



Spatially-Aware Information Retrieval on the Internet



SPIRIT is funded by EU IST Programme
Contract Number: IST-2001-35047

Evaluation of the Interim Prototype

Deliverable number:	D20 7202
Deliverable type:	R
Contributing WP:	WP 7
Contractual date of delivery:	1 July 2004
Actual date of delivery:	8 July 2004
Authors:	Guillaume Aumaitre Sandrine Balley, Stephen Levin, Bénédicte Bucher
Keywords:	Evaluation, Methodology, Usability

Abstract: This document describes the methodology used by the SPIRIT consortium to evaluate SPIRIT interim prototype.

Contents

1.INTRODUCTION: OBJECTIVES OF THIS EVALUATION	4
2.EVALUATION METHOD.....	5
2.1.Functionalities of the interim prototype	5
2.2.Method	6
Summary of the methodology proposed in D19	6
Adaptation of the methodology to the interim prototype	7
3.SYSTEM VALIDATION.....	7
3.1.Functionalities checking.....	7
3.2.Stability test	7
4.USER VALIDATION.....	8
4.1.Testing protocol	8
Introduction	9
Performing tasks	9
4.2.Results	11
4.3.Lessons learned concerning the evaluation methodology	12
5.CONCLUSION:	13
6.REFERENCES:.....	13
7.APPENDIX:.....	14
APPENDIX A: THE DOCUMENTS USED	15
APPENDIX B: THE RESULTS.....	28

Executive Summary

This report describes the evaluation of the SPIRIT interim prototype, in the context of the user evaluation task of WP7. The evaluation is based on the general evaluation methodology detailed in WP7-D19. This general methodology has been designed for the evaluation of both interim prototype and final prototype.

The first part of the report describes how the general methodology is specified to fit the interim prototype evaluation. The second part analyses the results of the evaluation tests.

This first evaluation shows that units implementation and integration have been successfully performed. It yielded positive results regarding the interface which was found easy to use. However, users pointed out the lack of functionalities and minor issues in the results presentation.

D20 7202

Evaluation of the Interim Prototype

1. Introduction: objectives of this evaluation

The interim prototype has been designed to demonstrate the operability of the concepts and search functionalities introduced in SPIRIT (ontology, spatial index, disambiguation, etc.). The main challenge of the interim prototype was the integration of various units implemented by different partners.

The interface of the interim prototype only provides textual display of information as shown in Figure 1. Although a user-evaluation could be seen as not adapted to the interim prototype, it remains very useful. As an intermediate evaluation, it enables us to detect potential problems at an early stage before development is completed, provides a deeper understanding of the users' expectations and impressions of the system, and give important feedback to improve the final evaluation.

The screenshot shows the SPIRIT web interface. At the top left is the SPIRIT logo, a blue circle with white arrows pointing outwards. To the right is a link for "About SPIRIT". Below the logo is a blue header bar with the text "Structured Query". The main search area has a light blue background. It contains a "Search for" field with the text "hotel". Below this is an "In" section with three input fields: "lancaster" (labeled "Town or City name"), an empty field (labeled "Region name"), and a dropdown menu showing "United Kingdom" (labeled "Country name"). A "search" button is located below the input fields. The search results are displayed in a light blue box, showing "You are searching for hotel in United Kingdom, Great Britain, England, Lancashire, Lancaster". Below this are four numbered links:

- 1 [Accommodation in Lancaster Hotels, Guest Houses, Bed and Breakfasts, Inns, Lodgings, Motels](#)
- 2 [Mushroom Guide to Lancaster, Lancashire, UK. Where to Buy the Lancaster map.](#)
- 3 [Mushroom Guide to Lancaster, Lancashire, UK. Transport and Travel in Lancaster.](#)
- 4 [Mushroom Guide to Lancaster, Lancashire, UK. The University of Lancaster.](#)

Figure 1: Interface of the interim prototype.

2. Evaluation method

A first test introduced in the general evaluation methodology is to compare the system functionalities that appear in the system specification and the functionalities that are really available in the system. It is described in section 2.1. This test recapitulates the functional specification of the interim prototype (see column 1 and 2 of Table 1), which will trigger the specification of the evaluation method in the following section (section 2.2). Indeed, the general evaluation method must be specified depending on the prototype functionalities.

2.1. Functionalities of the interim prototype

The following spreadsheet (Table 1) lists all the functionalities that were specified for SPiRiT search engine. The first and second column indicate whether or not the functionalities were specified in SPiRiT's specifications [WP1-D4] and in the interim prototype's specifications [WP1-D16]. The last column concerns the functionalities that were available in the interim prototype during the test period. "Available" means that the functionality has been implemented in the interim prototype and that it can be used. "Yes" only means that the functionality produces an action on the system. It doesn't mean that the functionality has been verified to perform as expected.

Functionalities	Specified in SPiRiT specifications (WP1-D4)	Specified in Interim Prototype specifications (WP1-D16)	Available in the implemented Interim Prototype
Textual interface (what-where-how)	Yes	Yes	Yes
Spatial relationships ("in", "North Of", "East Of", "South Of", "West Of", "Near")	Yes	Yes ("in" for spatial index only)	Yes ("in" for spatial index only)
Graphical interface	Yes	No	No
Sketching interface	Yes	No	No
Fuzzy geographical areas	Yes	No	No
Textual result presentation	Yes	Yes	Yes
Graphical result presentation	Yes	No	No
Result presentation choice	Yes	No	No
Usual relevance ranking	Yes	Yes	Yes
Ranking respect to spatial relation	Yes	Yes	Yes ₃
Relevance ranking choice	Yes	Yes	Yes ₃

Search refinement	Yes	No	No
Choice of number of results to display	N/A	Yes	Yes ₁
Query disambiguation (spatial)	Yes	Yes	Yes ₂
Query disambiguation (no spatial)	Yes	No	No
Query expansion	Yes	Yes	Yes ₂
Text indexing	Yes	Yes	Yes
Spatial indexing	Yes	Yes	Yes
Use of Google web API	Yes	No	No
Logging feature	Yes	Yes	Yes

(1): this functionality is available in the developer-interface ("alpha" interface)

(2): this functionality can be switched on or off on the developer-interface. User-interface doesn't enable users to switch on or off.

(3): developer-interface provides choice between several algorithms.

Table 1: Specifications and implemented functionalities of the interim prototype

The comparison between first and second column shows what needs to be done to reach the final functionalities.

This spreadsheet shows that no graphical functionality is available in the first interim prototype. This prototype enables the user to perform a spatial search by using the textual interface.

In addition, the ontology only contains data from United Kingdom. So the searches are limited to the UK area. Thus, to demonstrate the functionalities currently available, the Interim prototype has been implemented with a subset of the 1Tb collection (containing over than 94.5 million pages). This smaller collection called "Interim collection" contains 20000 pages from UK.

2.2. Method used

Summary of the methodology proposed in D19

The "Evaluation Methodology" [WP7-D19] proposes a comprehensive methodology to be used for the evaluation of SPIRIT prototypes. This methodology is composed of two major steps:

- the system validation, i.e. evaluating how well the system works, with its technical specifications as a reference, which is a rather quantitative approach,
- the user validation, i.e. checking if the system fits their requirements, which is a qualitative approach.

In SPIRIT, the system validation relies on two tests. One is to compare the system functionalities that appear in the system specification and the functionalities that are really available in the system (see section 2.1). The other is IR performance analysis based on the measuring of two criteria: "Recall" and "Precision". These two criteria respectively correspond to the probability that a relevant document is retrieved and the probability that a retrieved document is relevant. These measurements are made possible by indexing a collection of documents and searching for relevant documents inside it before running the performance test.

In addition to the measure of recall and precision, a stability test aims at measuring the effect of slight changes in the queries.

The user validation is based on an opportunity for users to use SPIRIT in a controlled environment. There are three main aspects to be considered: effectiveness, efficiency and user satisfaction. In the general methodology, two tests are proposed. In the first test, users are asked to perform search tasks that are explained in test scenarios. When the searches are completed, they are asked to fill in a satisfaction questionnaire. The test ends with an interview session. The second test, is a comparative usability test in which users compare SPIRIT's interface with other search engines' interfaces.

Adaptation of the methodology to the interim prototype

Regarding functionalities that are available in the Interim prototype, some of the tests described above won't be performed during the interim prototype evaluation. Another important factor is that the interim prototype indexes a subset of the 1Tb collection that is not representative of the information available on the Internet (only UK documents are available in the Interim collection).

- System validation

Regarding the measure of recall and precision, before running such a performance test, assessors have to manually search for relevant documents for several search topics. This is a time-consuming task that has to be redone each time the collection indexed by SPIRIT changes. Regarding the size of the interim collection, which is not representative of the Internet, and the amount of work to prepare a performance test, which will have to be redone for the evaluation of SPIRIT's final prototype [WP7-D31], it seems irrelevant to perform a performance test for the Interim prototype. However, the system stability can be rapidly tested. The tests performed in the system validation step and their results are described in section 3.

- User validation

Because the documents indexed by the interim prototype are not representative of the Web, the effectiveness of information retrieval may be biased. The interim prototype will mainly be tested regarding the efficiency and satisfaction aspects. There will be no comparison test with other spatial search engines as the interim prototype's interface is still in an early state and comparison test's results wouldn't be relevant. The tests performed in the system validation step and their results are described in section 4.

3. System validation

3.1. Functionalities checking

In section 2.1 the comparison of the two last columns of the spreadsheet (Table 1), taking WP1-D16 specifications as a reference, constitutes the first test of the system validation step.

The result of this comparison is that all the functionalities specified in WP1-D16 have been implemented in the interim prototype.

3.2. Stability test

A first stability test has been performed by introducing slight variations in the formulation of a query. The query was adapted to the interim prototype context. The introduced variations are:

- Spelling differences and use of special characters in the "Search for" search box,
- Spelling differences in the "In" search boxes,
- Inversions in the use of "town" and "region" boxes,

- Search of a place using the “Search for” box instead of the “In” boxes.

The queries were submitted by a single assessor, on the textual interface of the interim prototype shown in Figure 2. For each query, the first 20 returned results were recorded in a log file and compared after the session.

The image shows a search interface with a light blue background. At the top, there is a white rectangular box labeled "Search for". Below this box, the word "In" is centered. Underneath "In", there are three input fields: a wide one on the left, a medium one in the middle, and a dropdown menu on the right showing "United Kingdom". Below these fields are labels: "Town or City name" under the wide field, "Region name" under the medium field, and "Country name" under the dropdown menu.

Figure 2: Detailed view of the textual querying interface of the interim prototype.

The aim of this test was to have a precise idea of the system's behaviour, whether it is deliberate or not (i.e. caused by a bug). The following behaviours were pointed out:

- Misspellings in the non-spatial component of the query are not supported by the system and generate errors.
- Misspellings in the spatial component of the query do not generate errors but the system fails to recognise the place and returns no result.
- Several words are allowed in the non-spatial component. They can be separated by a comma without changing the returned results. Quotes and dashes generate errors.
- If a place is specified by one place name and one country name (e.g. Edinburgh, UK), the place name can be indifferently put in the “town” box or in the “region” box without changing the results.
- If a place name is specified by three components (e.g. Edinburgh, Scotland, UK), the order of query boxes (town, region, country) must be respected. If it is not, the system fails to recognise the place.
- The same place specified more or less precisely (e.g. “Edinburgh, UK” or “Edinburgh, Scotland, UK”), even unambiguous, returns a different list of results.
- When the spatial functionalities of SPIRIT are running, the place name must be specified in the spatial component of the query. If it is not (for instance, if the user types “hotel London” in the “Search for” box and nothing in the “In” boxes), an error occurs.

If some of these behaviours are system's features (as specifying a place indifferently in the “town” box or in the “region”), others may be caused by bugs, or may occur only because every possible query structure and users error were not anticipated in the interim prototype design. In any case, this stability test provides extremely useful information for the prototype improvement.

4. User validation

4.1. Testing protocol

8 users were involved in the test (4 in France, and 4 in Sheffield, UK).

For the final evaluation, users have to be representative of the end users. For SPIRIT, it means they belong to various fields of activity and are all familiar with the use of search engines.

Because tests are time-consuming, recruiting users from various domains (and thus, with various time schedules) is not an easy task. Moreover, users have to be different for the interim and for the final prototype's evaluation, which requires two distinct recruitments.

Among users we could recruit, we decided to keep a wide panel for the evaluation of the final prototype, and to keep fewer users for the evaluation of this interim prototype. Their domains are finances, information sciences, computer sciences, engineering. More then looking for a set of representative SPIRIT users, we recruited users that had a fairly good experience in using search engines (at least several times a week). In addition, all of them have already used mapping services on the internet.

Three nationalities were represented among the users. Three of them speak native english, three speak english fluently and two fairly well. A large range of ages is represented (from under 25 to above 54).

Each user session approximately lasted 1h30 and is composed of 4 major steps that rely on documents provided in appendix A of this report.

Introduction

Users are welcomed and are orally given general information about SPIRIT: subject and goals. They are also given details about the evaluation: how the evaluation will be running and what will be evaluated. These details are written in "instructions" document. (See Appendix A)

The introduction is also the step in which user profile information is collected through a "personal profile" questionnaire. (See Appendix A)

The introduction step is calibrated to last less than 10min.

Performing tasks

Users are given four scenarios. Each of them describes a specific kind of information that should be looked for using SPIRIT. This aims at guiding users to formulate queries consistent with the interim prototype capacities without directly giving them such explicit queries. The search topics were defined according to the indexed collection and the interim prototype's available functionalities. Practically, these can only be searches related to the United Kingdom with "in" as a spatial relationship.

While they are performing the searching tasks, users are asked to think aloud (the goal being to make sure they explain their interpretation of the contents of each screen, and their reason for making choices) and one assessor per user observes the interactions with the system and notes down every problem encountered. Users are not given any hints or assistance unless they are unable to complete the searching task. Each searching task is timed and limited to 15min so the performing tasks step doesn't exceed 1h.

The four search task (topics and scenarios) are detailed below:

- Non business topic:

This topic aims at retrieving general information principally focusing on tourism and involving a precise localisation search.

The topic used during the user test is "Walking in London" and the associated scenario is:

"Marie and John are visiting Britain this summer. They planned to stay only a couple of days in London, so they want to plan their trip as much as possible before they leave. They intend to do all their sightseeing by foot, and want to avoid museums and other indoor attractions wherever possible. Imagine you are Marie or John. Try and find as many relevant documents as you can relating to walking around London as a tourist, for example information about guided tours, details of other peoples' experiences, etc. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task."

- Vague items, large areas topic:

This topic intends to illustrate a search on a large area. Users may use SPiRiT as a helping tool to find information about vague items. The topic used during the test is “Castles in Wales” and the associated scenario is:

“John plans to visit Wales next Spring. He is particularly interested in visiting the various castles around the country, but he has no idea of where they are located and which are the ‘must see’ ones. Imagine you are John. Try and find as many relevant documents as you can relating to Welsh castles which may help you plan your trip. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.”

- Events, habits related topic:

This topic aims at finding information about events that took place in certain places or habits of certain locations. The topic used during the test is “Ceilidh dancing in Glasgow” and the associated scenario is:

“George and Laura are planning a visit to Scotland in 2001. They are particularly interested in the local cultures of the areas they intend to visit, and are planning on visiting Glasgow during summer, when there are several Scottish festivals. They particularly would like to discover ceilidh dancing. Imagine you are George or Laura. Try to find as many documents relating to ceilidh dancing as you can, which may help you to plan your stay in Glasgow, for example information about the festivals including ceilidh dancing sessions. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.”

- Business topic:

This topic aims at finding commercial services, as it could be done on Yellow Pages or Local Google search engines.

The topic used during the test is “Hotels in Cardiff” and the associated scenario is:

“Remy will be attending an academic workshop taking place in Cardiff later this month. He needs to book a hotel room for the duration of the conference. Imagine you are Remy. Try to find as many documents as you can relating to this topic. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.”

This scenario is completed with a search that involves the disambiguation feature:

“During the workshop, a social event is planned in a city located near Cardiff: Newport. Once he knows where to stay in Cardiff, Remy would also like to obtain general tourist information (things to see and to do) about this city. Try to find such general information. You do not need to bookmark documents.”

Satisfaction questionnaire

When users have completed their searching tasks, they are asked to fill in a satisfaction questionnaire.

The satisfaction questionnaire is based on QUIS 5.0 questionnaire which is composed of 5 parts: “1. Overall reaction to the system”, “2. Screen”, “3. Terminology and System Information”, “4. Learning” and “5. System capabilities”. [Chin et al., 1988]

This questionnaire has been enriched with some questions dealing with ergonomic aspects that are not taken into account in QUIS questionnaire. As QUIS is a very general questionnaire, some questions do not apply at all to SPiRiT and were removed.

The satisfaction questionnaire step is intended to last approximately 10min.

Both QUIS 5.0 and the satisfaction questionnaire used for the user test are provided in Appendix A.

Interview

The user test ends with the interview session. Assessors orally ask questions noted in a questionnaire and note down users' remarks (see Appendix A). The questionnaire used for the interview is functionally-centred.

Assessors can also ask questions about encountered problems or extreme answers done by the user in the satisfaction questionnaire.

4.2. Results

The satisfaction questionnaire gives an outlook on how the users feel with the interim prototype. Users Opinion is expressed by grades from 1 to 5 for each question. Even if it relies on 8 users only, we consider this outlook is reliable, as the standard deviation of each question is low (around 1). The average grade, calculated on all the users, is an indication of the satisfaction. The grade obtained is medium (around 3): obviously, users did not dislike the prototype, but were not attracted by it either.

To have a better understanding, two other grades are calculated. Indeed, some questions in the satisfaction questionnaire refer more to the efficiency and others refer more to the effectiveness. The grades calculated on these questions give an idea of the efficiency and an idea of the effectiveness.

It appears that the efficiency is judged better than the effectiveness (3.3 and 2.8 respectively), which means that the use of SPIRIT is globally satisfying and easy, but the results obtained are not good enough.

The plain results are available in Appendix B. A summary of these results, obtained with the confrontation between the satisfaction questionnaire, the interview and the assessors notes, is provided below :

- General appreciation

As a general appreciation, users hardly noticed innovative features introduced by SPIRIT. They thought the system doesn't provide more features than other systems like Google. This is corroborated by a low grade concerning the "overall reaction toward the system". They were surprised not to have a graphical interface, or at least, maps, that were judged to be an important element in a spatial search engine. The system was also considered as being too slow.

However, they appreciated the system's simplicity, place recognition feature and disambiguation, which may explain the medium grade of satisfaction.

- Querying interface

As indicated by the reasonably good efficiency grade, there is no major issue with the interface. The users' most appreciated aspect of the interface is its simplicity. They qualified its labelling and layout as being sober, basic and clear.

However there are some complaints about region and town boxes, which are too ambiguous (especially "region"). Indeed, some users tried to use the region box to express a fuzzy region field like "Region of Cardiff". In addition, some users formulated queries without using the location boxes like "Welsh Castles" instead of "Castles in Wales".

- Results display and relevance

The most common complaints concern the results presentation and relevance ranking.

First, users said that results were better presented in Google. They found too little information was provided: the number of retrieved documents is not displayed, and for each result, no description is provided but its title.

Users would like "snippets" (a few lines taken from the site body, like in Google), or at least the display of URLs. Several results may have the same title or, on the contrary, some identical documents may be listed with different titles.

A clustering of the results would also be appreciated.

In addition, users noticed that it was not possible to choose the number of displayed results and that the interface did not adapt to the screen size.

Secondly, relevance of documents retrieved by the system was judged to be too average. Some users had the feeling that retrieved documents were always the same. In general, they found relevance to be better in term of location than subject. But they also experienced some problems of badly localised documents (notably, when a location was mentioned in a different context in the document).

The ranking seemed insufficient but was very difficult to judge. No information concerning the ranking criteria (distance to the specified location or theme relevance) is provided and some users were not aware that the results were ranked. Some users had the feeling that the popularity should have been taken into account (as in Google).

Disambiguation was thought to be a useful feature. However, even if a place has been disambiguated, the search engine may retrieve some spurious documents. In addition, the disambiguation process does not include enough help features : when users are unfamiliar with the region, they don't know which place to choose. In such cases, they would appreciate the possibility to search on all or several places.

- Assistance

An important drawback raised by the users is the lack of help features. Indeed, there is no explanation on how to proceed. Some users felt confused at the beginning of the test because they didn't know whether they can use boolean operators, quotes or truncated queries.

In addition, in case of misspelling, they mentioned some bugs and a lack of help.

4.3. Lessons learned concerning the evaluation methodology

Some of the difficulties experienced by users were inherent to the way of testing. These troubles and the assessors' observations provide useful indications to refine and improve our methodology in prevision of the final prototype evaluation.

- Search scenarios

As defined in the general methodology (WP7-D19), the search scenarios were designed to guide both the usability evaluation and the performance evaluation. In fact, these scenarios proved to be too restrictive for the usability testing: users were given terms they just had to use in their queries. They had difficulties to judge the relevance of documents as they were not sure of the kind of responses the fictive characters were looking for. Users probably would be more implied and honest with free search scenarios or very vague topics, letting them choose their own subject.

Moreover, these defined scenarios did not suit performance evaluation well: they are not precise and restrictive enough. Users hesitated in bookmarking relevant documents, because they did not know what the fictive users were exactly looking for, and what the relevance criteria were.

- User-sessions and performance measures

These contradictory requirements regarding scenarios for usability and performance testing lead us to rethink our evaluation method: it is impossible to reconcile these two evaluation directions, which were however seen as two sessions of the same test in D19.

This necessary separation is corroborated by another fact: real users are not relevance assessors. Indeed, searching for relevant information and assessing the relevance of a list of retrieved documents are two different tasks. The analysis of log files and of bookmarked results showed how variously users react in real situation. Some of them stop the search as soon as they have found an answer to their question, while the others go on with the search and bookmark several documents providing approximately the same information. In addition, the log files showed that some users visit only documents whose title looks interesting, while others visit them all and only judge their content.

For the last prototype evaluation, which will include a performance assessment of the search engine, we will define two processes and scenarios, and definitely distinguish users and assessors.

- Information provided to users during the test

Our users were confused about some questions of the satisfaction questionnaire which are a somewhat vague. Some of them will be rephrased. Users were also confused about the number of broken links when they tried to go on the Web, and the fact that the search takes place within a limited documents collection. This information should be a part of the introductory speech.

5. Conclusion:

The evaluation of the interim prototype was fruitful. It yielded useful results regarding the evaluation of the work done so far, the prototype further design and the improvement of the evaluation methodology. It also let the SPIRIT team experiment evaluating a prototype. Last, the preparation of users tests launched an iterative process between WP7 and WP4 (User interface) to improve the prototype.

Concerning the interim prototype itself, it was working in due time which means the units implementation and integration had been performed successfully. The users globally praised the learnability and simplicity of the interface which is a very satisfying result that wasn't obvious before the test.

However, they were critical of the lack of functionalities. Indeed, given the early stage of the evaluation, minimal functions were actually available in this interim prototype.

Some clues regarding the improvement of the interface (see section 2.3: results), and the ranking were collected.

Concerning the evaluation methodology, this evaluation was a rather successful attempt to apply the theoretical methodology exposed in D19. Yet, there are some lessons learned that will lead to slight changes of this methodology or further specification of it in order to improve the evaluation method for the final prototype. The main one is that users need to experiment their own scenarios in order to truly evaluate the prototype. The final prototype should thus support a wide range of scenarios. Indeed, it will work on a larger collection (the 1 Tb collection), and have more functions. Last, presenting to users a prototype with so little functionality, which again was no surprise, motivated very much the evaluation team for presenting next year a prototype with all SPIRIT specified functionalities.

6. References:

1. Aumaitre, G., Balley, S., Levin, S. (2004) "Evaluation Methodology", WP7-D19-7201
2. Dave Finch (2004) "Specification of System Functionality", WP1-D4-1101
3. Dave Finch et al. (2004) "Interim Prototype", WP1-D16-1101
4. Chin, John P., Diehl ,Virginia A., Norman, Kent L., (1998) "Development of a Tool Measuring User Satisfaction of the Human-Computer Interface"

7. Appendix:

APPENDIX A: THE DOCUMENTS USED 15

A.1 Introduction document 15

A.2 Personal profile questionnaire 16
English version 16
French version 17

A.3 Scenarios..... 18

A.4 QUIS 5.0 questionnaire..... 19

A.5 Satisfaction questionnaire 20
English version 20
French version 23

A.6 Questionnaire used for the interview 26
English version 26
French version 27

APPENDIX B: THE RESULTS..... 28

B.1 Personal profile..... 28

B.2 Satisfaction questionnaire compiled results 36

B.3 Interview compiled results 37

Appendix A: the documents used**A.1 Introduction document***Instructions*

- Access SPIRIT (<http://cheese.geo.unizh.ch:28080/interface>) and submit a first query:
If username and password required use:
Username: « spirit »
Password: « vodka »
- Create 4 bookmarks files in your web browser that will be used for the bookmarking of the relevant documents of each scenario (make sure the users know how to use them.)
- Note your IP address
- Welcome the users and give them the following task instructions:
- Welcome to SPIRIT first prototype user-centred evaluation session.
- The session is composed of 3 parts: the searching tasks, the questionnaire and the interview.
- During the searching tasks, you will be asked to perform 4 searches using SPIRIT. The context of these 4 searches is explained in the booklet you have been given.
- For each search, you will be asked to find and note (using bookmark functionality of your web browser) all the documents you consider as relevant.
- Each search task will be timed and won't exceed 15min.
- After the searching tasks you will be asked to fill in a satisfaction questionnaire.
- The test will end with an interview.
- Remember that only SPIRIT is being evaluated... not you !!!

- Do not give any hints or assistance unless the users are unable to complete the tasks.
- Time each task. Each searching task will end after 15min maximum.
- Observe the user's behaviours and note any problem encountered.
- If possible, try to count the number of mouse's clicks to perform each task.
- If possible, use a dynamic screen capture to shoot the first searching task.
- Ask the users to think aloud (make sure they explain their interpretation of the contents of each screen, and their reason for making choices).
- At the end of the searching session, ask the users to complete the satisfaction questionnaire.
- Interview the users following the 2nd questionnaire's questions to gain general opinions and to ask about specific problems encountered.

A.2 Personal profile questionnaire

English version

Personal Profile

Education and qualifications:

Job:

Age:

Which language(s) do you speak read or write?

(List the language(s) and tick the relevant boxes)

	<i>Maternelle</i>	<i>Courant</i>	<i>Moyen</i>	<i>Basique</i>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?

(Tick one box only)

1 <i>Never</i>	2 <i>Occasionally</i>	3 <i>Once a week</i>	4 <i>Several times a week</i>	5 <i>Daily</i>
-------------------	--------------------------	-------------------------	----------------------------------	-------------------

How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?

(Tick one box only)

1 <i>Never</i>	2 <i>Occasionally</i>	3 <i>Once a week</i>	4 <i>Several times a week</i>	5 <i>Daily</i>
-------------------	--------------------------	-------------------------	----------------------------------	-------------------

How often do you perform searches on computerized library catalogues (e.g. your library)?

(Tick one box only)

1 <i>Never</i>	2 <i>Occasionally</i>	3 <i>Once a week</i>	4 <i>Several times a week</i>	5 <i>Daily</i>
-------------------	--------------------------	-------------------------	----------------------------------	-------------------

How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)? *(Tick one box only)*

1 <i>Never</i>	2 <i>Occasionally</i>	3 <i>Once a week</i>	4 <i>Several times a week</i>	5 <i>Daily</i>
-------------------	--------------------------	-------------------------	----------------------------------	-------------------

French version

Profil personnel

Formation et qualifications :

Emploi :

Age :

Quelles sont les langues que vous parlez, lisez ou écrivez ?

(Donnez la liste de ces langues et cochez la case correspondante)

	Maternelle	Courant	Moyen	Basique
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Utilisez-vous parfois des moteurs de recherche sur Internet (par ex : Google, AltaVista, Yahoo) ?

(N'entourez qu'une seule réponse)

1	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Utilisez-vous parfois des moteurs de recherche payants (par ex : Dialog, Lexis- Nexis) ?

(N'entourez qu'une seule réponse)

1	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Effectuez-vous parfois des recherches à l'aide d'index informatisés (par ex : dans une bibliothèque) ?

(N'entourez qu'une seule réponse)

1	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Effectuez-vous parfois des recherche à l'aide de sites proposant un service géographique (par ex mappi.fr, map24.com) ?

(N'entourez qu'une seule réponse)

1	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

A.3 ScenariosUser Scenarios

1.

Marie and John are visiting Britain this summer. They planned to stay only a couple of days in London, so they want to plan their trip as much as possible before they leave. They intend to do all their sightseeing by foot, and want to avoid museums and other indoor attractions wherever possible. Imagine you are Marie or John. Try and find as many relevant documents as you can relating to walking around London as a tourist, for example information about guided tours, details of other peoples' experiences, etc. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.

2.

John plans to visit Wales next Spring. He is particularly interested in visiting the various castles around the country, but he has no idea of where they are located and which are the 'must see' ones. Imagine you are John. Try and find as many relevant documents as you can relating to Welsh castles which may help you plan your trip. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.

3.

George and Laura are planning a visit to Scotland in 2001. They are particularly interested in the local cultures of the areas they intend to visit, and are planning on visiting Glasgow during summer, when there are several Scottish festivals. They particularly would like to discover ceilidh dancing. Imagine you are George or Laura. Try to find as many documents relating to ceilidh dancing as you can, which may help you to plan your stay in Glasgow, for example information about the festivals including ceilidh dancing sessions. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.

4.

Remy will be attending an academic workshop taking place in Cardiff later this month. He needs to book a hotel room for the duration of the conference. Imagine you are Remy. Try to find as many documents as you can relating to this topic. Bookmark any documents that you consider to be relevant. If you feel you have found enough relevant documents or cannot find anymore, you can move onto the next task.

During the workshop, a social event is planned in a city located near Cardiff : Newport. Once he knows where to stay in Cardiff, Remy would also like to obtain general tourist information (things to see and to do) about this city. Try to find such general information. You do not need to bookmark documents.

A.4 QUIS 5.0 questionnaire

(For each of the following questions choose one among: 1 2 3 4 5 6 7 8 9)

A. Overall Reactions to the system

1. (terrible/wonderful)
2. (frustrating/satisfying)
3. (dull/stimulating)
4. (difficult/easy)
5. (inadequate power/adequate power)
6. (rigid/flexible)

B. Screen

1. Characters on the computer screen (hard to read/easy to read):
2. Highlighting on the screen simplifies task (not at all/very much):
3. Organisation of information on screen (confusing/very clear):
4. Sequence of screens (confusing/very clear):

C. Terminology and system information

1. Use of terms throughout system (inconsistent/consistent)
2. Computer terminology is related to the task you are doing (never/always)
3. Position of messages on screen (inconsistent/consistent)
4. Messages on screen which prompt user for input (confusing/very clear)
5. Computer keeps you informed about what it is doing (never/always)
6. Error messages (unhelpful/helpful)

D. Learning

1. Learning to operate the system (difficult/easy)
2. Exploring new features by trial and error (difficult/easy)
3. Remembering names and use of commands (difficult/easy)
4. Tasks can be performed in a straight-forward manner (never/always)
5. Help messages on the screen (unhelpful/helpful)
6. Supplemental reference materials (confusing/clear)

E. System capabilities

1. System speed (too slow/fast enough)
2. System reliability (unreliable/reliable)
3. System tends to be (noisy/quiet)
4. Correcting your mistakes (difficult/easy)
5. Experienced and inexperienced users' needs are taken into consideration (never/always)

A.5 Satisfaction questionnaire

English version

Overall reactions to the system

1.

1	2	3	4	5
<i>terrible</i>				<i>wonderful</i>

2.

1	2	3	4	5
<i>frustrating</i>				<i>satisfying</i>

3.

1	2	3	4	5
<i>dull</i>				<i>stimulating</i>

4.

1	2	3	4	5
<i>difficult</i>				<i>easy</i>

Screen

1. Reading characters on the screen:

1	2	3	4	5
<i>hard</i>				<i>easy</i>

2. Highlighting on the screen simplifies task:

1	2	3	4	5
<i>not at all</i>				<i>very much</i>

3. Organisation of information on the screen:

1	2	3	4	5
<i>confusing</i>				<i>very clear</i>

4. Position of information on the screen:

1	2	3	4	5
<i>inconsistent</i>				<i>consistent</i>

5. Sequence of screens:

1	2	3	4	5
<i>confusing</i>				<i>very clear</i>

6. You think the system is under your control:

1	2	3	4	5
<i>never</i>				<i>always</i>

Terminology and System Information

1. Use of terms throughout the system:

1	2	3	4	5
<i>inconsistent</i>				<i>consistent</i>

2. Computer terminology is related to the task you are doing:

1	2	3	4	5
<i>never</i>				<i>always</i>

3. System informs about its progress:

1	2	3	4	5
<i>never</i>				<i>always</i>

4. Error messages:

1	2	3	4	5
<i>unhelpful</i>				<i>helpful</i>

Learning

1. Learning to operate the system:

1	2	3	4	5
<i>difficult</i>				<i>easy</i>

2. Performing tasks is straightforward:

1	2	3	4	5
<i>never</i>				<i>always</i>

3. Help features:

1	2	3	4	5
<i>unhelpful</i>				<i>helpful</i>

System capabilities

SPiRiT project*Evaluation Method*

IST-2001-35047

D20 7202

1. *System speed:*

1	2	3	4	5
<i>too slow</i>				<i>fast enough</i>

2. *System reliability:*

1	2	3	4	5
<i>unreliable</i>				<i>reliable</i>

3. *Correcting your mistakes:*

1	2	3	4	5
<i>difficult</i>				<i>easy</i>

4. *Experienced and inexperienced users' needs are taken into consideration:*

1	2	3	4	5
<i>never</i>				<i>always</i>

French version

Réactions d'ordre général au système :

1.

1	2	3	4	5
<i>extrêmement mauvais</i>				<i>excellent</i>

2.

1	2	3	4	5
<i>frustrant</i>				<i>apporte de la satisfaction</i>

3.

1	2	3	4	5
<i>ennuyeux</i>				<i>stimulant</i>

4.

1	2	3	4	5
<i>difficile</i>				<i>facile</i>

L'écran :

1. La lecture des informations à l'écran est : difficile / facile

1	2	3	4	5
<i>difficile</i>				<i>facile</i>

2. La mise en valeur visuelle (styles, couleurs...) permet une utilisation facile :

1	2	3	4	5
<i>pas du tout</i>				<i>beaucoup</i>

3. L'organisation de l'information à l'écran (son emplacement, sa densité) est :

1	2	3	4	5
<i>confuse</i>				<i>très claire</i>

4. Le positionnement des informations au cours de la navigation est :

1	2	3	4	5
<i>incohérent</i>				<i>cohérent</i>

5. Lors de l'utilisation, l'enchaînement des écrans est :

1	2	3	4	5
<i>déstabilisant</i>				<i>très logique</i>

6. Vous avez le sentiment de contrôler le système :

1	2	3	4	5
<i>jamais</i>				<i>toujours</i>

Terminologie et information sur le système :

1. L'utilisation des termes dans l'application est :

1	2	3	4	5
<i>incohérente</i>				<i>cohérente</i>

2. Les termes utilisés sont appropriés au contexte d'utilisation :

1	2	3	4	5
<i>jamais</i>				<i>toujours</i>

3. Le système informe l'utilisateur de sa progression lorsque une action est effectuée :

1	2	3	4	5
<i>jamais</i>				<i>toujours</i>

4. Les messages d'erreurs sont :

1	2	3	4	5
<i>d'aucun secours</i>				<i>très utiles</i>

Prise en main et apprentissage :

1. Apprendre à utiliser le système fut :

1	2	3	4	5
<i>difficile</i>				<i>facile</i>

2. Effectuer une recherche est une tâche simple :

1	2	3	4	5
<i>jamais</i>				<i>toujours</i>

3. Les fonctionnalités d'aide sont :

1	2	3	4	5
<i>insuffisantes</i>				<i>très utiles</i>

Performances et possibilités du système :

1. La vitesse du système est :

1	2	3	4	5
<i>trop lente</i>				<i>assez rapide</i>

2. La fiabilité du système est :

1	2	3	4	5
<i>très mauvaise</i>				<i>très bonne</i>

3. En cas d'erreur de votre part, la correction est une tâche :

1	2	3	4	5
<i>difficile</i>				<i>facile</i>

4. Les besoins des utilisateurs novices et expérimentés ont été pris en compte :

1	2	3	4	5
<i>jamais</i>				<i>toujours</i>

A.6 Questionnaire used for the interview

English version

Interview

User Interface

- 1. What did you think about the SPIRIT user interface in general?*
- 2. What aspects of the user interface did you particularly like?*
- 3. What aspects of the user interface did you particularly dislike?*
- 4. What did you think about the layout of the interface?*
- 5. What did you think about the system's ability to recognise place names and the ability to disambiguate between places with the same name?*

Search Results

- 6. What did you think about the results returned by SPIRIT?*
- 7. What did you think about the relevance of the results returned in terms of subject and location searched for?*
- 8. What did you think about the way search results were presented?*
- 9. What did you think about the information accompanying each search result?*
- 10. What did you think about the ranking of displayed search results?*

Using SPIRIT

- 11. Did you experience any difficulties when using SPIRIT?*
- 12. What did you think about the amount of support the system gave you when completing your tasks?*

Conclusion

- 13. How do you feel about the SPIRIT system in general?*
- 14. What aspects of the system did you particularly like?*
- 15. What aspects of the system did you particularly dislike?*
- 16. Are there any other comments you would like to make about your experience with the SPIRIT system?*

French version

Interview

L'interface

1. De façon générale que pensez-vous de l'interface de SPIRIT ?
2. Quels sont les aspects de l'interface qui vous ont particulièrement plu ?
3. Quels sont les aspects de l'interface qui vous ont particulièrement déplu ?
4. Que pensez-vous de l'agencement général de l'interface ?
5. Que pensez-vous de la capacité du système à reconnaître les noms de lieux et sa capacité de désambiguïser les lieux ayant le même nom ?

Les résultats de la recherche

6. Que pensez-vous des résultats trouvés par SPIRIT ?
7. Que pensez-vous de la pertinence des résultats en ce qui concerne le thème et la localisation ?
8. Que pensez-vous de la façon dont les résultats sont présentés ?
9. Que pensez-vous des informations fournies pour chaque résultat trouvé ?
10. Que pensez-vous du classement des résultats ("ranking") ?

L'utilisation de SPIRIT

11. Avez-vous rencontré des difficultés lors de votre utilisation ?
12. Que pensez-vous de la quantité d'aide et de support fournie par le système pendant que vous effectuiez une tâche ?

Conclusion

13. Quelle est votre impression générale sur SPIRIT ?
14. Quels aspects du système avez-vous particulièrement apprécié ?
15. Quels aspects du système vous ont particulièrement déplu ?
16. Y-a-t-il un commentaire que vous souhaiteriez ajouter à propos de votre expérience avec SPIRIT ?

Appendix B: the results

B.1 Personal profile

Profil personnel – U_BB_1

Formation et qualifications : Bac +8 , computer science, PHd

Emploi : computer scientist

Age : 25-34

- **Quelles sont les langues que vous parlez, lisez ou écrivez ?**
(Donnez la liste de ces langues et cochez la case correspondante)

	Maternelle	Courant	Moyen	Basique
_____ French _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____ English _____	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- **Utilisez-vous parfois des moteurs de recherche sur Internet (par ex : Google, AltaVista, Yahoo) ?**
(N'entourez qu'une seule réponse)

1	2	3	4	X
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Utilisez-vous parfois des moteurs de recherche payants (par ex : Dialog, Lexis-Nexis) ?**
(N'entourez qu'une seule réponse)

X	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherches à l'aide d'index informatisés (par ex : dans une bibliothèque) ?**
(N'entourez qu'une seule réponse)

X	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherche à l'aide de sites proposant un service géographique (par ex : mappi.fr, map24.com) ?**
(N'entourez qu'une seule réponse)

1	2	3	4	X
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Profil personnel – U_BB_2

Formation et qualifications : Bac +8 , « Institut des hautes études en sciences politiques »

Emploi : retired

Age : > 54

- **Quelles sont les langues que vous parlez, lisez ou écrivez ?**
(Donnez la liste de ces langues et cochez la case correspondante)

	Maternelle	Courant	Moyen	Basique
_____ French _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____ English _____	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- **Utilisez-vous parfois des moteurs de recherche sur Internet (par ex : Google, AltaVista, Yahoo) ?**
(N'entourez qu'une seule réponse)

1	2	3	X	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Utilisez-vous parfois des moteurs de recherche payants (par ex : Dialog, Lexis-Nexis) ?**
(N'entourez qu'une seule réponse)

X	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherches à l'aide d'index informatisés (par ex : dans une bibliothèque) ?**
(N'entourez qu'une seule réponse)

X	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherche à l'aide de sites proposant un service géographique (par ex : mappi.fr, map24.com) ?**
(N'entourez qu'une seule réponse)

1	2	X	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Profil personnel – U_SB_1

Formation et qualifications : Bac +5 , engineering

Emploi : computer engineer

Age : 25-34

- **Quelles sont les langues que vous parlez, lisez ou écrivez ?**
(Donnez la liste de ces langues et cochez la case correspondante)

	Maternelle	Courant	Moyen	Basique
<u>French</u>	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<u>English</u>	<input type="checkbox"/>	X	<input type="checkbox"/>	<input type="checkbox"/>
<u>German</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
<u>Spanish</u>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X

- **Utilisez-vous parfois des moteurs de recherche sur Internet (par ex : Google, AltaVista, Yahoo) ?**
(N'entourez qu'une seule réponse)

1	2	3	4	X
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Utilisez-vous parfois des moteurs de recherche payants (par ex : Dialog, Lexis-Nexis) ?**
(N'entourez qu'une seule réponse)

X	2	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherches à l'aide d'index informatisés (par ex : dans une bibliothèque) ?**
(N'entourez qu'une seule réponse)

1	X	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

- **Effectuez-vous parfois des recherche à l'aide de sites proposant un service géographique (par ex : mappi.fr, map24.com) ?**
(N'entourez qu'une seule réponse)

1	X	3	4	5
Jamais	Occasionnellement	Une fois par semaine	Plusieurs fois par semaine	Tous les jours

Personal Profile – U_SL_1

Education and qualifications: MSc in Library and Information Science

Job: PhD student

Age: 35-44

Which language(s) do you speak, read or write?
(List the language(s) and tick the relevant boxes)

	Native	Fluent	Fair	Basic
Danish _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
English _____	<input type="checkbox"/>	<input type="checkbox"/>	X	<input type="checkbox"/>
German _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?
(Tick one box only)

1	2	3	X	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?
(Tick one box only)

1	X	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on omputerized library catalogues (e.g. your library)?
(Tick one box only)

1	2	3	X	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)? (Tick one box only)

1	2	3	X	5
Never	Occasionally	Once a week	Several times a week	Daily

Personal Profile U_SL_2

Education and qualifications: MSc Information Management

Job: PhD Student

Age: 35-44

Which language(s) do you speak, read or write?
(List the language(s) and tick the relevant boxes)

	Native	Fluent	Fair	Basic
English _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
French _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
German _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?
(Tick one box only)

1	2	3	4	X
Never	Occasionally	Once a week	Several times a week	Daily

How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?
(Tick one box only)

X	2	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on omputerized library catalogues (e.g. your library)?
(Tick one box only)

1	X	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)? (Tick one box only)

1	2	X	4	5
Never	Occasionally	Once a week	Several times a week	Daily

Personal Profile U_SL_3

Education and qualifications: BA (Hons) & MSc

Job: PhD student

Age: 25-34

Which language(s) do you speak, read or write?
 (List the language(s) and tick the relevant boxes)

	Native	Fluent	Fair	Basic
English _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
French _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
German _____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	X
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?
 (Tick one box only)

1	2	3	4	X
Never	Occasionally	Once a week	Several times a week	Daily

How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?
 (Tick one box only)

X	2	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on omputerized library catalogues (e.g. your library)?
 (Tick one box only)

1	2	X	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)? (Tick one box only)

1	2	3	X	5
Never	Occasionally	Once a week	Several times a week	Daily

Personal Profile U_SL_4

Education and qualifications: Currently studying towards Meng in Software Engineering

Job: Student

Age: < 25

Which language(s) do you speak, read or write?
 (List the language(s) and tick the relevant boxes)

	Native	Fluent	Fair	Basic
English _____	X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
_____	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?
 (Tick one box only)

1	2	3	4	X
Never	Occasionally	Once a week	Several times a week	Daily

How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?
 (Tick one box only)

X	2	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on computerized library catalogues (e.g. your library)?
 (Tick one box only)

1	X	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)? (Tick one box only)

1	X	3	4	5
Never	Occasionally	Once a week	Several times a week	Daily

B.2 Satisfaction questionnaire compiled results

Satisfaction Questionnaire											Results per question and group of questions		
		French				English				Average	écart type	Average	
		User1(bb)	User2(sb)	User3(ga)	User4(bb)	User1	User2	User3	User4				
Overall reaction to the system	Question1	3	2	3	3	2	2	3	3	2,63	0,52	2,72	
	Question2(e)	2	3	4	2	1	1	4	2	2,38	1,19		
	Question3	2	2	2	3	1	2	3	2	2,13	0,64		
	Question4(i)	5	4	3	4	2	4	4	4	3,75	0,89		
Screen	Question1(i)	5	5	5	5	5	4	5	4	4,75	0,46	3,75	
	Question2(i)	2	4	3	2	3	4	1	4	2,88	1,13		
	Question3(i)	5	4	1	5	1	2	5	2	3,13	1,81		
	Question4(i)	4	5	5	4	2	4	3	3	3,5	1,04		
	Question5(i)	5	5	5	5	2	3	5	4	4,25	1,16		
	Question6(i)	5	4	4	4	2	3	4	4	3,75	0,89		
Terminology and system information	Question1(i)	5	4	5	5	3	3	5	4	4,25	0,89	2,84	
	Question2(i)	5	4	4	4	2	3	5	4	3,88	0,99		
	Question3(i)	1	4	2	2	3	3	1	2	2,25	1,04		
	Question4(i)	1	1	1	1	1	1	1	1	1	0		
Learning	Question1(i)	5	4	4	5	3	4	4	4	4,13	0,64	3,17	
	Question2(i)(e)	5	5	3	4	1	4	4	4	3,75	1,28		
	Question3(i)	1	1	1	1	1	3	1	4	1,63	1,19		
System capabilities	Question1(e)	1	1	2	2	2	3	1	2	1,75	0,71	2,5	
	Question2(e)	3	2	4	4	2	2	2	4	2,88	0,99		
	Question3(e)	1	2	3	3	4	4	4	4	3,13	1,13		
	Question4(i)	1	1	3	2	1	3	3	4	2,25	1,16		
Average per questionnaire		3,19	3,19	3,19	3,33	2,1	2,95	3,24	3,29				
Global Average (indice of satisfaction)												3,06	
Indice of Efficiency (i)												3,31	
Indice of Effectiveness (e)												2,78	

B.3 Interview compiled resultsInterview**L'interface / User Interface**

1. De façon générale que pensez-vous de l'interface de SPiRiT ? / What did you think about the SPiRiT user interface in general?

- (U_BB_1): Very basic
 (U_BB_2): Very sober
 (U_SB_1): Pretty and clear, excepted from the "where" part of the query interface : I did not get the difference between "town and city" and I was not sure about which fields were compulsory.
 (U_GA_1): Austere (too few colors). A clustering should be used for the display of the results.
 (U_SL_1): Some aspects were very simple - you enter a natural language query and there is only one selection box. But, you don't have a drop-down menu for region and town – this should be a required feature. If the user specifies country, the system could then automatically narrow down the list of regions and towns.
 (U_SL_2): It seemed OK. It was reasonably clear.
 (U_SL_3): Nice, clear, consistent and simple. My only complaint would be 'region' is too ambiguous.
 (U_SL_4): Seemed fairly straightforward. I could work out what was meant to go were. UK should have been split into England, Wales, Scotland, etc.

2. Quels sont les aspects de l'interface qui vous ont particulièrement plu ? / What aspects of the user interface did you particularly like?

- (U_BB_1): The logo
 (U_BB_2): The simplicity
 (U_SB_1): Its simplicity. The strip summing up the sent query.
 (U_GA_1): The query reformulation above the results.
 (U_SL_1): The simple layout.
 (U_SL_2): None.
 (U_SL_3): The fact it's uncluttered and all labelled and the fact it's clear and has been designed but not over-designed.
 (U_SL_4): There wasn't really anything particularly innovative. It seemed very similar to most search engines, which is not necessarily a bad thing.

3. Quels sont les aspects de l'interface qui vous ont particulièrement déplu ? / What aspects of the user interface did you particularly dislike?

- (U_BB_1): It does not adapt to the screen size
 (U_BB_2): It does not adapt to the screen size
 (U_SB_1): The lack of detail to describe the provided links.
 The impossibility to choose the number of displayed results.
 If you type 5 successive queries about Newport, you have to disambiguate the place each time.
 (U_GA_1): The lack of detail to describe the provided links
 The space between the result-lines is too big
 "Next 10" sign is placed too low in the result page (user has to scroll down to see it)
 The query reformulation's colors make it very difficult to read.
 (U_SL_1): The interface is too simplified – you cannot see the total number of search results or the criteria used for ranking.
 (U_SL_2): The error messages and the fact it wouldn't allow use of phrase operators, dates or numbers, and there was no indication that you could use them.
 (U_SL_3): Town and city are easy to define but region could mean county or country, etc.
 (U_SL_4): Again, nothing really.

4. Que pensez-vous de l'agencement général de l'interface ? / What did you think about the layout of the interface?

- (U_BB_1): Ordinary
 (U_BB_2): I like it
 (U_SB_1): Good
 (U_GA_1): Ordinary, not destabilizing

- (U_SL_1): The basic principle of 1 retrieve document per line is a good format. You cannot see if you are searching for single words or phrases.
- (U_SL_2): Reasonably clear. Needed more information under results titles. I didn't know if the document was repeated or if it was a different document with the same title.
- (U_SL_3): See above.
- (U_SL_4): It seemed straightforward. It seemed to flow reasonably well.

5. Que pensez-vous de la capacité du système à reconnaître les noms de lieux et sa capacité de désambiguer les lieux ayant le même nom ? / What did you think about the system's ability to recognise place names and the ability to disambiguate between places with the same name?

- (U_BB_1): It did not work. When I asked info about Newport in UK I got infos about Newport in the states.
- (U_BB_2): I did not notice it especially.
- (U_SB_1): Not bad. But it is not a new feature, other Web sites do the same. I fear this filter could be a bit dangerous (couldn't it filter too much documents ?) Anyway, with or without this functionality, the only point for users is to get good results.
- (U_GA_1): It worked fine, expecting some documents from the states... But it is impossible to ask everything from a search engine.
- (U_SL_1): Because you can't see how the engine works and can't specify, you can't disambiguate search results and spurious documents are sometimes retrieved. You could restrict results to certain domains, i.e. .co.uk, etc. I think the disambiguation feature should be automatic from a drop-down box if you have already selected the country.
- (U_SL_2): I wasn't impressed when it didn't find 'Renfrew' (a place in Scotland). It seems it can't find places as a separate entity. It wouldn't accept places as the search subject rather than the city or region. But in this case I was interested in the 'Renfrew ferry' and not the actual place.
- (U_SL_3): Good idea. Depends on how precise you're doing your query.
- (U_SL_4): The disambiguation feature was very useful, because you often get places with the same name.

Les résultats de la recherche / Search Results

6. Que pensez-vous des résultats trouvés par SPiRiT ? / What did you think about the results returned by SPiRiT?

- (U_BB_1): Similar to those found with other search engines.
- (U_BB_2): Very good results on the scenario proposed, but for Cardiff (no maps) and for Newport (I got some results on the US)
- (U_SB_1): Not bad but not phenomenal.
- (U_GA_1): Not bad.
- (U_SL_1): It doesn't seem that popularity is taken into account, i.e. the number of links to that page elsewhere on the web. The quality of the sources is questionable. There are too many duplicated. You should make it so the index page only is retrieved and not multiple pages from the same source.
- (U_SL_2): It seemed OK.
- (U_SL_3): Good. It's hard when doing testing because you're trying to search like you would normally but can't go any deeper into the resulting web page than the first page you get back.
- (U_SL_4): I didn't think they were great. There were some good result hits, but some seemed to be quite ambiguous and only slightly related to the topic.

7. Que pensez-vous de la pertinence des résultats en ce qui concerne le thème et la localisation ? / What did you think about the relevance of the results returned in terms of subject and location searched for?

- (U_BB_1): Ordinary
- (U_BB_2): Nice (but for some US documents when I asked for "top ten gardens, UK", "outdoors, London, UK", "tourist information, Newport, UK")
- (U_SB_1): Better for the place than for the theme.
- (U_GA_1): Some results were irrelevant but the average is correct.
- (U_SL_1): Precision was very low – there were some spurious results, particularly for the Glasgow query. If the system is a geo search engine you don't expect this.
- (U_SL_2): Most of the documents on festivals were out of date.
- (U_SL_3): All OK apart from Glasgow which came up with links in the USA. Just typing in Cardiff seemed to bring in irrelevant results, but it could have been my searching. On the whole it was good, you never expect results to be perfect.
- (U_SL_4): I didn't feel they tied to a location well. The results weren't necessarily relevant to a location, they might have just mentioned it in a different context.

8. Que pensez-vous de la façon dont les résultats sont présentés ? / What did you think about the way search results were presented?

- (U_BB_1): Very sober.
 (U_BB_2): Not much spatial.
 (U_SB_1): The disposition is OK.
 (U_GA_1): The space between lines is too big.
 (U_SL_1): I like it. You need more information about ranking and total number of results and you should have grouping of documents from the same source but in sub-hierarchies.
 (U_SL_2): Apart from the lack of sub-titles it was OK.
 (U_SL_3): Fine. I guess you could present like that or like Google where it takes a bit of the page, but I didn't click on a title thinking it was relevant only to find it wasn't.
 (U_SL_4): They should have given a bit more information, rather than just the title, which could be rather misleading.

9. Que pensez-vous des informations fournies pour chaque résultat trouvé ? / What did you think about the information accompanying each search result?

- (U_BB_1): I would have expected some spatial relevance weights to describe the answers instead of the URL title. On the documents retrieved, the highlighting of the searched words is very useful.
 (U_BB_2): There is not much information, you have to click on a link to get an overview.
 (U_SB_1): Not enough details
 (U_GA_1): Not sufficient. Should provide the web address of the site and also snippets (like Google). A clustering of the result would be useful.
 (U_SL_1): See above.
 (U_SL_2): See above.
 (U_SL_3): It was fine.
 (U_SL_4): See above. I would maybe include passages of text from inside the result.

10. Que pensez-vous du classement des résultats ("ranking") ? / What did you think about the ranking of displayed search results?

- (U_BB_1): Not very spatial.
 (U_BB_2): Correct, I never had to go to the 2nd page.
 (U_SB_1): I did not understand the criteria. It would be good to have an indication. The first results was not always relevant. Anyway, it is hard to judge because of the lack of information describing links.
 (U_GA_1): Not very spatial.
 (U_SL_1): Since no information was given about the ranking functions, I can't comment on this.
 (U_SL_2): I didn't really go through enough screens to give it a fair test. Generally, there were a fair few relevant hits on the first screen so rather than going on, I amended my search to look at different aspects, so I cannot fairly comment on this.
 (U_SL_3): I wasn't aware that they were ranked. They've obviously got numbers next to them. For the first 30 results, it often wasn't clear why things were being ranked the way they were, so documents appeared less relevant.
 (U_SL_4): I wasn't aware they were ranked.

L'utilisation de SPiRiT / Using SPiRiT

11. Avez-vous rencontré des difficultés lors de votre utilisation ? / Did you experience any difficulties when using SPiRiT?

- (U_BB_1): Some bugs message when I ill typed a word (breafeast instead of breakfast).
 When there is such a bug and you click on the return button then you go back to the homepage instead of getting your query back (so you have to type everything again)
 When I wanted to change my query. i.e. I send a query and then think of modifying so I change it and send it again but the system will first handle the initial query anyway instead of cancelling it automatically if I have sent a new one from the same page. I must then use the mozilla "stop" button and then type my new query and send it.
 (U_BB_2): Some bugs : I got US documents sometimes whereas my query was on UK
 I don't know place names in UK (region and so on) so I would like the system to help me and put the region name himself when I have written a city Name

- I don't like going to the homepage again, couldn't each list of answers be displayed in a new page?
- (U_SB_1):
 No heavy difficulty
 Difficulties to find Newport : it is very restrictive and I did not know exactly where was the Newport I was looking for.
 No place name was proposed in case of misspelling
 I would have liked to submit several place names for the same query
 I tried to do an "advanced" search ("or", "and", "and not" operators in the theme field) but I had errors.
- (U_GA_1): No, except the fact that no correction is provided when misspelling.
 (U_SL_1): See my previous comment about not being able to specify what field or domain.
 (U_SL_2): Yes. I received errors when trying to use operators. It was very restricted.
 (U_SL_3): Only technical difficulties but not searching difficulties.
 (U_SL_4): None whatsoever.

12. Que pensez-vous de la quantité d'aide et de support fournie par le système pendant que vous effectuez une tâche ? / What did you think about the amount of support the system gave you when completing your tasks?

- (U_BB_1): Zero help.
 (U_BB_2): I did not feel I was helped.
 (U_SB_1): We do not know anything. There is a serious lack of explanation.
 (Only the errors help a bit to understand). By chance, this is a simple system and the lack of explanation does not prevent from using it.
- (U_GA_1): Zero help provided.
 (U_SL_1): You should display the ranking method, percentage relevance, total results, etc.
 (U_SL_2): The only time support was needed was when error messages were produced and they weren't meaningful.
 (U_SL_3): I wasn't aware of any.
 (U_SL_4): Didn't really notice any support, it just seemed pretty straightforward.

Conclusion / Conclusion

13. Quelle est votre impression générale sur SPiRiT ? / How do you feel about the SPiRiT system in general?

- (U_BB_1): Similar to google.
 (U_BB_2): Nice but could give much more assistance with the spatial stuff.
 (U_SB_1): I'm doubtful.
 (U_GA_1): Do not provide more features than other search engine such as Google, but very interested by future version (if graphical features are provided).
 (U_SL_1): The problem is the quality of the sources of results was simply too low and there were spurious results about location. You cannot specify phrases and search fields, i.e. title or whole document, in the results.
 (U_SL_2): Pretty neutral. Not sure if it would have done anything Google couldn't do.
 (U_SL_3): Fine. I guess because it's only got a bit of the Internet cached. In terms of integrating that though it seemed good, but as it gets bigger it would maybe become less accurate.
 (U_SL_4): It just seemed like another search engine.

14. Quels aspects du système avez-vous particulièrement apprécié ? / What aspects of the system did you particularly like?

- (U_BB_1): Simple to use.
 The fields are simple.
 (U_BB_2): It is simple and easy to learn.
 (U_SB_1): The place recognition and disambiguation. I would have liked some maps.
 (U_GA_1): Geographical aspect, even if it was difficult to appreciate with the scenarios that are too guiding and let not the user free enough.
 (U_SL_1): Disambiguation of place names with a drop-down menu.
 (U_SL_2): None.
 (U_SL_3): Interface was obviously good for geographical related information. I like the fact you can specify areas.
 (U_SL_4): None really.

15. Quels aspects du système vous ont particulièrement déplu ? / What aspects of the system did you particularly dislike?

- (U_BB_1): It is slow.
There were some ill understanding of the query (newport).
- (U_BB_2): I had to know city or region names in UK.
- (U_SB_1): It is slow.
Not enough description of the provided links.
Not enough search options
- (U_GA_1): No maps.
Not enough user-friendly.
- (U_SL_1): You should be able to see the results spatially on a map which you can search on different levels.
- (U_SL_2): The error reporting.
- (U_SL_3): The fact it was slow.
- (U_SL_4): None.

16. Y-a-t-il un commentaire que vous souhaiteriez ajouter à propos de votre expérience avec SPiRiT ? / Are there any other comments you would like to make about your experience with the SPiRiT system?

- (U_BB_1): Concerning the evaluation methodology : it is difficult as a user to evaluate the quality of the spatial search on an area I don't know.
There is no map.
The scenarios are too rigid and oriented. I was implicitly given the words to type on the interface and felt I had not the possibility to really evaluate by myself the system.
- (U_BB_2): I think a map is important to express the location I am looking for.
- (U_SB_1): I don't see what the system adds to Google.
- (U_GA_1): The scenarios are too rigid and thus it is difficult to talk about satisfaction.
- (U_SL_1): The system should be at a point where pictures are displayed before I could comment.
- (U_SL_2): No. I think it should be linked to a graphical display if possible.
- (U_SL_3): No.
- (U_SL_4): It seemed alright. It was usable.

Moderator notes

(U_BB_1) Moderator Notes :

Task1 : 10mn.

Task2 : 5mn

Task3 : 5mn

Task4 : stop after 15mn

The user often writes a query map (in) London. i.e he does not use the spatial relationship "in".

He often queries sites writing "top ten..."

(U_SB_1) Moderator Notes :

Task1 : 12mn

Task2 : 6mn

Task3 : 14mn

Task4 : 14mn

The user clicked on the link "about spirit" but there is no information.

He tried some complex searches (with "-" operator)

He had the feeling, for each scenario, that the retrieved documents were always the same.

Isolated error when typing "Glasgow, UK" (place not found) Later, it works.

The user says Google is better because the links are better presented. He says several times "I do not see what it adds to Google"

Sometimes two links have the same title. Some have the title "no title" or "directory". Some links go to the same Web page.

He tries the query "hotel Cardiff" on Google and finds it works very well.

The Newport disambiguation : the user has some difficulties. He would have liked some indications helping him to choose the right Newport. His reflex is to search for information about different Newport on Google.

He tries a fuzzy region in the "region" field ("region of Cardiff").

About the methodology:

The user tried to navigate out of the collection.

For the first scenario he was not sure about what he had to search for.

Some difficulties about the satisfaction questionnaire (questions about the consistency of terms and about the information disposition)

About various users needs (Quis question D.4): no advanced search option

About useless error messages (Quis question C.4): there are two cases of error message:

A computation error (no help)

A "stupid" error (e.g. Glasgow not found)

About spatial relationships : "In" is a good one but some other would be useful for "advanced" users.

(U_GA_1) Moderator Notes :

Task1 : 13mn.

Task2 : 12mn

Task3 : 14mn

Task4 : 9mn

The user is experiencing many difficulties because he performs as an ISJ assessor would and thus he is frustrated by the dead links. He also often goes out of the collection.

In case of misspelling, the user is confused because he doesn't think to check first for this mistake and get an java error.

The user tried Welsh Castles instead of Castles IN Walles.

To understand the meaning of "ceilidh", the user used Google and not Spirit.

The user is used to quering with only singular forms (e.g. Hotel and not Hotels) because he thinks he will get both results from "hotel" and "hotels"

(U_SL_1) Moderator Notes :

U_SL_1:

- Tried phrase searching with "" --> error message.
- Asked if system accepted truncated queries.
- Complained that they couldn't see how many documents had been retrieved.
- Complained that results were being repeated in the top ten.
- Tried to follow links within a result to judge relevance.
- System crashed.
- Complained that they received results for Cambridge & Aberdeen despite searching for Glasgow.
- Dead links.
- Disambiguation screen took a long time to load.
- Complained that yellow highlighting of found search terms was too bright.

U_SL_2:

- Use of capital letters in query.
- Tried phrase searching with "" --> error message.
- Query too long --> error message.
- "Descriptions need to be better. Same documents with different titles keep coming up"
- Tried to use a '-' symbol --> error message.
- Place name ('Renfrew') in subject box --> error message (because it didn't occur in any doc's?).

U_SL_3:

- No region inputted --> error message.
- "Is Wales a region?!"
- Browser crashed when clicking back button on one of the results.

U_SL_4:

Nothing to report.