to speech (TTS) (Mecanovic & Shi, 2005). The navigational structure of a voice interface is represented by a call flow diagram (Dahl & Froumentin, 2012). It represents how the user navigates through the application from beginning to end.

4. PROBLEM STATEMENT & APPROACH

The problem is that people in developing countries are low literate and do not have Internet access. But they do have a mobile phone. Wikipedia Zero uses text messages to provide access to Wikipedia articles. This is not helpful for low literate people. Therefore, in this research we developed voice interfaces to make Wikipedia articles available on mobile phones. The information provided by the voice interfaces is Wikipedia information, because 70% of the mobile accessed websites provide general-purpose data, like the information on Wikipedia (Schmiedl, Seidl, & Temper, 2009). The research question is: How can information from Wikipedia efficiently be made available using voice interfaces for GSM?

The following subquestion will help to answer the research question:

1. What are the requirements of a good voice interface for Wikipedia and DBpedia concepts?
2. What are good methods for converting Wikipedia and DBpedia concepts to voice interfaces?
3. How do users perform on the Wikipedia and DBpedia voice interface in terms of speed, error rate and usability?

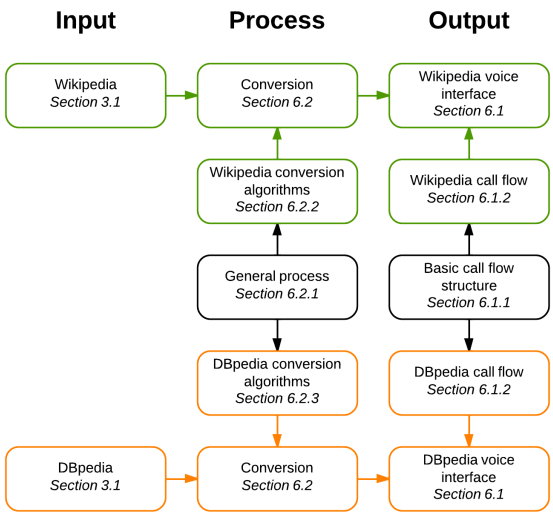
To answer the research questions we did a literature study, to find out more about voice interfaces and the differences between Wikipedia and DBpedia. To answer subquestion 1, we did requirements elicitation from literature. To answer subquestion 2, we developed a general process and conversion algorithms for Wikipedia and DBpedia concepts. The algorithms are the foundation for the two voice interface prototypes, one for Wikipedia and one for DBpedia. These prototypes are in English and contain only a small part of Wikipedia and DBpedia, to narrow down the scope of the research. They are not developed to be used in developing countries. To answer subquestion 3, users evaluated these prototypes. The results of these evaluations are used to study possible differences between the two in terms of speed, error rate and usability.

See figure 1 for an overview of how the different parts fit together. This diagram contains the section numbering to be able to see where in the overall picture a certain section is located. The diagram is an overview of the overall process of the Wikipedia and DBpedia voice interface made in this project. The overall process consists of input, process and output.

5. REQUIREMENTS ELICITATION

This section describes the requirements of good voice interfaces. Section 5.1 describes the input requirements, section 5.2 describes the process requirements and section 5.3 describes the output requirements.

Figure 1: Overall process of conversion with corresponding section numbers for each step



5.1 Input requirements

5.1.1 *DTMF input*

A requirement for low literate users living in developing countries is to use DTMF, which is dialed input, as an input source for voice interfaces. The reason for using DTMF, is that low literate users prefer DTMF input instead of speech input (Sherwani et al., 2009), DTMF also outperforms speech input in terms of task completion rate and learnability, and users reported significantly less difficulty providing input using DTMF than using speech (Patel et al., 2009). Also not all languages are available for natural language systems (Sharma Grover, Stewart, & Lubensky, 2009). For example, under-resourced languages in developing countries, where there is limited linguistic knowledge and resources, technical expertise, computing infrastructure and funding to build natural language systems, are not available (Sharma Grover et al., 2009).

The voice interfaces of this paper also use DTMF as input source, because of the better performance than speech input, the higher task completion rate and because it is easier to learn and is less difficult to use. Also for future work, when using these voice interfaces for low literate users in developing countries, it is the preferred input source.

5.1.2 *Local phone line*

An important requirement for voice interfaces in developing countries is that the voice interface should be accessible on a local phone line (Bon et al., 2013). A local phone line is an important requirement, because this saves the user quite some costs (Bon et al., 2013). Local phone lines are cheaper than other phone lines.

The use of a local phone line is not a requirement for the voice interfaces developed in this research paper, because these voice interfaces are not for developing countries. In this paper we only want to see if it is possible to make Wikipedia and DBpedia accessible via voice interfaces on a mobile phone. We will use a Skype number instead of a local phone number. In future work, a local phone number can be used, when making these voice interfaces for developing countries.

5.2 Process requirements

5.2.1 *Providing an overview of the page*

When a user is visually impaired, he uses a screen reader to read web pages. The way screen readers read web pages to a user is also a good method for making voice interfaces for web pages. When accessing a web page with a voice interface or a screen reader it is important that the user gets an idea of what is on the page, a kind of table of contents. In Yesilada et al. (2004) they made a table of contents for screen readers using the headings within the page. They add links from the table of contents to the headings and also back. Adding a table of contents improves the intra (within the page) mobility support, which then again also improves the inter (between the pages) and collection wide (within the site) mobility support (Yesilada et al., 2004).

An overview of the pages is also implemented in the voice interfaces of this paper, because an overview gives the user an idea of what kind of information he can expect to find in the voice interface.

5.2.2 *Eliminating repetitions*

Some parts of web pages are the same on all pages of a website, like the header and the footer. If users access web pages with screen readers or voice interfaces, they will read the whole web pages to the user. A user accessing the web pages on a computer often skips these repetitive parts. Therefore Yesilada et al. (2004) removed these repetitive parts on web pages, to be able to provide the user a shorter and concise web page. This is particularly useful if a user accesses the web page more than once (Yesilada et al., 2004).

We implemented this requirement in the voice interfaces of this paper, because these repetitive parts do often not contain any interesting information about the content of the page.

5.2.3 *Feedback*

When designing voice interfaces, it is important that users always get immediate and informative feedback (Yankelovich, Levow, & Marx, 1995). It is important that the user knows if the system has understood the input the user gave.

We added this informative and immediate feedback to the voice interfaces, because it helps the user to know

what is happening. After entering a number, the system tells the user the chosen option, for example 'You have chosen Maize'.

5.2.4 Error recovery

If a user makes an error, the system should guide the user to recover from this error (Sharma Grover et al., 2009). For low literate users, it is very important to provide robust error management (Sharma Grover et al., 2009). To help the user recover from errors, the system should give clear error recovery feedback. Without clear error recovery feedback, the learning curve for low literate users of the voice interface will become much harder and the use of the interface will become a frustrating experience (Sharma Grover et al., 2009). To help users recover from errors the system should use explicit and implicit confirmation prompts (Sharma Grover et al., 2009). The system should use implicit confirmations for actions that can easily be undone and are noncritical, and explicit confirmations for critical actions (Cohen, 2004). Voice interfaces should maintain a careful balance between these implicit and explicit confirmations (Sharma Grover et al., 2009).

For the voice interfaces of this paper we added clear error recovery feedback for possible errors, which could occur. For example when pressing an impossible number for a DTMF menu, the user will be notified of this and asked to press another number, after which the choices are repeated.

5.3 Output requirements

5.3.1 System's voice

A requirement for the output is that it should be a female pre-recorded human voice. In the experiment of Mecanovic & Shi (2005) 75% of the users preferred a female voice over a male voice and 67% of the users preferred a pre-recorded human voice over a TTS computer voice. In the case of dynamic contents a pre-recorded voice is no option, in that case a TTS voice is preferred (Mecanovic & Shi, 2005).

A pre-recorded human voice is no option for the voice interfaces developed for this research, because the content of the voice interfaces is dynamic. Therefore we use a female TTS computer voice.

5.3.2 Nonverbal sounds

Users form a mental image of a personality or character of the voice interface based on the application's voice and language. To keep up the mental image, the voice interface should sound more human. Therefore another output requirement is that there should be some consistent suitable nonverbal sounds and background music throughout the whole dialog, to increase the clarity of prompts and messages (Schnelle & Lyardet, 2006).

We did not implement this requirement, because it is not possible to add background sounds to a voice interface using the voice application language VXML.

6. DEVELOP VOICE INTERFACES

6.1 Output: Voice user interface

After analyzing the requirements for voice interfaces, the user interface was developed, which is the output part of figure 1. To develop the user interface we made a basic call flow structure. This basic call flow structure we used to make the Wikipedia and DBpedia call flow, which is shown in figure 2. Both call flows will therefore have the same underlying structure, to be able to compare them. The only difference between the two call flows will be the data and data format used. Section 6.1.1 describes the basic call flow structure and section 6.1.2 describes the Wikipedia and DBpedia call flows.

6.1.1 Basic call flow structure

The basic call flow structure contains four parts:

1. The system plays the welcome message and page menu. The user chooses a Wikipedia or DBpedia page, for example the page about maize.
2. The system plays the section menu of the chosen page. The user chooses a section, for example the abstract or history section.
3. The system plays the subsection menu of the chosen section. The user chooses a subsection.
4. The system reads the chosen (sub)section to the user. At the end the system redirects the user to the main menu, which is the menu in step 1.

A design choice was to only use a small amount of pages the user can choose from in the voice interface. This is different compared to a user who searches on Google for one of 4.5 million English Wikipedia pages (Wikipedia, 2014). We chose to only use a small amount of pages to scale down the experiment. After opening a page, a user has to choose the section he wants to read. When looking at a webpage in a browser, the user can scroll down to the section he wants to read. In the voice interface the user gets a menu with all sections on the page, of which he can choose the one he wants to read and if applicable also the subsection. This is part two and three in the call flow structure. At the end the user with a browser starts reading the chosen section. In our call flow, the voice interface will read the section to the user using TTS, which is part four of the call flow structure.

To provide an overview of the page, which is one of the requirements, see section 5.2.1, the call flow gets the sections and subsections of the page and displays them as menu options to the user. This overview is part two and three of the call flow structure. In these

parts the call flow sums up the different (sub)sections and displays these as DTMF menu options.

The call flow also eliminates repetitions, which is a requirement in section 5.2.2, by not displaying parts, like the header and copyright section. It only displays the real content of the page, which is the actual content of the page, excluding menus, header sections, footer sections, copyright sections and sidebar content. The actual content is different for every page.

After the user chose a page or a section, the user gets immediate feedback about what he has chosen. Providing feedback to the user was also a requirement as stated in section 5.2.3.

The call flow also provides error feedback, which is a requirement as stated in section 5.2.4. The call flow structure checks if the user has given input. If not it asks the user to enter a number and repeats the menu options. Also the system checks if the user has entered valid input. If not, the system tells the user that the input given is no option and asks him to try again. Thereafter the system repeats the menu options.

6.1.2 Wikipedia and DBpedia call flow

Based on the call flow structure we made the Wikipedia and DBpedia call flow. Figure 2 is the call flow diagram of both call flows, because the structure of both call flows is exactly the same. The only difference is that they both use different input sources. The Wikipedia call flow uses Wikipedia data and the DBpedia call flow uses DBpedia data. Section 3.1 describes the difference between these two input sources. In more detail, the parts of the call flows that are different, are the menu of the different sections and subsections and the content of the different (sub)sections. The rest is the same. In figure 2, the differences are indicated with sliced blocks in which the green text is for the Wikipedia call flow and the orange text is for the DBpedia call flow.

6.2 Process: Conversion algorithms

The previous section explained the call flow of the voice interfaces, which is the output of the voice interfaces. Wikipedia or DBpedia pages are the input. The remaining question is how to get from the input to the output, which is the process part shown in figure 1. This section explains this remaining step, which also answers subquestion 2. Section 6.2.1 explains the general process and sections 6.2.2 and 6.2.3 describe the conversion algorithms.

The conversion algorithms convert Wikipedia and DBpedia concepts in such a way they can be used as input for the voice interfaces. The general process uses the converted input to generate the call flow. Both voice interfaces have the same general process. The steps 'Play section options', 'Play subsection options' and 'Read (sub)section' are conversion steps. These three steps differ for the Wikipedia and DBpedia voice interfaces, because of the different data input sources.

6.2.1 General process

The activity diagram in figure 3 represents the general process. The whole process will be the same for both voice interfaces except the parts indicated in green and orange. The green parts are only present in the Wikipedia voice interface and the orange parts in the DBpedia voice interface. As can be seen in figure 3, these are only the different input sources for the different steps and one check to skip the subsection menu. The rest of the general process is the same for both voice interfaces so we are able to compare the two voice interfaces in the experiment.

The process of the voice interfaces is live, which means when a user calls the voice interface, the voice interface will retrieve the Wikipedia and DBpedia data during the process. The live retrieval of data is to ensure that the data is up-to-date. A disadvantage of this is that if the input source is offline, the voice interfaces cannot retrieve data.

The process involves three parties, the user, system and Wikipedia/DBpedia. The user provides input to the system, like starting the voice interface and choosing a (sub)section. Wikipedia and DBpedia provide the system with data input. Without this data the voice interface would be a voice interface with empty menus.

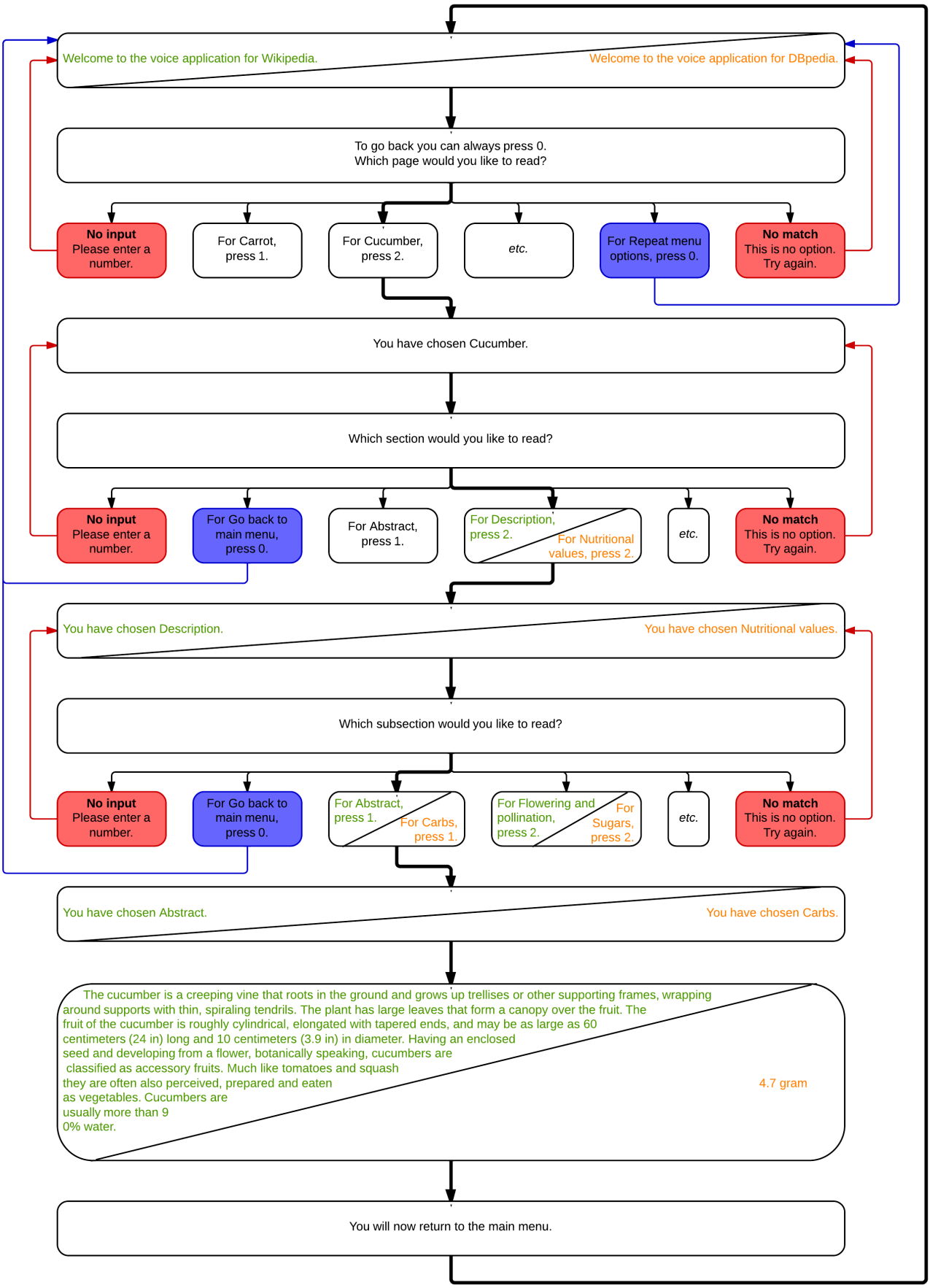
The process generates the call flow, described in the previous section. The call flow has four parts, therefore also the process can be divided into four parts, see figure 3. These parts correspond to each other. Every part in the process generates the same part in the call flow.

After the user chose a page, section or subsection, the system always checks if the user has given any input, if the input is valid and if the user has pressed the back button, before going on with the process.

In the second part of the process the system confirms the chosen page. After that the system plays the sections menu. The operation of the step 'Play section options' differs for the DBpedia and Wikipedia voice interfaces. This is the first conversion step, described in section 6.2.2.1 and section 6.2.3.1.

The third part is optional in the Wikipedia and DBpedia voice interface. The Wikipedia voice interface will only have a subsection menu if the chosen section has subsections. Otherwise the user will skip this part and go to part four in which the system reads the section to the user. The DBpedia voice interface will only have a subsection menu if the chosen section is not the abstract. If the chosen section is the abstract, the user

Figure 2: Wikipedia and DBpedia call flow diagram



will skip this part and continue with part four. In the third part the step 'Play subsection options' contains a conversion step for both the DBpedia and Wikipedia voice interface, which is described in section 6.2.2.2 and 6.2.3.2.

In the last part the system reads the content of the chosen section or subsection to the user. The step 'Read (sub)section' differs for the DBpedia and Wikipedia voice interface, because also for this step the Wikipedia and DBpedia concepts have to be converted to data input the voice interface can use. Section 6.2.2.3 and 6.2.3.3 explain this step for each voice interface. After the system has read the section to the user, the system returns to the main menu.

For coding the whole process VXML and PHP are used. The different parts of the process are separate code files, which are on https://github.com/ennair/TalkingLinkedData. VXML is used to generate the call flow and PHP is used for the conversion of the data. Apart from VXML and PHP, also query languages are used. For each voice interface the used query language is different. In the DBpedia voice interface the query language SPARQL is used. To query over the DBpedia dataset we made use of the public SPARQL endpoint at http://dbpedia.org/sparql. For the Wikipedia voice interface DOMXPath queries are used.

6.2.2 Wikipedia conversion

The Wikipedia voice interface contains three steps in which Wikipedia data is converted to data the voice interface can use. The three conversion steps are 'Play section options', 'Play subsection options' and 'Read (sub)section' of the general process in figure 3. Appendix A displays the diagrams of these conversion steps.

6.2.2.1 Wikipedia - Play section options

The step 'Play section options' extracts the section headers from the Wikipedia page and converts them in such a way that the information can be played as section menu.

All headers in the Wikipedia page have the class name mw-headline. This is the case for all pages on Wikipedia. The algorithm removes the heading 3 (`h3`) and heading 4 (`h4`) elements, to be able to extract only the section headers and not the subsection headers. After that the system plays the back and abstract options. The abstract of Wikipedia has no header. Therefore the abstract section option is a static menu option. After that the algorithm loops over the text to search for elements with the class mw-headline. If it finds such an element, it reads the header and checks if it is 'See also'. If not, it will generate a VXML option for this header.

The loop stops after finding the 'See also' header to exclude the sections 'See also', 'Notes', 'References', 'Further reading' and 'External links'. This is a design decision. These sections contain long lists with references and do not include any sentences or context. When a voice interface would read this to a user, it would be of no use to the user, because it does not contain any relevant contextual data about the subject. We also excluded the information gathered from these sections by DBpedia to keep the voice interfaces as similar as possible.

6.2.2.2 Wikipedia - Play subsections options

The 'Play subsection options' step converts the Wikipedia page to be able to generate the subsection menu. This is an optional step in the voice interface. If the chosen section has subsections, the system will execute this step. To do this, the system needs the chosen Wikipedia page. After that it executes a query to get all headings 3, which are the subsection headings between the chosen section header and the next section header. Thereafter it prints the static back option and abstract option and checks for subsections. If it finds a subsection it prints it as a VXML menu option, otherwise it stops.

6.2.2.3 Wikipedia - Read (sub)section

The step 'Read (sub)section' first retrieves the chosen Wikipedia page to be able to read the chosen section to the user. Next it removes the `sup` elements, which are the reference numbers in the text. The system deletes the reference numbers, because it was hard to distinguish the reference numbers from the numbers in the text, when listening to the voice interface. Also the sentences were constantly interrupted when the voice interface read the reference numbers, it was therefore hard to understand what a sentence was about.

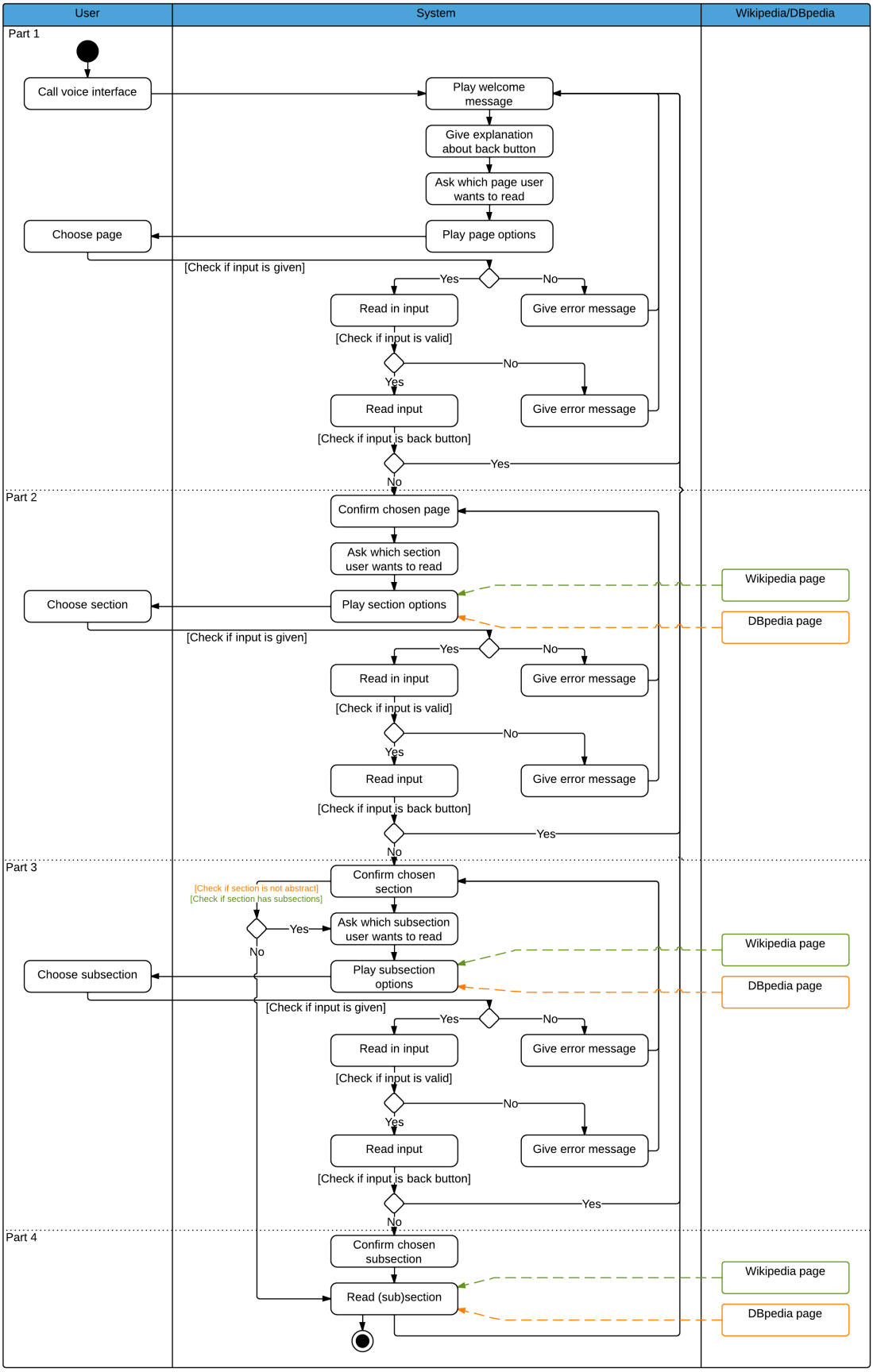
After removing the reference numbers, the system checks if the chosen section header is the abstract. If the chosen section header is the abstract, it queries all paragraphs (`p`) and list items (`li`) before the first section. If the chosen subsection is abstract, it queries all paragraphs and list items between the chosen section header and the first subsection. The system reads these items to the user. If the chosen (sub)section header is not the abstract, it queries all paragraphs and list items between the chosen (sub)section header and the next (sub)section header and reads this to the user.

6.2.3 DBpedia conversion

This section explains the conversion steps 'Play section options', 'Play subsection options' and 'Read (sub)section' for the DBpedia voice interface. Appendix B displays the diagrams of these steps.

The DBpedia voice interface is domain specific. If this voice interface will be used for another domain, the properties of that domain should be studied first. After that the section menu, which will be explained in section 6.2.3.1, and all SPARQL queries in the sections 6.2.3.2

Figure 3: General process diagram



and 6.2.3.3, except the one for the abstract, should be changed. The rest of the code remains the same. For the comparison of the two voice interfaces in section 7, this does not matter. However, when one would like to expand the voice interfaces to cover whole Wikipedia and DBpedia, a domain independent voice interface is easier to expand.

6.2.3.1 DBpedia - Play section options

The section menu of the DBpedia voice interface is a static menu, which means it is the same for every page. After we studied the DBpedia properties of crops, we added the sections 'Abstract', 'Nutritional values', 'Biological classification' and 'Associated food, persons and organizations' to the voice interface. This is a domain specific menu. The only domain independent section is 'Abstract', because every resource has an abstract. The domain of this research is scaled down to crops, which will be explained in section 7.2.

In the first DBpedia voice interface version, the menu was one big list of properties without subsection menu. This was a domain independent menu, but it was not user-friendly, because the user had to listen to one long list of options. After a pilot experiment, the user was not satisfied with this voice interface. Therefore we divided the properties into sections. A disadvantage of DBpedia is that the different properties are hard to distinguish. The properties hardly have any properties themselves to distinguish them from each other. Therefore at the end we chose for a domain specific menu.

6.2.3.2 DBpedia - Play subsection options

The 'Play subsection options' step extracts DBpedia property labels to be able to generate a subsection menu. This is only for the sections 'Nutritional values', 'Biological classification' and 'Associated food, persons and organizations'. The 'Abstract' section is only one property, therefore this section has no subsections. To make the subsection menu, the system first retrieves the chosen DBpedia page. Next it checks the chosen section, because every section has another SPARQL query.

The query of the 'Nutritional values' section retrieves all property labels of the chosen DBpedia page of which the value has data type gram or of which the value is a double and the label contains 'mg'. The query orders these properties descending by the value size. First it orders all values with data type gram descending and after that the properties that are in milligram.

To get all properties belonging to the section 'Biological classification', the query retrieves all object properties with an English label of the chosen DBpedia page. It orders these properties alphabetically.

The query for the 'Associated food, persons and organizations' section retrieves all English property labels with domain `Food`, `Organisation` or `Person`. It orders these properties also alphabetically. Most properties do not have a domain in DBpedia, therefore this section only has a maximum of three properties.

Thereafter the system executes the SPARQL query to get the query results. It plays a back option, after which it checks for `binding` elements. Every property returned by the query is a `binding` element. When it finds a `binding` element, it prints it as a VXML section option.

6.2.3.3 DBpedia - Read section

After the user has chosen the 'Abstract' section or one of the subsections, the system reads it to the user. The same SPARQL queries as in section 6.2.3.2 are used. The only difference is that the queries also retrieve the value for each label.

It is possible that a label occurs multiple times with different values. To tackle this, the system sets an integer `i` to one and a boolean `print` to `false`. Next it checks for `binding` elements. A `binding` element can be a label or a value. Therefore after finding a `binding` element it checks if `i` is odd. If yes, this means that it is a label, because first the query prints the `binding` element of the label and then the `binding` element of the value of that label. The system checks if the label is the same as the chosen subsection header. If it is not the same, this label is not the one searched for. If it is the same, this label is the one searched for. The system then changes the `print` value to `true`, to indicate that it should print the next `binding` element, which is a value element. After the `print` boolean is set to `true`, it adds one to `i` and searches for the next `binding` element. This `binding` element will be even, because the previous one was odd. Next it reads the `print` boolean, if it is `true` it reads the content of the `binding` element to the user. If it is `false`, it skips this `binding` element, adds one to `i` and searches for the next `binding` element until there are no left.

If the chosen section is the 'Abstract', the query retrieves the `abstract` value of the chosen page. After receiving this query result, the system reads the `binding` element on index zero to the user.

6.3 DBpedia backup

While building the voice interfaces and during the pilot experiments, DBpedia was often down. Therefore we could often not use the DBpedia voice interface. To solve this problem, we made a backup DBpedia voice interface, which uses pre-saved query results. The backup voice interface does not execute queries anymore, but excesses the static query results. The call flow remains the same. In case DBpedia is down when doing the experiment, we will use this backup version. The code can be found on the following website: https://github.com/

ennair/TalkingLinkedData.

6.4 Implementation

To be able to call the voice interfaces we used Voxeo Evolution (Corporation, 2014). Voxeo Evolution hosts the applications and adds phone numbers to them. The Wikipedia voice interface has the following Skype number: 00990009369991486423. The Skype number of DBpedia is: 00990009369991488325. Next to the Skype numbers, they also have a local American phone number. As the experiment will be held in the Netherlands, the American numbers are expensive to call, therefore for the experiment the Skype numbers are used.

7. EXPERIMENT

This experiment compares the Wikipedia and DBpedia voice interface in terms of speed, error rate and usability. The goal is to answer research subquestion 3. Users therefore tested the two voice interfaces and evaluated the usability of them. The speed of use and the error rate are measured by the experimenter.

7.1 Participants

For the user test, 16 students tested the Wikipedia and DBpedia voice interface.

We calculated the minimum sample size per group for the two-tailed student t-test by first calculating the anticipated effect size with Cohen's d (Soper, 2006). To calculate Cohen's d, we assumed that if the participants need three minutes for the Wikipedia version then they would need one minute for the DBpedia version. Also the standard deviation within the groups we assumed to be 0.5. The result is an effect size (Cohen's d) of four. The desired statistical power level is 0.95. The minimum sample size per group, which we calculated with Cohen's d and the desired power level, is three. Because there were four groups, as section 7.4 will explain, we needed at least 12 participants.

7.2 Domain

We scaled down the domain for this experiment to crops, to limit the scope of the experiment. Crops have sufficient information on Wikipedia and DBpedia. To test the voice interfaces the participants received a question set consisting of three questions, which required the participant to find certain information about a certain crop. The two voice interfaces provided the information to answer all questions. There were two question sets with different questions so participants would not get the same questions for the two voice interfaces they had to test. The participant tested the Wikipedia and DBpedia version. When doing the same questions for both versions, this could have influenced the results, because they would already know the answer and could already have an idea where to look for it.

The first question set consists of the following questions: 'How many carbohydrates (carbs) does a potato contain?', 'Is the carrot a flowering plant?' and 'Of what is maize an ingredient? Name 2'. The second question set consists of the following questions: 'Does a pumpkin contain zinc?', 'Is Gnocchi made from potatoes?' and 'What are the three main varieties of cucumbers?'.

7.3 Material

The experiment took place at the Vrije Universiteit Amsterdam, to ensure the Wi-Fi connection, needed for Skype, is the same. Between different Wi-Fi networks the speed and robustness could be different. Also every participant used the same smartphone to call the voice interfaces, to ensure the speed and user interface, like keyboard, are the same. Therefore participants could not do the experiment simultaneously. The smartphone had the Skype app installed, to be able to call the voice interfaces with Skype. All participants used the same Skype account to call the voice interfaces. We created a new Skype account for this experiment, to ensure users did not get distracted by their friends on their personal Skype account. With a Skype account especially for this experiment, it was also easier to keep track of the times the user needed to find an answer with a voice interface. Next to the smartphone we also needed a tablet on which the user could fill in the surveys and answers to the questions. These we all combined into one form using Google Forms (Google, 2011).

7.4 Procedure

We assigned participants randomly into one of four groups. The groups differed in the order they tested the different versions and in the question set they got for each version, to prevent that the order of the versions influenced the results. Two of these groups first tested the Wikipedia version and the other two groups first tested the DBpedia version. Each of these groups had another version of the form to fill in, see https://github.com /ennair/TalkingLinkedData. The four groups are:

- First Wikipedia version with question set 1 and then DBpedia version with question set 2 (W1D2);
- First Wikipedia version with question set 2 and then DBpedia version with question set 1 (W2D1);
- First DBpedia version with question set 1 and then Wikipedia version with question set 2 (D1W2);
- First DBpedia version with question set 2 and then Wikipedia version with question set 1 (D2W1).

Before the participants started, they received a verbal explanation about what they will be doing. This explanation explained that they were going to test two voice interfaces. For each voice interface they had to answer three questions and afterwards fill in a questionnaire

about what they thought of the voice interface. After the explanation the user started the experiment.

During the experiment the user first got a basic questionnaire in which the user had to fill in some basic information, like gender, age, purposes of mobile phone usage and usage of voice interfaces. After that the user had to test one of the two voice interfaces and had to answer the questions of one of the two question sets. For every question the user had to call the voice interface again. The time it took to find the answers to the question was measured. After finding the answers to these questions using the voice interface, the user had to fill in a questionnaire.

For the questionnaire we used the standardized Post-Study System Usability Questionnaire, which is one of IBM's computer usability satisfaction questionnaires (Lewis, 1995). This scale has a reliability of 0.97. A standardized scale is used, to be sure the right elements are measured. The scale measures the overall satisfaction, system usefulness, information quality and interface quality.

After the participant had answered the questions and filled in the questionnaire, he started testing the next voice interface. For this voice interface the user also had to answer three questions and fill in the same questionnaire, only this time for the second voice interface instead of the first voice interface tested.

8. RESULTS

After the participants conducted the experiment, we analyzed the results. The answers we retrieved from the questionnaires and the measured times, can be found on https://github.com/ennair/TalkingLinkedData. Section 8.1 explains the data cleaning process, in section 8.2 we analyzed the time it took participants to answer the questions, in section 8.3 we analyzed the number of questions answered correct and in section 8.4 we analyzed the results of the usability questionnaire.

To analyze the results we used the independent sample t-test, because the Wikipedia and DBpedia sample are independent of each other and the data comes from a normal distribution. They are independent, because of the different question sets used for both Wikipedia and DBpedia, and the changing order of the tested voice interfaces.

8.1 Data cleaning

Before analyzing the data we cleaned the data, which left us with 10 time results and 16 survey results for the Wikipedia voice interface, and 10 time results and 15 survey results for the DBpedia voice interface.

Five Wikipedia and four DBpedia time results were deleted, because participants did not hang up after finding an answer to a question, which resulted in one time for all three questions of a question set. We did not delete the survey results of these participants, since they were still able to evaluate the voice interfaces. We deleted another two time results of DBpedia, because one participant used the backup DBpedia voice interface, which is faster than the real DBpedia voice interface, and the other used the real DBpedia voice interface, but from his comments it was clear that DBpedia was offline while testing. The DBpedia survey results of this last participant were also deleted. For another participant, we removed the time results of the Wikipedia voice interface, because the time of the first question was almost four times as high as that of other participants. We assumed something went wrong.

8.2 Speed

The average time it took the participants to answer a question with the Wikipedia voice interface is 2:53 minutes and with the DBpedia voice interface is 2:22 minutes. The independent sample t-test shows this difference is not significant (0.097). We can conclude that both voice interfaces are equally fast. Also for individual questions no significant difference was found between the two voice interfaces. From the comments in the questionnaires participants saw both voice interfaces as time-consuming.

We calculated the significant difference between the first and last question of each voice interface with the independent sample t-test to check if there is a learning curve. There was no significant difference found, which means there is no learning curve for both voice interfaces.

8.3 Finding correct answers

In general the independent sample t-test shows no significant difference (0.106) between the two voice interfaces for the questions answered correct and incorrect. When checking the answers to the questions, the users had to answer with each voice interface, spelling errors and errors like only naming one item when two items are asked are ignored. Figure 4 shows how many questions the participants answered correct and incorrect for each voice interface. For both voice interfaces the same amount of participants could not find the answer. All questions could be answered with both voice interfaces.

Also there is no significant difference found for the correct and incorrect scores for each question of the question sets between the two voice interfaces, except for question 2 of question set 1 (0.006). Question 2 of question set 1 is: Is the carrot a flowering plant? For the Wikipedia voice interface all eight participants gave the correct answer. For the DBpedia voice interface only one participant gave the correct answer and two participants gave an incorrect answer, the other participants did not find the answer. The answer to this question in

Figure 4: Error rates Wikipedia and DBpedia

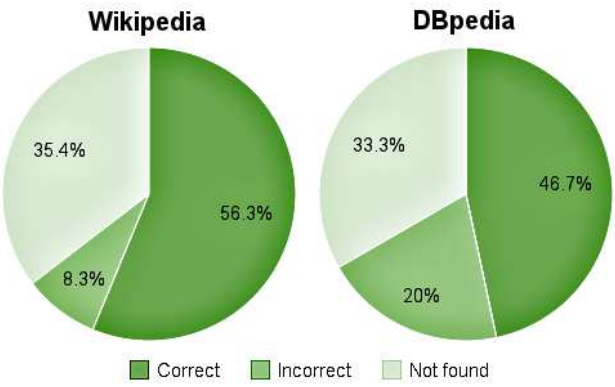


Table 1: Scores of questionnaire for Wikipedia (W) and DBpedia (D) voice interface

		OVER-ALL	SYS-USE	INFO-QUAL	INTER-QUAL
Question set 1	W	3.44	3.39	3.51	3.46
	D	3.38	3.09	3.72	3.38
Question set 2	W	4.09	4.20	4.08	3.63
	D	3.96	4.25	3.61	3.86
Tested first	W	3.32	3.39	3.40	3.00
	D	4.20	4.25	4.11	4.13
Tested second	W	4.21	4.20	4.20	4.08
	D	3.05	2.98	3.13	3.07
Total	W	3.76	3.80	3.80	3.54
	D	3.65	3.63	3.67	3.60

the DBpedia voice interface is located under section 'Biological classification' in subsection 'division'. This is probably not the most logical place to search for the answer to this question. Therefore DBpedia scored worst on this question. The name 'division' is given to the property by DBpedia. This is not something we could have changed.

8.4 Usability

To measure usability, we first measured if there is a significant difference between the Wikipedia and DBpedia voice interface results of the questionnaire. The independent sample t-test showed there is no question where there is a significant difference between the results for the Wikipedia and DBpedia voice interface.

For the IBM post-study system usability questionnaire used in the experiment, we can calculate four scores: overall satisfaction (OVERALL), system usefulness (SYSUSE), information quality (INFOQUAL) and interface quality (INTERQUAL) (Lewis, 1995). To calculate these scores we need to average the results of particular questions. The overall satisfaction score is the average of all questions. The system usefulness score is the average of the questions 1 through 8. The information quality score is the average of questions 9 through 15. The interface quality score is the average of the questions 16 through 18. If a participant has answered a question with not applicable, the remaining questions should be averaged. Low scores are better than high scores. Table 1 contains the calculated scores.

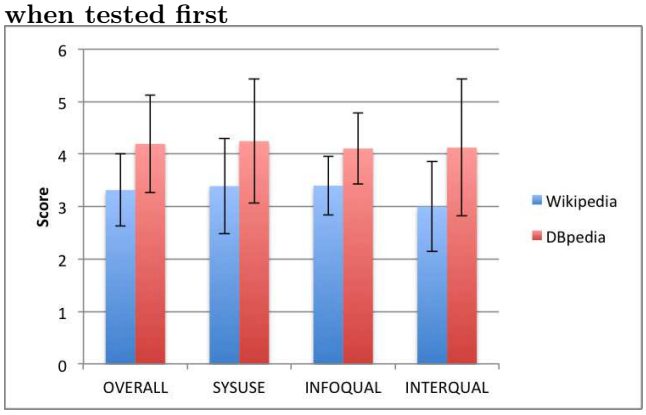
There is no significant difference between Wikipedia and DBpedia voice interfaces for any one of the four scores. However when we compared these scores between Wikipedia and DBpedia based on if they were tested first or last, there is a significant difference for the overall score (0.049) and the information quality score (0.039) between Wikipedia when tested first and DBpedia when tested first. For the other scores and the scores when both are tested second, there is no significant difference.

An interesting fact is, that between all the different scores there is a significant difference between the group where Wikipedia is tested first and DBpedia second and the group where DBpedia is tested first and Wikipedia second. However there is no significant difference between Wikipedia and DBpedia within these groups. The same participants tested both voice interfaces within one group. A difference between participants of the two groups or a certain influence of the first voice interface tested, can cause the significant difference between the Wikipedia tested first group and the DBpedia tested first group. With a certain influence of the first voice interface tested, we mean that the impression of the first voice interface can influence how the user scores the second voice interface. The second voice interface the participants almost rated the same as the first voice interface in that group.

Figure 5 is a plot of the scores of Wikipedia and DBpedia when tested first with their standard deviation. The overall score differs with almost one whole point, in which Wikipedia is better, and the information quality score differs with a bit more than 0.5 point, in which Wikipedia is also better.

In addition to the numbers of the questionnaire, the participants also added some comments. A lot of participants thought that the information was read too fast and was sometimes not easy to understand, but in general it was clear. Some participants found it hard to find the answers to the questions, because they did not know in which sections they had to search. Also sometimes there were too many menu options. For the Wikipedia voice interface some participants thought the content of the sections was too long. In the DBpedia version a participant said that the subsections were not clear, because they are different for each product.

Figure 5: Scores of Wikipedia and DBpedia when tested first



8.5 Summary of results

In short, both voice interfaces are equally fast also for the individual questions and both voice interfaces do not have a learning curve.

The Wikipedia voice interface scores significantly better for question 2 of question set 1 than the DBpedia voice interface, because in the DBpedia voice interface it is not clear where to search for the answer to the question. However in general there is no significant difference between the error rates of the two voice interfaces.

In general there is no significant difference between the Wikipedia and DBpedia voice interface in overall usability, system usability, information quality and interface quality. However the Wikipedia voice interface when tested first has a significantly higher overall usability and information quality score compared to DBpedia voice interface when tested first. Also there is a significant difference for all four scores between the group where Wikipedia is tested first and DBpedia second and the group where DBpedia is tested first and Wikipedia second. These differences can be caused by a difference between participant groups or the impression of the first voice interface influences how the user looks at the second voice interface.

9. DISCUSSION

During the experiment there were some problems, for example with the Internet connection. Sometimes the Internet connection did not work or during the experiment stopped working. This caused some problems, like not being able to precede the questionnaire or to call the voice interface.

Another problem, which occurred, was that DBpedia was offline a number of times. We noticed this already when building the voice interfaces, therefore we made a backup DBpedia voice interface by saving the answers to the DBpedia queries. This is possible, because we only used a small part of DBpedia. When making a voice interface for whole DBpedia, this will not be that easy, because it takes a lot of storage space and time to make such a backup. The backup DBpedia voice interface was faster than the real DBpedia voice interface, because it did not run queries. Therefore the times of this voice interface we cannot compare with the real voice interface.

After the experiment a few participants stated that they thought the menus were too long. They forgot the first options when they reached the end. Also in the current voice interfaces, the participants were not able to repeat parts of text and to go back one step, instead of going back to the main menu. Some participants also indicated they would have preferred to use their voice to navigate.

Some participants stated that DBpedia was hard to understand, probably because DBpedia only contains terms without any context. Wikipedia instead contains whole sentences. If a participant misses out on one word, the rest of the sentence could help in understanding what the sentence is about. A disadvantage of this is that the participants thought the content of the Wikipedia sections was too long. It takes long to come across what they wanted to hear.

The participants of this experiment are used to accessing a textual version of the Web. Some of the participants therefore preferred text, pictures and graphs over voice, because they are used to it and could therefore find the information they are searching for faster. However they still think the voice interfaces can be helpful for people who cannot access the textual web, because they have trouble with sight or cannot read.

While constructing the voice interfaces and thinking of questions for the different question sets, we noticed that DBpedia contains very little information. It was very hard to find questions, which can be answered with Wikipedia and DBpedia. Wikipedia contains much more information than DBpedia, and therefore also much more useful information.

Also the DBpedia voice interface is domain specific, because the different properties in DBpedia are hard to distinguish. The properties hardly have any properties themselves to distinguish them from each other. A disadvantage of the voice interface being domain specific is that it takes more time to add a new domain, because before adding a new domain first the domain has to be studied and the queries and section menu have to be changed for each new domain.

10. CONCLUSION & FUTURE WORK

We elicited requirements to answer research subquestion 1. From this we can conclude, that requirements of good voice interfaces for Wikipedia and DBpedia con-

cepts are, using DTMF as input source, provide the user with an overview of the page, eliminate repetitions, provide the user with informative and immediate feedback and error recovery feedback, and use a female TTS voice.

To answer research subquestion 2, we made conversion algorithms for converting Wikipedia and DBpedia concepts, which are functioning efficiently. For Wikipedia we used the HTML heading elements to make the section and subsection menus, and all paragraphs and list items between two heading elements as section content. For DBpedia we used SPARQL queries to ask for the needed information.

To be able to answer research subquestion 3, we compared the two voice interfaces based on three aspects: speed, error rate and usability. For speed there is no significant difference between the Wikipedia and DBpedia voice interface. None of the two is faster than the other. Also they do not have a learning curve.

Besides speed, there is also no significant difference in error rate, except for question 2 of question set 1. For this question, using the Wikipedia voice interface resulted in more participants finding the right answer. This is probably caused by the answer being located at a strange place in the DBpedia voice interface, which is done by DBpedia itself.

Also there is in general no significant difference in usability between the two voice interfaces, except when we compared the usability of the Wikipedia and DBpedia voice interface, when they both are tested first by the participants. In that case Wikipedia has a higher overall satisfaction score and a higher information quality score. This difference can be caused by a difference between participant groups or the impression of the first voice interface influences how the user looks at the second voice interface.

The answer to the main research question is that to make information from Wikipedia efficiently available using voice interfaces for GSM, the above mentioned requirements should be met and the methods used in this research to convert Wikipedia and DBpedia should be considered, because they are functioning efficiently. Also it does not matter if normal data, Wikipedia, or Linked Data, DBpedia, is used to make the information from Wikipedia available via voice interfaces. The voice interfaces of these two sources do not differ in speed, error rate and usability.

However the DBpedia voice interface is domain specific, which means more work is needed to change it to be used for another domain compared to the domain independent Wikipedia voice interface. If DBpedia would add more super-properties to the DBpedia properties, this problem could be solved. Also when the DBpedia voice interface is set up, it works as efficient as the Wikipedia voice interface. Another disadvantage of DBpedia is that it contains less useful information than Wikipedia, but DBpedia is still growing and improving, in the future this disadvantage will probably disappear. According to this, our conclusion is that it does not matter which one is used, they are both efficient methods to make Wikipedia information available using voice interfaces.

The voice interfaces can be used in developing countries, but to be able to use them there they should be changed. They should use the local languages of the country where it will be used and have a local phone number. Apart from that they should also be tested with locals of these countries outside a lab. Also research should be done, to see what information these people would like to gain access to.

In this research, the scope was narrowed down to only information about crops. In further work the voice interfaces can be changed to have access to whole Wikipedia and DBpedia.

A few additions we recommend to be added to the voice interfaces are a repeat option and an option to go back only one step instead of to the main menu. Another addition can be an option to also input voice instead of only key input.

Acknowledgement

I would like to thank all people who helped me throughout this master project. A special thanks to my supervisor Victor de Boer, without him this project could have never been done. Also I would like to thank Chris van Aart for being my second reader and the company SG websites for providing me with a piece of server space for the voice interfaces.

References

Aron, J. (2011). How innovative is apple's new voice assistant, siri? *New Scientist*, *212*(2836), 24.

Bischoff, A. (2006). Podcast based m-learning with pediaphon-a web based text-to-speech interface for the free wikipedia encyclopedia. In *7th international conference'virtual university'vu'06* (pp. 173–176).

Bon, A., de Boer, V., Gyan, N. B., van Aart, C., De Leenheer, P., Tuyp, W., ... others (2013). Use case and requirements analysis in a remote rural context in mali. In *Requirements engineering: Foundation for software quality* (pp. 331–346). Springer.

Cohen, M. H. (2004). *Voice user interface design.* Addison-Wesley Professional.

Corporation, V. (2014). *Voxeo evolution.* Retrieved 13 January 2014, from http://evolution.voxeo.com/

Dahl, D., & Froumentin, M. (2012, February). *Voice application development.* World Wide Web Foundation.

De Boer, V., De Leenheer, P., Bon, A., Gyan, N. B., van Aart, C., Guéret, C., ... Akkermans, H. (2012). Radiomarché: distributed voice-and web-interfaced market information systems under rural conditions. In *Advanced information systems engineering* (pp. 518–532).

De Boer, V., Gyan, N. B., Bon, A., Tuyp, W., van Aart, C., & Akkermans, H. (2013). A dialogue with linked data: Voice-based access to market data in the sahel.

Fendelman, A. (2014). *What is 2g? definition of 2g: Cell phone glossary.* Retrieved 12 January 2014, from http://cellphones.about.com/od/phoneglossary/g/2g.htm

Foy, D. (2013, October). *Airtel wikipedia zero partnership to pilot wikipedia via text.* Retrieved 19 October 2013, from http://blog.wikimedia.org/2013/10/24/airtel-wikipedia-zero-text-trial/

Goble, C., Harper, S., & Stevens, R. (2000). The travails of visually impaired web travellers. In *Proceedings of the eleventh acm on hypertext and hypermedia* (pp. 1–10).

Google. (2011). *Google documents.* Retrieved 11 June 2014, from http://www.google.com/google-d-s/createforms.html

Jacobs, I., & Walsh, N. (2004, December). *W3c.* Retrieved 12 October 2013, from http://www.w3.org/TR/webarch/

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., ... others (2013). Dbpedia-a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal.*

Lewis, J. R. (1995). Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*(1), 57–78.

Lucas, B. (2000). Voicexml. *Communications of the ACM*, *43*(9), 53.

Mecanovic, D., & Shi, H. (2005). Voice user interface design for a telephone application using voicexml. In *Web technologies research and development-apweb 2005* (pp. 1058–1061). Springer.

Mendes, P. N., Jakob, M., & Bizer, C. (2012). Dbpedia: A multilingual cross-domain knowledge base. In *Lrec* (pp. 1813–1817).

Miniwatts Marketing Group. (2012, June). *Internet world stats - usage and population statistics.* Retrieved 29 December 2013, from http://www.internetworldstats.com/

Patel, N., Agarwal, S., Rajput, N., Nanavati, A., Dave, P., & Parikh, T. S. (2009). A comparative study of speech and dialed input voice interfaces in rural india. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 51–54).

Schmiedl, G., Seidl, M., & Temper, K. (2009). Mobile phone web browsing: a study on usage and usability of the mobile web. In *Proceedings of the 11th international conference on human-computer interaction with mobile devices and services* (p. 70).

Schnelle, D., & Lyardet, F. (2006). Voice user interface design patterns. In *Europlop* (pp. 287–316).

Sharma Grover, A., Stewart, O., & Lubensky, D. (2009). Designing interactive voice response (ivr) interfaces: localisation for low literacy users.

Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N., & Rosenfeld, R. (2009). Speech vs. touch-tone: Telephony interfaces for information access by low literate users. In *Information and communication technologies and development (ictd), 2009 international conference on* (pp. 447–457).

Soper, D. (2006). *Statistics calculators.* Retrieved 1 January 2014, from http://www.danielsoper.com/statcalc3/

UNCTAD. (2007). Science and technology for development: the new paradigm of ict. In *United nations conference on trade and development.*

UNESCO. (2010). *Index mundi 2010.* Retrieved 29 December 2013, from http://www.indexmundi.com/facts/indicators/SE.ADT.LITR.ZS/map/africa

Wikipedia. (2013). *Wikipedia:wikiproject spoken wikipedia.* Retrieved 13 January 2014, from http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Spoken_Wikipedia

Wikipedia. (2014). *Wikipedia.* Retrieved 19 June 2014, from http://www.wikipedia.org/

Yankelovich, N., Levow, G.-A., & Marx, M. (1995). Designing speechacts: Issues in speech user interfaces. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 369–376).

Yesilada, Y., Harper, S., Goble, C., & Stevens, R. (2004). Screen readers cannot see. In *Web engineering* (pp. 445–458). Springer.

APPENDIX

A. Wikipedia conversion algorithms

This Appendix contains the activity diagrams for the three conversion steps of Wikipedia. Figure 6 displays the diagram of the 'Play section options' step, figure 7 displays the diagram of the 'Play subsection options' step and figure 8 displays the diagram of the 'Read (sub)section' step.

Figure 6: Wikipedia 'Play section options' step

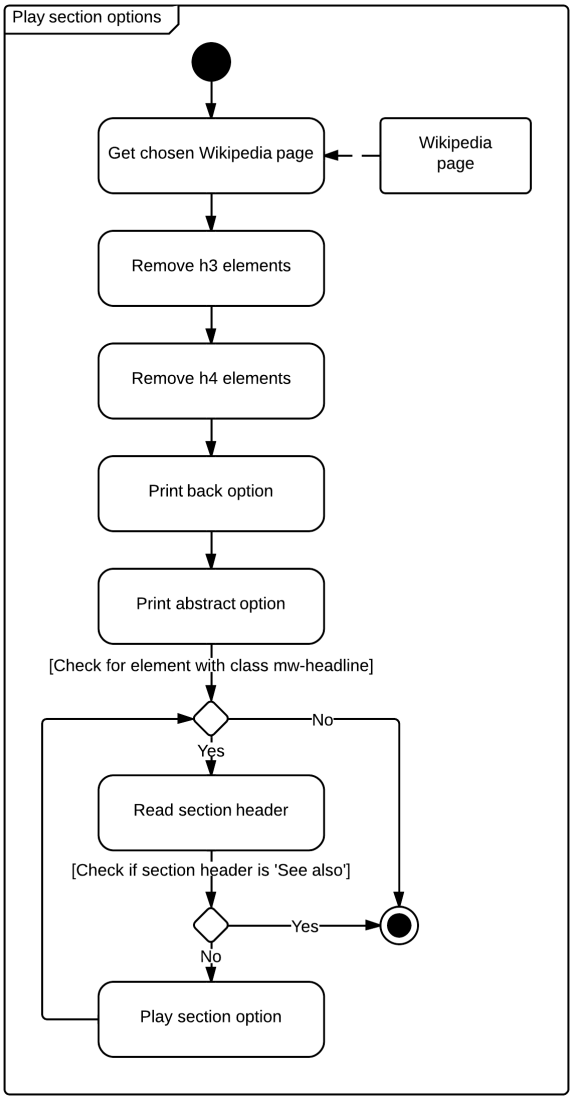


Figure 7: Wikipedia 'Play subsection options' step

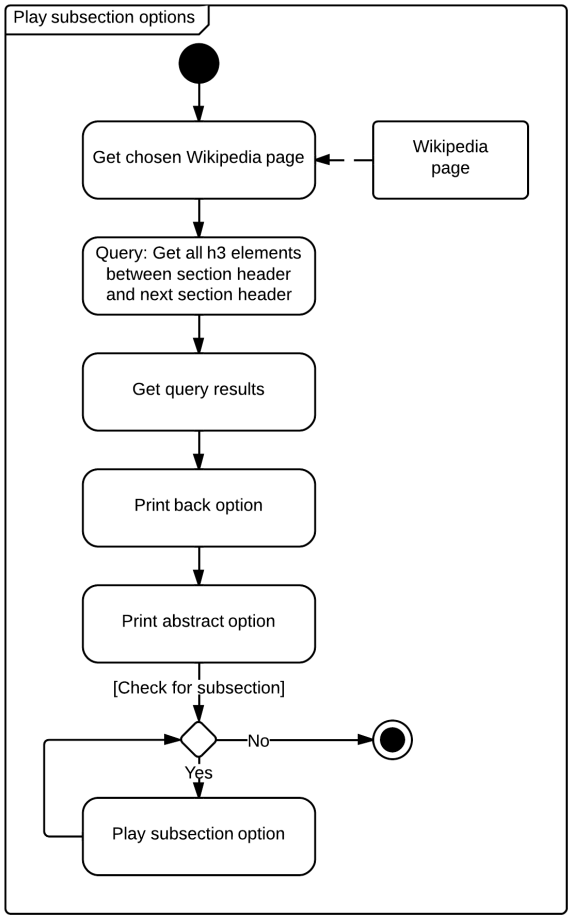
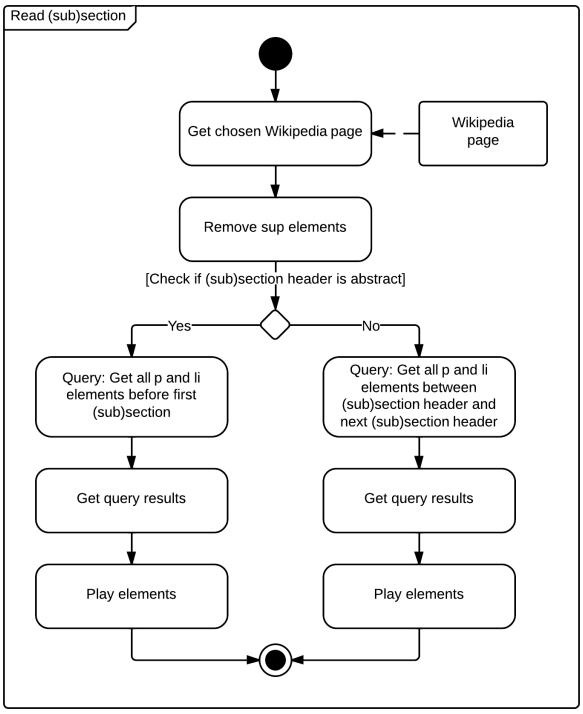


Figure 8: Wikipedia 'Read (sub)section' step



B. DBpedia conversion algorithms

This Appendix contains the activity diagrams for the three conversion steps of DBpedia. Figure 9 displays the diagram of the 'Play section options' step, figure 10 displays the diagram of the 'Play subsection options' step and figure 11 displays the diagram of the 'Read (sub)section' step.

Figure 9: DBpedia 'Play section options' step

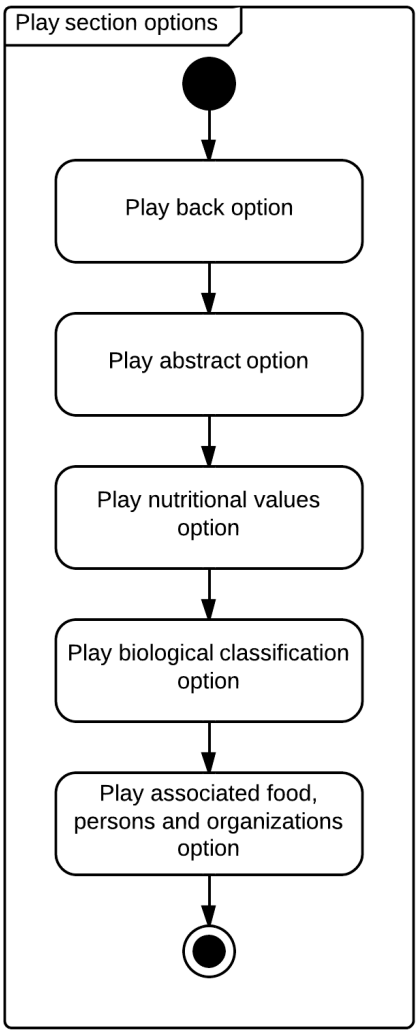


Figure 10: DBpedia 'Play subsection options' step

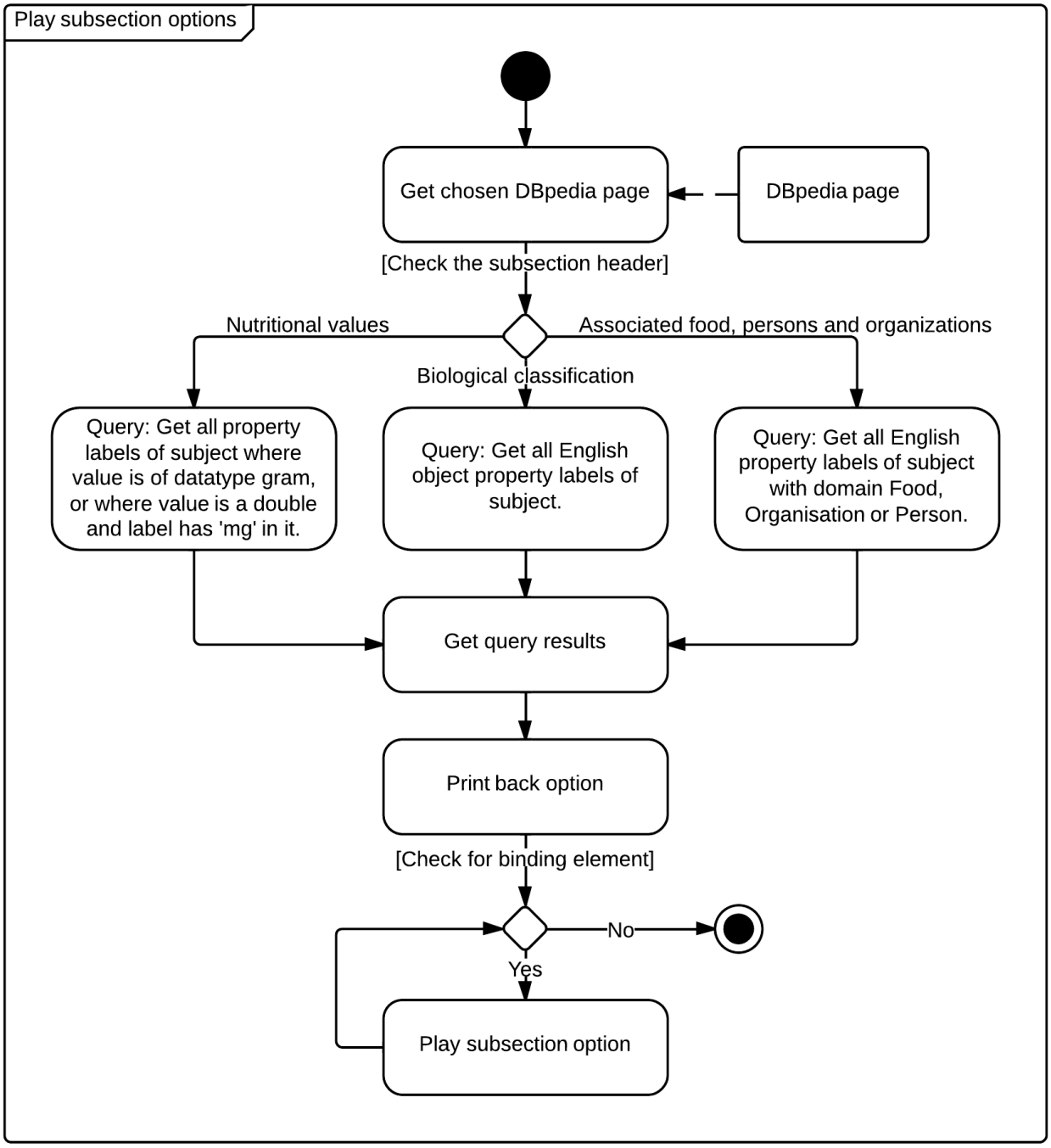


Figure 11: DBpedia 'Read (sub)section' step