British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Semantic Information Retrieval for Geoscience Resources: results and analysis of an online questionnaire of current web search experiences

Informatics Programme

Internal Report OR/16/047

BRITISH GEOLOGICAL SURVEY

INFORMATICS PROGRAMME

INTERNAL REPORT OR/16/047

Semantic Information Retrieval for Geoscience Resources: results and analysis of an online questionnaire of current web search experiences

I Nkisi-Orji

Contributor/editor

R Heaven

Keywords

Information retrieval; Ontology; Semantic; Search engine; Questionnaire.

Front cover

Bibliographical reference

NKISI-ORJI, I, HEAVEN, R. 2016. Semantic Information Retrieval for Geoscience Resources: results and analysis of an online questionnaire of current web search experiences. *British Geological Survey Internal Report*, OR/16/047. 22pp.

Keyworth, Nottingham British Geological Survey 2016

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

BGS Central Enquiries Desk

Tel 0115 936 3143 Fax 0115 936 3276

email enquiries@bgs.ac.uk

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3241 Fax 0115 936 3488

email sales@bgs.ac.uk

The Lyell Centre, Research Avenue South, Edinburgh EH14 4AP

Tel 0131 667 1000 Fax 0131 668 2683

email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090 Fax 020 7584 8270

Tel 020 7942 5344/45 email bgslondon@bgs.ac.uk

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff CF15 7NE

Tel 029 2052 1962 Fax 029 2052 1963

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800 Fax 01491 692345

Geological Survey of Northern Ireland, Department of Enterprise, Trade & Investment, Dundonald House, Upper Newtownards Road, Ballymiscaw, Belfast, BT4 3SB

Tel 028 9038 8462 Fax 028 9038 8461

www.bgs.ac.uk/gsni/

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500 Fax 01793 411501

www.nerc.ac.uk

Website www.bgs.ac.uk

Shop online at www.geologyshop.com

Foreword

This report presents the results and analysis of the questionnaire "*Semantic web searches for geoscience resources*" completed by staff of British Geological Survey (BGS). The questionnaire was designed to better understand current web search habits, preferences, and the reception of semantic search features. It will be used primarily to understand user requirements and inform research direction for the PhD project "Semantic Information Retrieval using Domain Ontologies". The PhD research is based at Robert Gordon University (School of Computing Science and Digital Media), and is jointly funded by BGS.

Acknowledgements

The authors would like to thank the BGS staff that volunteered their time to complete the questionnaire.

We would like to acknowledge the help of the PhD lead supervisor, Dr Nirmalie Wiratunga, and the rest of the supervisory team at Robert Gordon University - Dr Stewart Massie and Dr Kit Ying-Hui - for their help in designing the questions.

SurveyMonkey software [1] was used to create the survey, collect the responses and analyse the results.

Ikechukwu Nkisi-Orji collated the results, created the figures and wrote the results and analysis sections in this report. Rachel Heaven helped design the questionnaire and edited the text into this BGS standard format, writing the summary and acknowledgements.

Contents

Figures

Summary

An online questionnaire "*Semantic web searches for geoscience resources*" was completed by 35 staff of British Geological Survey (BGS) between 28th July 2015 and 28th August 2015. The questionnaire was designed to better understand current web search habits, preferences, and the reception of semantic search features in order to inform PhD research into the use of domain ontologies for semantic information retrieval.

The key findings were that relevance ranking is important in focussed searches that seeks the answer to a specific question, because 50% of people only look at the first 10 results. Relevance ranking is important but not so critical for broad reaching literature and data gathering searches because 88% of respondents would typically assess more than 10 results in this case. A large majority of respondents usually or sometimes had to perform multiple searches or construct advanced searches in order to include all relevant variations in terminology, and an optional feature in the search engine that expanded the search terms for them would be beneficial and desirable. All respondents reported that their search results were at some point, dominated by irrelevant result entries. Asked how a feature that disambiguates terms in search queries should be implemented, 81% will like to be able to specify intended context/meaning of search terms but only when such terms are ambiguous.

The conclusion was that the collected responses, though a small sample, indicated vast support for the implementation of semantic search features to add narrower or equivalent terms to original search intent and to specify the context/meaning of ambiguous search terms, but the respondents preferred to be in control of whether or not those features were implemented on each search.

1 Introduction

This report is on findings from the questionnaire "Semantic web searches for geoscience resources" completed by staff of British Geological Survey (BGS). The questionnaire was designed to better understand current web search habits, preferences, and the reception of semantic search features.

1.1 RESPONSE STATISTICS

Responses to this opt-in survey were collected online between 28 July 2015 and 28 August 2015. Thirty five responses were received with 22 completed (that is, 13 respondents had one or more questions left unanswered). On average, 25 responses were received for each question.

1.2 QUESTIONNAIRE STRUCTURE

Questions are numbers Q1, Q2, etc. Two types of search activity were defined during design of the questionnaire, so a number of questions were repeated with respect to each type of search activity.

Section 1 of the questionnaire (Q1-4) relate to web searches where one wants to find a comprehensive list of all relevant results (e.g. literature search, data gathering), so completeness of results is the most important measure of success.

Section 2 of the questionnaire (Q5-8) relate to web searches where one is looking for the answer to a specific question, i.e. the relevance ranking of the results is the most important measure of success.

Section 3 (Q9-16) posed questions to assess respondents' reception of semantic search features and their preferences in the implementation of such features.

Each question with results and analysis of each question is described in the relevant section below.

2 Analysis of responses to questions

2.1 SECTION 1: LITERATURE OR DATA GATHERING SEARCHES

Questions in this section (Q1-4) relate to web searches where one wants to find a comprehensive list of all relevant results (e.g. literature search, data gathering), so completeness of results is the most important measure of success.

2.1.1 Q1

Q1: For this first sort of search, which search applications are useful to you?

- ❑ Popular search engine (e.g. Google, Bing, Yahoo)
- ❑ Publication citations (e.g. Google Scholar, Science Direct)
- ❑ Cross discipline data portal (e.g. data.gov, INSPIRE geoportal, Scottish SDI)
- ❑ Earth Sciences catalogue (e.g. NERC Data Catalogue, NERC library, NORA)
- ❑ Discipline/community specific catalogue (e.g. MEDIN for marine data, ESDAC for soil data etc.)
- ❑ BGS intranet tools (dtSearch for text resources, discovery metadata)
- ❑ Other (please specify)

2.1.1.1 RESULTS AND ANALYSIS

Fifty nine percent (59%) of respondents chose publication citations (e.g. Google Scholar and Science Direct) as the most useful application for this sort of search, see Figure 1.

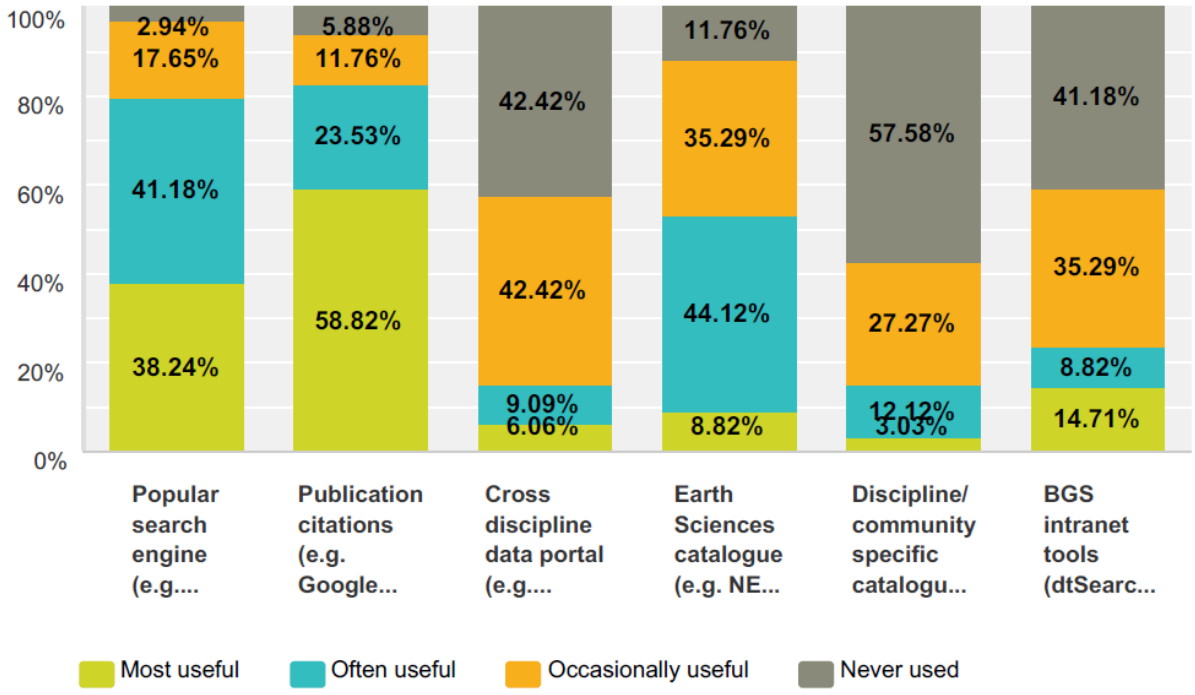

Figure 1 Search engine preferences for literature searches

Eighty two percent (82%) found publication citations most or often useful for literature and data gathering search. This is expected because these are repositories for journal articles and other scholarly works. The popularity of publication citations is closely matched by popular search engines (79% most or often useful) for this sort of search. It is not unusual to begin literature search using popular search engines since they often look in publication citations as well. For example, Google search often displays results from Google Scholar. Earth science catalogues are fairly popular as 53% found them most or often useful.

Over 57% never used discipline specific catalogues and no search application was reported as most useful by 26% of respondents. Wikipedia, BGS Library Catalogue and Web of Science were identified as other useful search tools.

2.1.2 Q2

Q2: For this first sort of search, how often are you satisfied with the results after:

- Using a small number (<5) of words in a free text search
- Using a large number (>5) words in a free text search
- Using logical operators in a free text search (AND/NOT/OR etc.)
- Using advanced search features to search within specific metadata fields (keywords, title, author etc.)

2.1.2.1 RESULTS AND ANALYSIS

Fifty percent (50%) of respondents were usually satisfied with search results of short queries (that is, when less than 5 words were used in search).

Usual satisfaction with search results dropped to 44% for long queries (that is, above 5 words) as seen in Figure 2. Often, search applications only return documents which contain most of the search terms. Hence, fewer search terms will increase the number of hits. Since search intent is to gather a comprehensive list of relevant results, more results may translate to greater satisfaction. A likely alternative explanation is that the use of long queries indicates search

instances where users find it difficult to properly express their information needs. Search results will not be satisfactory if the search terms used fail to appropriately convey an information need.

Thirty six percent (36%) were usually satisfied with results of advanced search features (that is, search within specific metadata fields); 47% had never used logical operators and results from their use were the least satisfactory. The use of logical operators in search requires additional skills which may seem excessive to many users.

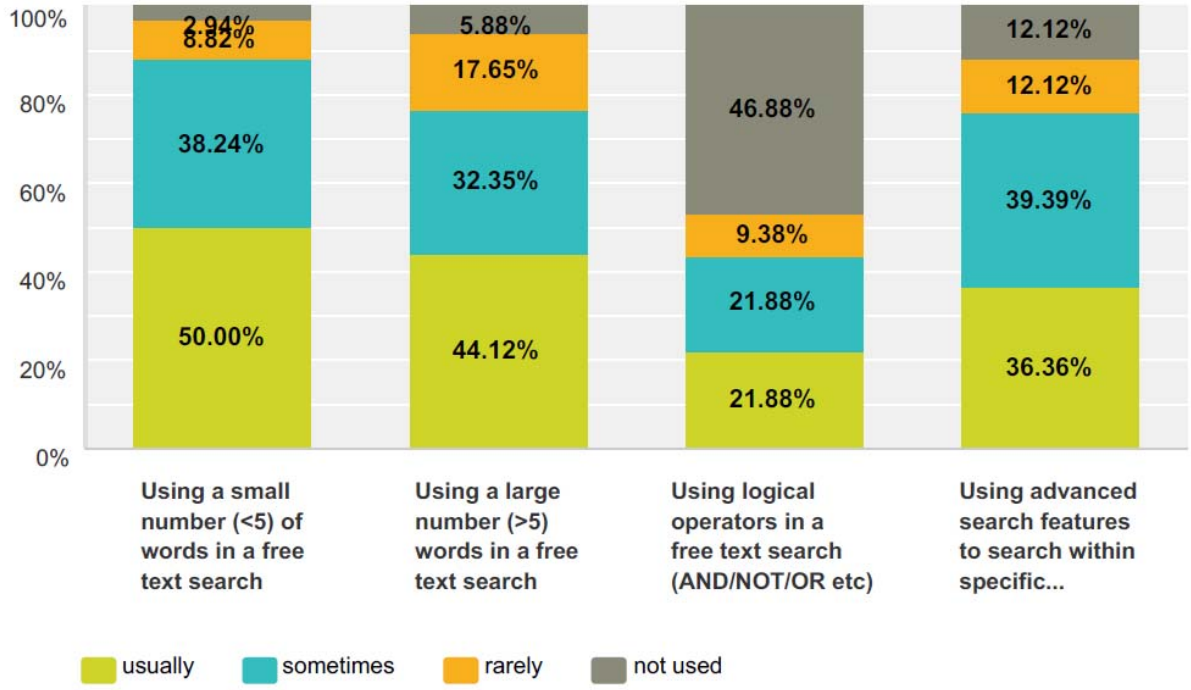

Figure 2 Satisfaction with search results according to query length or search engine features used

2.1.3 Q3

Q3: For this first sort of search, what is the maximum number of search results you are willing to assess, rather than refining your search criteria or changing the search engine?

❑ 1-10 ❑ 10-20 ❑ 20-50 ❑ 50+

Eighty eight percent (88%) of respondents will assess more than the first 10 search results.

As shown in Figure 3, majority of respondents (59%) will assess 10 to 20 search results. With default settings, this is the second search result page of popular search applications (e.g. Google, Bing, and Yahoo). Since search intent is to identify multiple relevant documents, there is increased tendency to assess more than the first few results. A significant 24% will go on to assess between 20 and 50 search result entries.

2.1.3.1 RESULTS AND ANALYSIS

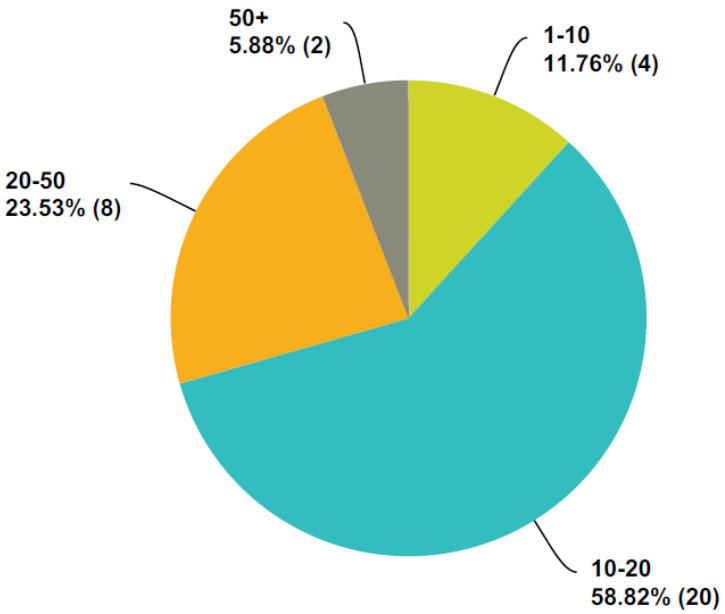

Figure 3 Number of search results assessed

2.1.4 Q4

Q4: For this first sort of search, could you give a few examples of some recent searches you conducted, and any comments on the relevance of results returned?

Forty two (42) sample queries were collected which showed a mixture of navigational and informational queries.

Highlighted sample queries in Figure 4 (e.g. #9: Garcia et al 2006 Tetrapods) are instances where search intent seem navigational. Navigational queries are queries whose underlying information needs are specific web sites or publications which are known (or assumed) to exist. On the other hand, the intent of informational queries is to find the best documents that meet an information need. Our interest is in informational queries which are more pertinent to information retrieval (IR) research. Query #7 (shale gas mechanics) is an interesting example as the respondent noted that search results from a different domain influenced search precision.

Most of the sample queries are short queries (<5 words) with an average query length of 2.8 words. Query #36 in Figure 4 shows an interesting search strategy where search results are limited to documents which contain exact phrases in quotes. This is an implicit use of the AND logical operator and can help to filter off irrelevant documents.

2.1.4.1 RESULTS AND ANALYSIS

#	Query with comments
1	Evaluation of Unconventional Natural Gas Prospects > got the paper I was looking for in 2 clics
2	gutter casts crevasse splay
3	Extreme natural hazards - better on publication citation sites
4	Groundwater flooding - ok in google scholar
5	borehole data
6	"ice streams tweed" Author: Everest (in Scholar)
7	shale gas mechanics, a lot of the papers are associated with social science and it is ahrd to filter these out
8	hampshire author:melville very relevant
9	Garcia et al 2006 Tetrapods
10	Lawson, J.D. & Straw, S.H. 1956. The Ludlovian rocks of the Welsh Borderland. Advanced Science. 12 pp563-570.
11	tsunami
12	sinkhole gower subsidence
13	porosity chalk
14	Characterization of inclusions in Emeralds from Afghanistan
15	st vincent citizen science
16	BGS BRK (result was not actual answer but close enough)
17	Hill Lombardi 2002 shale > got the paper I was looking for in 2 clics
18	synaeresis cracks
19	Applecross - better on specific catalogue (ie. NERC)
20	Radon as groundwater tracer - ok in google scholar
21	geological maps
22	sedimentary palaeozoic (in Science Direct)
23	streaming potential porous media relevant
24	Tsunamis of Volcanic Origin: Summary of causes
25	Advmt. Sci, Br. Ass
26	deposit
27	cromford sough
28	correlation length chalk
29	volcano crowd sourcing
30	BGS Lexicon LC (result was correct)
31	william smith cross sections
32	Recent volcanic eruptions in the afar rift - Oublication citations as i knew it was a paper.
33	Groundwater abstraction statistics uk - proved really difficult
34	geophysical logs
35	"Gossan" in Wikipedia
36	"barton on sea" landslide very relevant
37	paleosols in clastic sedimentary rocks
38	The british association for the advancement of science
39	North Sea
40	gunn et al peak hydrology
41	porosity chalk UK
42	new zealand disaster management

Figure 4 Examples of search queries with comments. Highlighted entries are where specific content are being sought

2.2 SECTION 2: SEARCHES THAT ASK A SPECIFIC QUESTION

Questions in this section (Q5-8) relate to web searches where one is looking for the answer to a specific question, i.e. the relevance ranking of the results is the most important measure of success.

2.2.1 Q5

Q5: For this second sort of search, which search applications are useful to you?

- ❑ Popular search engine (e.g. Google, Bing, Yahoo)
- ❑ Publication citations (e.g. Google Scholar, Science Direct)
- ❑ Cross discipline data portal (e.g. data.gov, INSPIRE geoportal, Scottish SDI)
- ❑ Earth Sciences catalogue (e.g. NERC Data Catalogue, NERC library, NORA)
- ❑ Discipline/community specific catalogue (e.g. MEDIN for marine data, ESDAC for soil data etc.)
- ❑ BGS intranet tools (dtSearch for text resources, discovery metadata)
- ❑ Other (please specify)

2.2.1.1 RESULTS AND ANALYSIS

Seventy nine percent (79%) of respondents chose popular search engines (e.g. Google, Bing and Yahoo) as their most useful application for this sort of search as shown in Figure 5.

Ninety five percent (95%) said that popular search engines were most or often useful for this sort of search. The next popular search applications are publication citations (e.g. Google Scholar) which are most useful to about 33% of respondents. Earth Sciences catalogue and BGS intranet tools were occasionally useful to a significant proportion of respondents (47% and 35% respectively). Cross discipline data portals (e.g. data.gov) and Discipline/community specific catalogues (e.g. MEDIN for marine data) were never used by majority of respondents. Wikipedia, GSW, GDI and COPAC are other useful search applications that were identified by respondents.

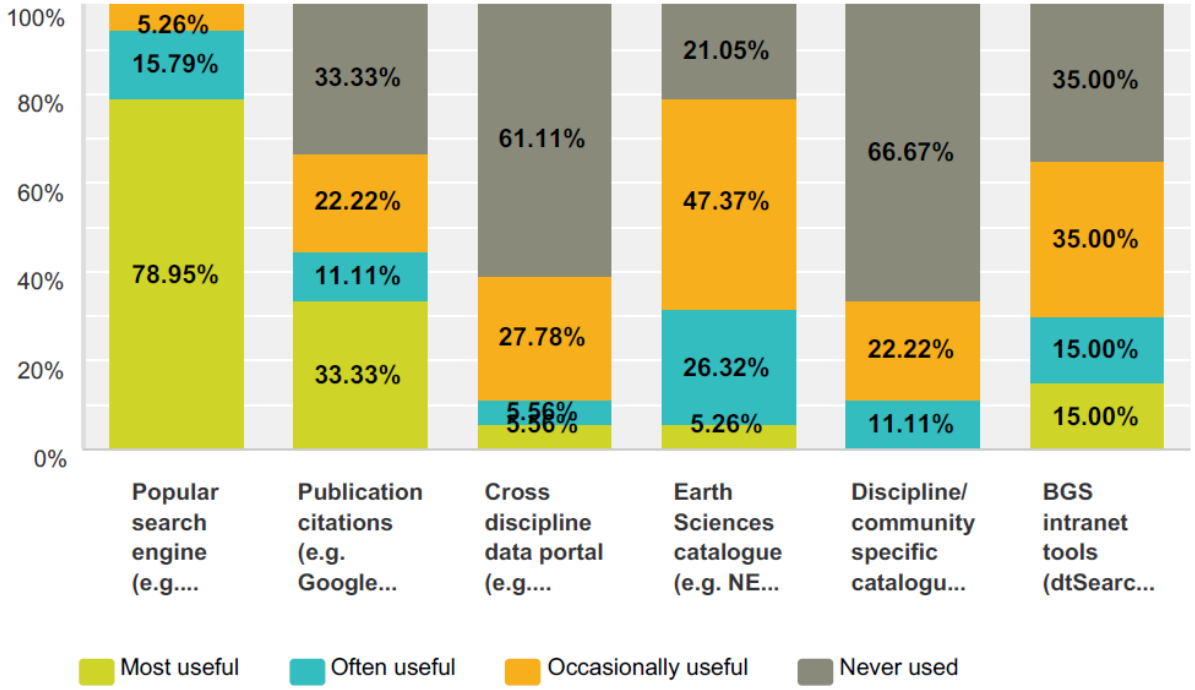

Figure 5 Search engine preferences to ask specific questions

2.2.2 Q6

Q6: For this second sort of search, how often are you satisfied with the results after:

- Using a small number (<5) of words in a free text search
- Using a large number (>5) words in a free text search
- Using logical operators in a free text search (AND/NOT/OR etc.)
- Using advanced search features to search within specific metadata fields (keywords, title, author etc.)

2.2.2.1 RESULTS AND ANALYSIS

Sixty one percent (61%) of respondents were usually satisfied with search results when more than 5 words were used in search query.

As seen in Figure 6, there was no clear distinction for search result satisfaction between short queries (< 5 words) and long queries (> 5 words). Sixty one percent (61%) were usually satisfied with results of long queries and is slightly higher than 58% for usual satisfaction with results of short queries. However, the proportion of respondents who were sometimes satisfied was higher for short queries. Similar to the previous type of search (see Section 1, Q2), the use of logical operators produced the least satisfaction among respondents who use it.

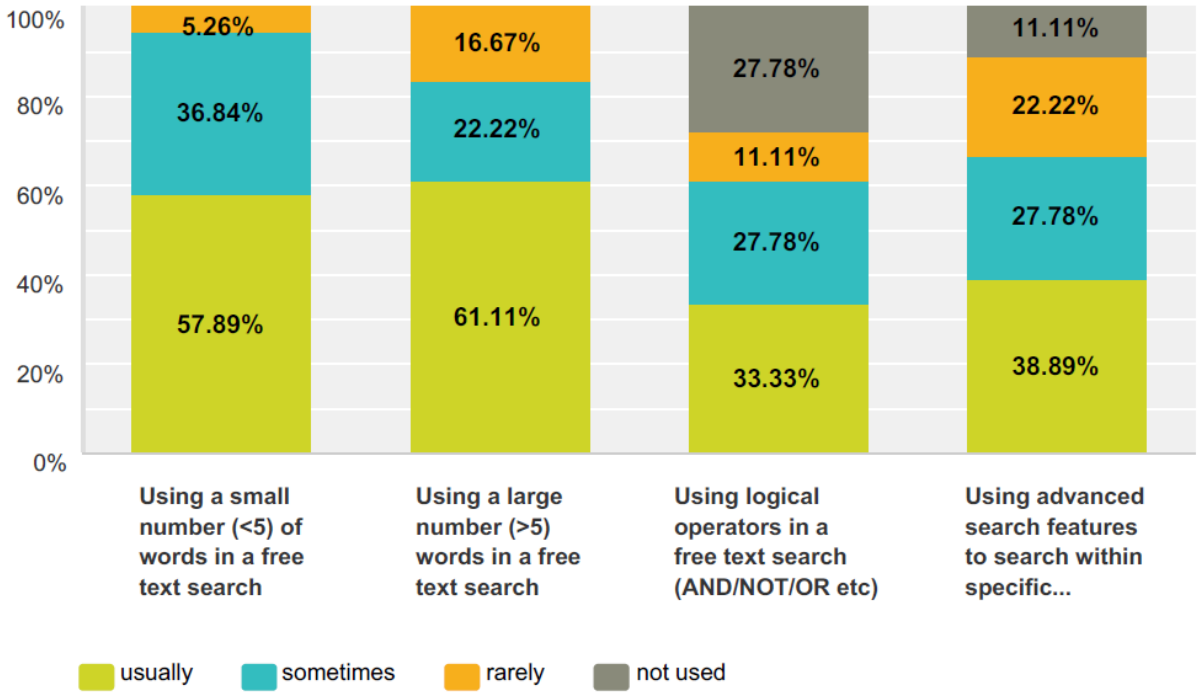

Figure 6 Satisfaction with search results according to query length or search engine features used

2.2.3 Q7

Q7: For this second sort of search, what is the maximum number of search results you are willing to assess, rather than refining your search criteria or changing the search engine?

❑ 1-10 ❑ 10-20 ❑ 20-50 ❑ 50+

2.2.3.1 Results and analysis

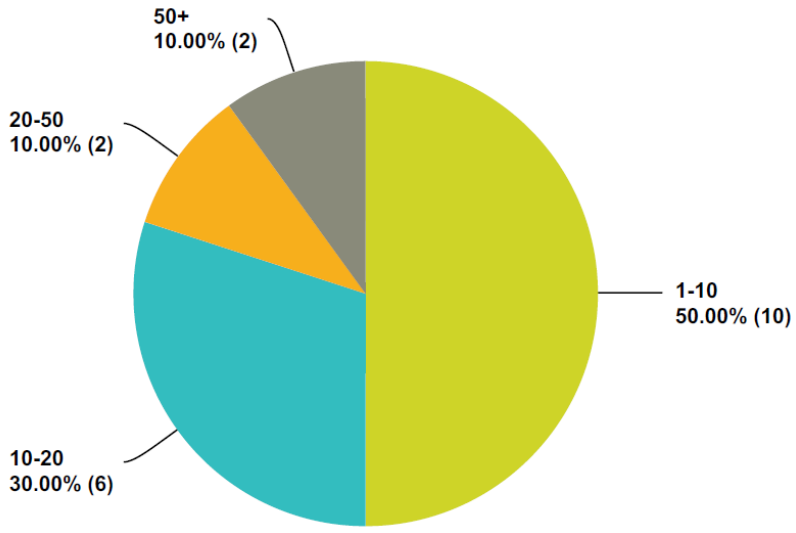

Figure 7 Number of search results assessed

Figure 7 shows that 50% of respondents assess the first 10 search results only. Eighty percent (80%) will not assess more than 20 search results.

Fifty percent (50%) will not assess more than 10 search result entries before attempting other search strategies in order to obtain better results. Thirty percent (30%) assesses 10 to 20 search results while 20% assesses more than 20 results. Compared to the previous sort of search (see

Section 1, Q3), the number of search results which respondents assess is less for this sort of search. Respondents are either unwilling to assess many search results or have no need to assess many search results. The latter is a possible explanation since a relevant entry among the first few search results can make assessing additional results unnecessary.

2.2.4 Q8

Q8: For this second sort of search, could you give examples of some recent searches you conducted, and any comments on the relevance of results returned?

2.2.4.1 Results and analysis

Fourteen sample queries were collected with associated comments.

Query #11 (deposit) in Figure 8 is an example where the search term have popular alternative senses. When using popular search engines (e.g. Google), results that describe the financial sense of “deposit” are expected to be popular in search results thereby reducing search precision (if the geological sense was intended). Such interference is expected to be less pronounced when searching in domain-specific document collections. The average query length in collected examples is 3.5 words per query.

#	Query with comments
1	Pavlek harness pelvis formation > got results dealing either with the P harness or the pelvis
2	bgs bussels no 7 - google takes me straight to relevant bgs webpage, but i knew it existed
3	usually copy full title
4	boge sobi borehole classification very relevant
5	difficulties identifying bouma sequences in core
6	tsunami
7	BGS GSI3D FME
8	3D geology modelling methods and software
9	American Petroleum Institute one page summary - Google - didn't find what i wanted!
10	geological time scale very relevant
11	deposit
12	3D databases at geological national state survey
13	Leaflet js poerformance - google search gave me reasonable results
14	North Sea

Figure 8 Examples of recent searches to ask specific questions

2.3 SECTION 3: SEMANTIC SEARCH FEATURES

Questions in this section assess respondents’ reception of semantic search features and how they want such features to be implemented.

2.3.1 Q9

Q9: How often do you have to perform multiple searches or construct an advanced search query in order to also search all the narrower/child terms of your original search intent?

❑ always ❑ usually ❑ sometimes ❑ seldom ❑ never

2.3.1.1 Results and analysis

Although no one performed searches to include narrower/child terms to search intent all the time, 64% of respondents usually or sometimes used this search strategy.

Narrower/child terms are more specific terms used to describe a concept. For example, when search intent is the "Longmyndian Supergroup" one may also use the narrower "Wentnor Group" in search terms which is a more specific rock unit. This way, relevant documents which did not mention "Longmyndian Supergroup" are also discovered. Twenty seven percent (27%) of

respondents include narrower terms to search intent sometimes and 36% usually does this. Fourteen percent (14%) have never attempted to include narrower terms to original search intent.

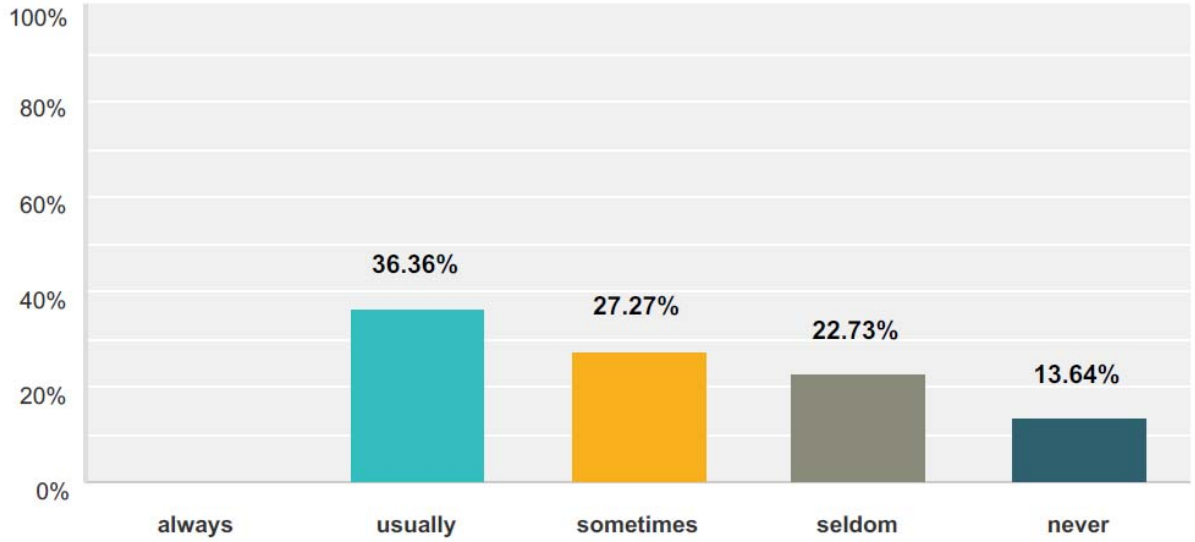

Figure 9 Tendency to perform multiple searches or use advanced search features to include narrower terms to original search intent

2.3.2 Q10

Q10: How often do you have to perform multiple searches or construct an advanced search query in order to include all the equivalent terms or alternative spellings of your original search intent?

❑ always ❑ usually ❑ sometimes ❑ seldom ❑ never

2.3.2.1 Results and analysis

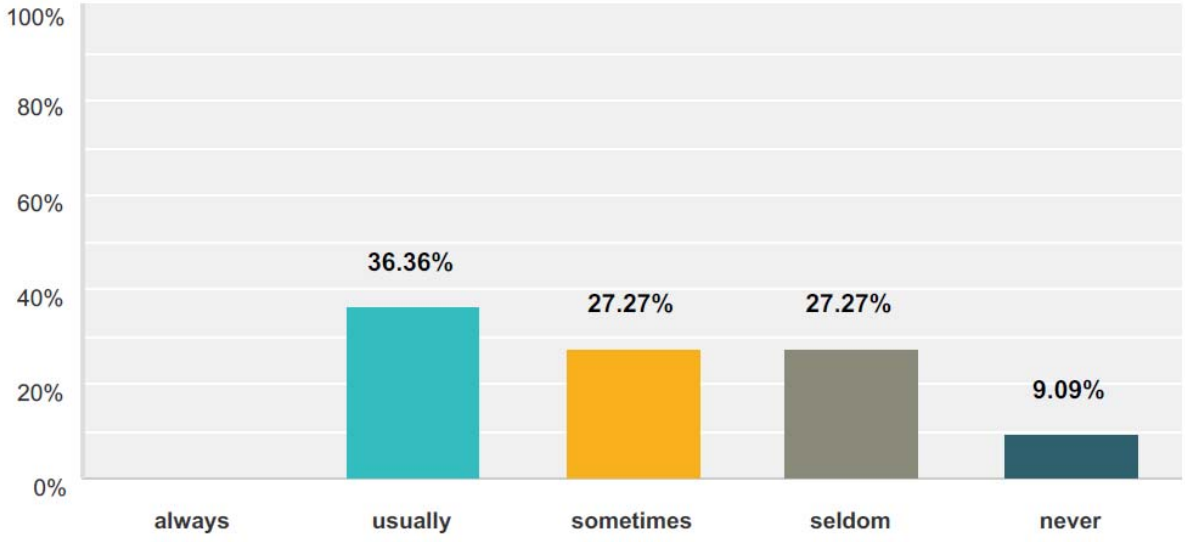

Figure 10 Tendency to perform multiple searches or use advanced search features to include equivalent terms to original search intent

Sixty four percent (64%) of respondents usually or sometimes include equivalent terms in search results.

While no respondent includes equivalent terms or alternative spellings to original search intent at all times, only 9% have never attempted this search strategy (see Figure 10). Although summary responses here are very similar to those of Q9 (that is, tendency to include narrower or child

terms to original search intent), as much as 55% responded differently between Q9 and Q10. This reflects differences in how respondents go about meeting their information needs and a need to separate features that can assist in both search strategies.

2.3.3 Q11

Q11: If a search feature was available that could include the narrower and equivalent terms from controlled vocabularies, would you prefer that this functionality was

- ❑ always included implicitly
- ❑ included by default but can be turned off by the user
- ❑ not included by default but can be turned on by the user
- ❑ not included at all, not of benefit to me

2.3.3.1 Results and analysis

Ninety five percent (95%) of respondents think that a search feature to include narrower or equivalent terms from controlled vocabularies to original search intent is beneficial.

Ninety percent (90%) of those who want narrower or equivalent terms included from controlled vocabularies prefer to have control over its use (that is, ability to turn the feature off). The other 10% want such feature included implicitly. Forty eight percent (48%) of respondents prefer that such feature be included by default with the ability to turn it off while 38% do not want it turned on by default. Only 5% of respondents think that such feature is not of benefit to them. Considering that a significant 43% do not want this feature as default search option or do not deem it beneficial, it may be most appropriate to include it as an optional search feature which a user can turn on.

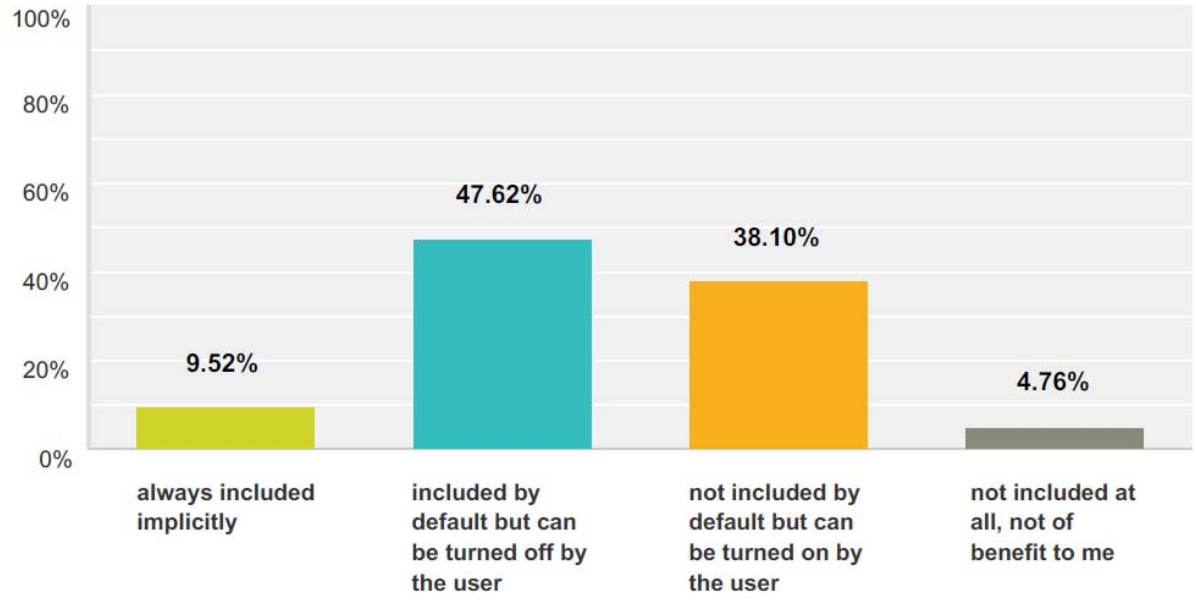

Figure 11 How a search feature to add narrower or equivalent terms to search intent should be included

2.3.4 Q12

Q12: How often do you find that your search results are dominated by results that are not relevant?

❑ always ❑ usually ❑ sometimes ❑ seldom ❑ never

2.3.4.1 RESULTS AND ANALYSIS

All respondents have at some point, found their search results dominated by irrelevant result entries.

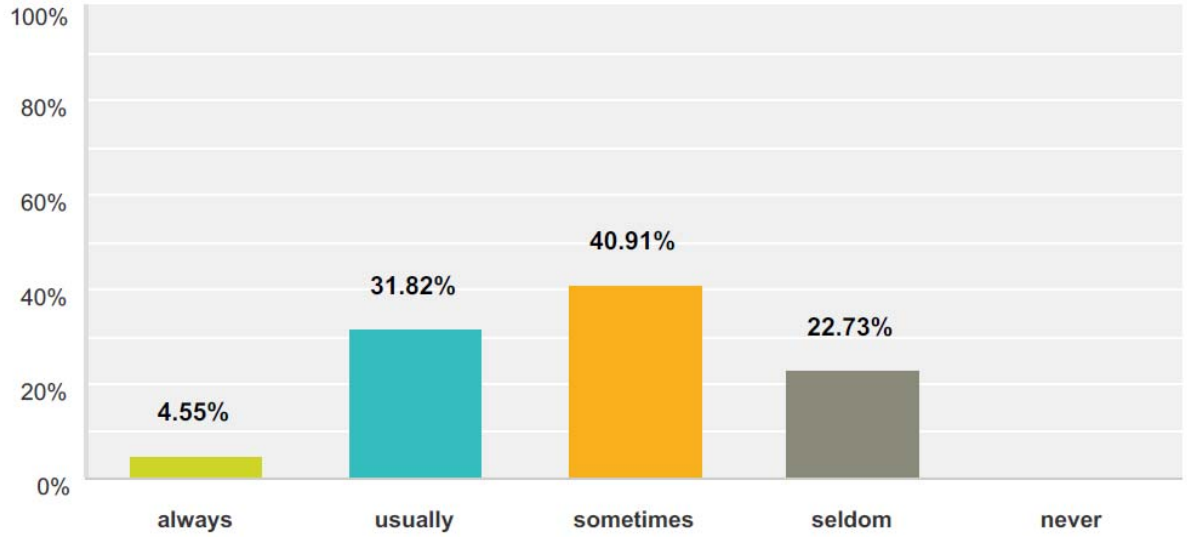

Figure 12 How often search results are dominated by results that are not relevant

Seventy seven percent (77%) of respondents always, usually or sometimes find their search results dominated by irrelevant content. Figure 12 shows that although this does not happen always for most respondents, only 23% said it happened rarely. Hence, a feature that can filter out irrelevant search results will be beneficial.

2.3.5 Q13

Q13: If a search function was available that could search on the intended context/meaning of the search term entered, rather than just matching the term as typed, would you prefer to

- ❑ always specify the context/meaning of your search terms as you build the search (e.g. pick them from a controlled vocabulary)
- ❑ specify the context/meaning of your search terms only if there is ambiguity (e.g. pick the correct definition from a list)
- ❑ let the search engine decide which context/meaning to use, depending on my previous actions or preferences
- ❑ not have this feature, not of benefit to me

2.3.5.1 RESULTS AND ANALYSIS

Eighty one percent (81%) of respondents want to be able to select intended context or meaning of search terms only when there is ambiguity.

Responses to this question indicate a strong support for a feature that allows users to select from a list of competing definitions whenever there is ambiguity in search terms. As shown in Figure 13, about 10% want intended context/meaning of search terms to be decided by the search engine. Only 5% think that a search function to resolve ambiguity in search terms is of no benefit to them.

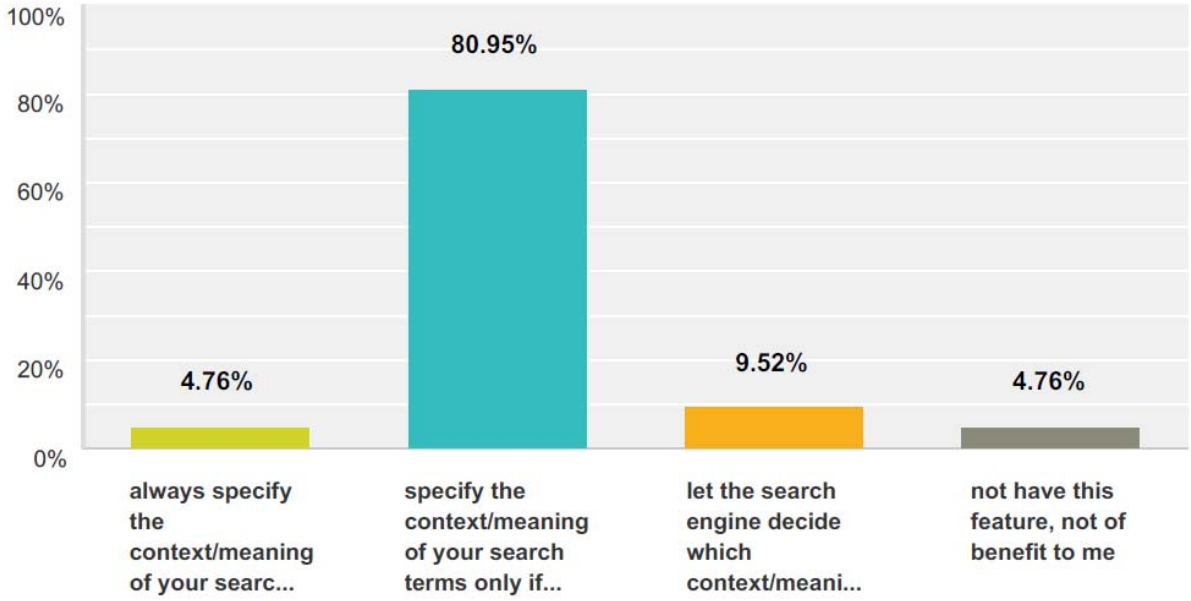

Figure 13 How a search feature to disambiguate search terms should be included

2.3.6 Q14

Q14: Which vocabularies would be useful to you in the sort of semantic search functionality described above?

2.3.6.1 Results and analysis

A list of geoscience-related vocabularies was presented to respondents so that they selected as much as they thought useful. As shown in Figure 14, about 78% of respondents selected the Geoscience thesaurus as a useful vocabulary for implementing sematic search features. This thesaurus describes general geoscience-related concepts so it is not surprising that is was thought useful by most respondents. This is unlike more specialised vocabularies like Chemical analytes (selected by 5.6%) and Fossil taxonomy (selected by 11.1%) which may not be relevant to most respondents.

The usefulness distribution of selected vocabularies can provide an indication of which vocabularies to prioritise if a sub-selection is to be used for implementing semantic search features. As an added comment to responses, it was pointed out that a semantic search feature which requires users to have knowledge of the content of vocabularies being used will be too complicated for users.

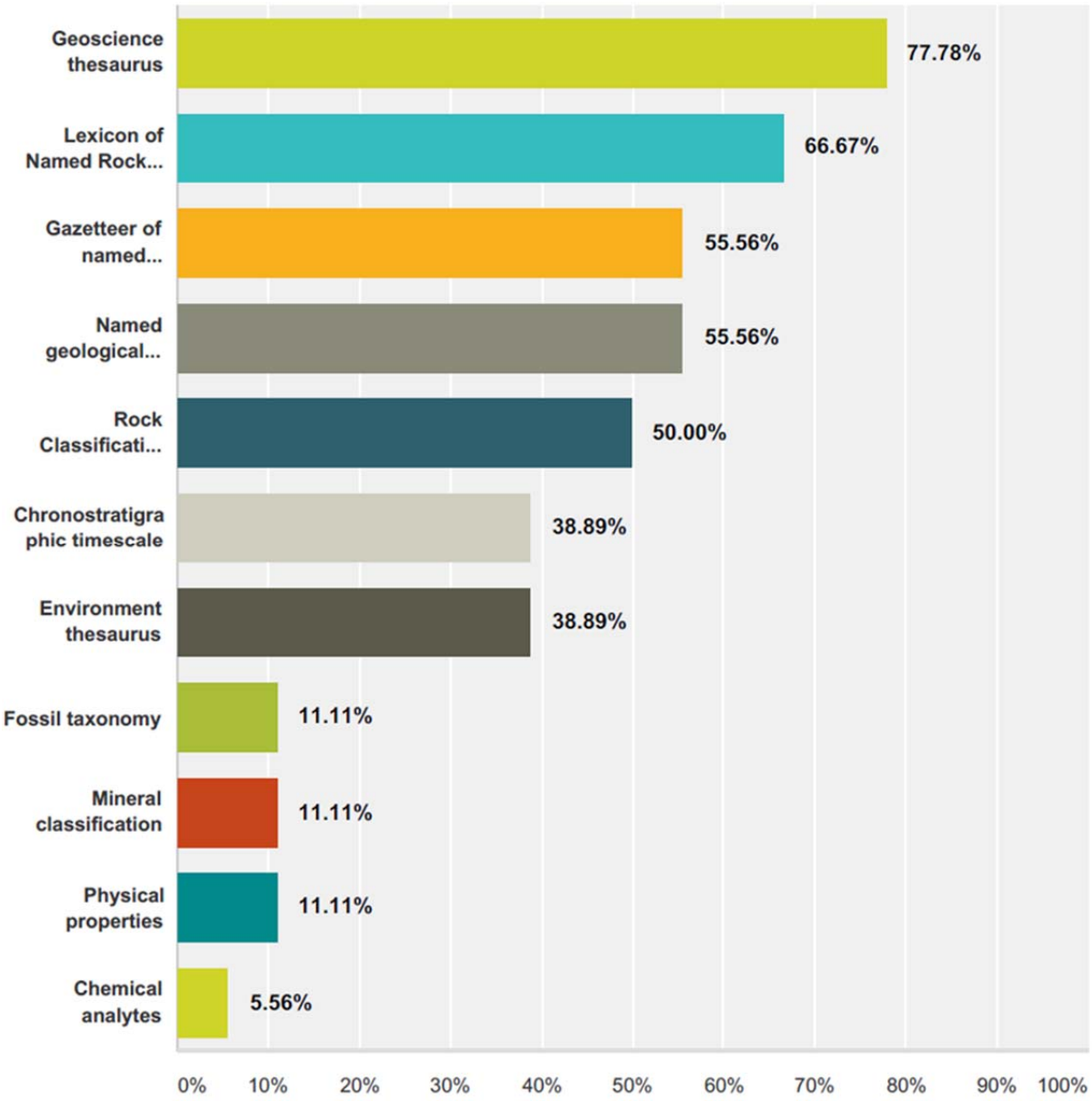

Figure 14 Preference of vocabularies to implement semantic search

2.3.7 Q15

Q15: Might you be willing to volunteer 1 hour of your time to help evaluate a search tool which implements features like the above?

2.3.7.1 RESULTS AND ANALYSIS

Eight (8) respondents indicated willingness to volunteer to help in semantic search tool evaluation.

2.3.8 Q16

Q16: Please provide any other relevant comments such as current search challenges, features you value in a search engine (existing or desired), preferred search engines not mentioned in questionnaire etc. mentioned in questionnaire, etc.

2.3.8.1 RESULTS AND ANALYSIS

There were 4 comments received from respondents. The importance of reusing existing tools when implementing a search feature and the need to allow for different search strategies were pointed out.

3 Key Findings

In this section, literature and data gathering searches (Section 1) are referred to as type 1 search while searches that ask specific questions (Section 2) are referred to as type 2. The key findings from responses collected in this survey are as follows.

1. Popular search engines (e.g. Google) and publication citations (e.g. Science Direct) were reported as most useful applications for search. While publication citations were slightly favoured (59%) for type 1 search, popular search engines were found most useful (79%) for type 2 search.
2. There was a higher tendency for respondents to assess more search results in type 1 searches than in type 2 searches. While about 88% will assess more than 10 results in type 1, only 50% will do same in type 2. However, in both search types, most respondents will not assess more than 20 search results.
3. Eighty two percent (82%) usually or sometimes performed multiple searches or constructed advanced search queries in an attempt to include narrower or equivalent terms to original search intent.
4. Ninety five percent (95%) think that a search feature that expands search terms with narrower or equivalent terms will be beneficial. Ninety percent (90%) of this proportion will prefer control over when to use such feature.
5. All respondents reported that their search results were at some point, dominated by irrelevant result entries. An irrelevant result can be a document which contains supplied search terms but in non-intended sense. This was not a rare occurrence for 77% of respondents.
6. About 95% think that a feature which disambiguates search terms will be beneficial. Eighty one percent (81%) will like to be able to specify intended context/meaning of search terms but only when such terms are ambiguous.

4 Recommendations

The following are recommendations on the implementation of semantic search features based on the findings of this survey.

1. The ranking of relevant search results should focus on the top 20 results as most users may not assess results beyond that. This can provide a cut-off point when evaluating the performance of a search application.
2. A feature which includes narrower or child terms of original search intent will be beneficial in a search application. Also, searches for equivalent terms or alternative spellings of original search intent will benefit users. These should be separate optional search features which users can turn on and off. Narrower/child terms are usually available in domain ontologies (thesauri, controlled vocabularies, etc.). Equivalent terms and alternative spellings can also be described in ontologies as alternative textual labels of concepts. The Geoscience thesaurus and Lexicon of Named Rock Units were identified as the most useful ontologies for this purpose.
3. It will be beneficial to restrict search results to only return relevant documents, omitting irrelevant entries which users find are often returned. Irrelevant entries in search results can

make it difficult to locate relevant documents. This is especially true when irrelevant results dominate the top ranks of search results.

4. The resolution of ambiguous search terms will be beneficial to many users. It is preferable to present search application users with alternative meanings of ambiguous search terms from which they select intended meanings. The disambiguation of search terms should translate to the return of documents which express the intended context/meaning only.

5 Conclusion

While most search applications only match strings of characters between search terms and documents being searched, the outcome of this survey shows that individuals often consider semantics. There is vast support for the implementation of semantic features to help assist in search; these include features to add narrower/child terms, equivalent terms or alternative spellings to original search intent and; features to specify the context/meaning of ambiguous search terms. The majority of respondents will like to decide when to use these features for a search. Instead of an automated process which tries to disambiguate terms, most respondents prefer to select the intended meaning of search terms but only when there can be ambiguity.

This survey was designed in the initial stages of research on semantic search. Results from the analysis of collected responses, which have been discussed in this report, will contribute towards the formulation of research questions and hypotheses. As this was an opt-in survey, we are unable to tell if there was any selection bias in respondents.

References

[1] SurveyMonkey Inc. [cited 20 September 2016]. Palo Alto, California, USA. Available from www.surveymonkey.com

[2] Web of Science. [cited 20 September 2016]. Available from wok.mimas.ac.uk

[3] Broder, A., 2002, September. A taxonomy of web search. In *ACM Sigir forum* (Vol. 36, No. 2, pp. 3-10). ACM.