CHALMERS

Designing Search Analytics with a User-Centered Approach

Master of Science Thesis

Mikael Hjort
Magnus Spånggård

Chalmers University of Technology
University of Gothenburg
Department of Computer Science and Engineering
Göteborg, Sweden, October 2011

Designing Search Analytics with a User-Centered Approach

Mikael Hjort
Magnus Spånggård

Examiner: Olof Torgersson

Chalmers University of Technology
University of Gothenburg
Department of Computer Science and Engineering
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Department of Computer Science and Engineering
Göteborg, Sweden, October 2011

Abstract

Over the last decade, as the amount of digital information has exploded, search has emerged as one of the most important tools available to help us find what we are looking for. But search is not a simple silver bullet that solves the problem - creating an effective and useful search service requires a lot of work on many different levels.

A Search Analytics application is one part of a successful search service. It provides information about how the search service is used; who is using it, what are they looking for, what are they (not) finding, etc. A Search Analytics application takes quantitative data gathered from the search service, processes it and presents it as different metrics which can be tracked and analysed.

This master thesis from the Interaction Design Programme at Chalmers University of Technology aims to investigate relevant aspects regarding the design of a search analytics application from a user centered perspective.

A description of potential users and their level of knowledge, work situation, goals and attitude regarding Search Analytics was created, as well as a list of metrics and tasks which a Search Analytics application should offer to support the users and their organisations in their work. A prototype of a concept design, containing a sub-set of these metrics and tasks, was also designed, implemented and tested.

The work shows that a lot of attention must be directed at the needs of the potential users and their organisations - there is no one solution that works across the board.

This master thesis was done with the support of Findwise AB, a consultancy firm working with enterprise search solutions, and conducted at their Gothenburg offices.

Keywords: Search Analytics, User Centered Design, Web Analytics, Information Retrieval, Enterprise Search

Acknowledgements

The authors would like to thank Findwise AB, especially our supervisor Björn Klockljung Johansson, for the opportunity and help in conducting this thesis work. Furthermore, the authors would like to thank Olof Torgersson at Chalmers University of Technology for being our examiner and supervisor.

Finally the authors would like to extend a grateful thank you to the people who graciously lent us some of their time to participate in interviews and user tests.

Vocabulary

organic search result - A search result whose ranking has been determined through the algorithms of the search engine, and not manipulated by hand.

Key Performance Indicator (KPI) - A measure of performance used by organisations to define and evaluate how successful they are, typically in terms of making progress towards their long-term goals.

search operator - A term that is used to more exactly define a query in some way, e.g. the Boolean-operators (AND, OR, NOT, etc.) or Google's link-operator.

search facet - A predetermined search operator, usually included in the search GUI.

Contents

F Summary of User Test Two

1 Introduction

Search has become one of the most popular ways of finding information, as illustrated by the worldwide number of searches for December 2009 reaching 131 billion (comScore, Inc., 2010).

Modern search originated from more traditional information retrieval (IR) systems, such as digital libraries or other online databases, where the users often were experts and comfortable with complex search systems. These kinds of legacy IR systems have over the last decade more or less been superseded by Web search, which has greatly changed what people expect from a search service and how they use it. For example, Jansen et al. (2000) argues that Web search is different from traditional IR search in the following ways; query modifications are not as common, number of search terms are many times lower in web search, few users look further than the first few pages of results and Boolean-operators are rarely used; similar findings were made by Spink et al. (2001).

These factors of course influence how a modern search service should be designed. They were found through analysis of Search Analytics data - information gathered from the search service describing its use. Search Analytics is ultimately concerned with learning more about people and their needs, through analysis of their actions and behaviour, to create a better search service.

The need for good and effective search service systems has grown in tandem with the amount of information - in today's world, the amount of information generated is growing at an enormous rate, faster and faster each day (EMC Corporation, 2011). The limitations imposed by cost of storage hardware and computing power that existed ten or twenty years ago is no longer a problem, as we now have CPU-cycles, data-bandwidth and storage space for capturing and recording the data. While this opens up many possibilities for improved understanding, the risk of information overload is very much real. To cope with the amounts of data now available one requires models, processes and methods for analysing the raw data and produce tangible knowledge and answers that can be used to solve real-world problems. Keim et al. (2008) pose four questions that any such system must be able to answer:

- Who or what defines the "relevance of information" for a given task?
- How can appropriate procedures in a complex decision making process be identified?
- How can the resulting information be presented in a decision- or task-oriented way?
- What kinds of interaction can facilitate problem solving and decision making?

(ibid, p. 155) The challenge is then to understand which subsets of data that is

relevant to the party concerned and how it should be processed and presented to facilitate positive change.

1.1 Purpose

The purpose of this thesis is to explore relevant aspects regarding the design of a Search Analytics application with a user centered approach. This will require three steps: first, an overview of the field of Search Analytics, both of research and existing products; second, user studies of the different stakeholders, discovering their needs and requirements; third, creating a concept design to illustrate and test the findings.

1.2 Goal

This thesis has two main goals; first, to examine the current state of the Search Analytics field by studying available literature and comparing existing Search Analytics solutions. Secondly, to research, design and implement a Search Analytics application with web technologies with the potential customers of Findwise AB as a basis for the requirements gathering and the design. More specifically, this thesis aims to provide three deliverables that supports the goals:

1. **Tasks and Metrics**. A list of tasks that the process of Search Analytics entails, as well as the data metrics that need to be gathered and processed to support each task.
2. **User descriptions**. A description of the different types of users that were encountered during the requirements gathering phase. Also discussing the needs, requirements and knowledge-level of the users.
3. **Concept Designs**. A functional prototype based on conceptual designs of tasks identified as being relevant to Search Analytics.

1.3 Scope and Delimitations

This thesis has focused on Search Analytics pertaining to Enterprise Search; that is, it is not strictly web search where several different sites are included in the search index (e.g. Google, Bing), and not strictly an intranet where the content is not available to the general public. Enterprise Search means that it is one (outward-facing) organisation which controls the content of the site, which may or may not be publicly available. This has had an impact on the types of requirements that were gathered.

The implementation-phase of this thesis was also limited due to the data available on which the analysis and report generation was performed, as this thesis did not aim to implement data capture or data logging tools. Therefore, the implemented tasks were limited to what the data available could provide. The fact that the data came from a non-commercial source also had some implications on the requirements, which should be noted.

Search Analytics is merely one step in the process of improving the search experience of a site; after the analysis is done, one should proceed to use the newly gathered information to improve the search service in some way. This task is highly dependent on the technical solution that is powering the search service in terms of which changes that can or cannot be made. As such, and due to time constraints, this thesis only deals with the presentation and analysis of the data, and does not concern itself with the design of how any changes are made.

2 Background

2.1 Search Analytics

At its core, Search Analytics is analysing search engine logs for interesting patterns that can be leveraged to improve the search experience for the end user - i.e. what did people search for, and did they find what they were looking for?

Another name for Search Analytics is Search Log Analysis or Transaction Log Analysis, although the most used term nowadays is Search Analytics. Jansen described in 2006 a process for Search Log Analysis that contains three stages: collection, preparation and analysis. This process has a more technical focus where data preparation and manipulation is placed in the forefront, and the underlying motivations for why one should do this, and what to analyse is not dealt with much. This can be contrasted with other more recent processes (among them Jansen's web analytics process, see Jansen (2009)) where, naturally, the data contained in the search logs form the basis for the analysis, but where the data is only considered to be as good as the analysis performed on it. This is also clear from looking at Search Analytics as defined by Hurst (2010), which describes Search Analytics from a user and organisational perspective, where it is described as a process with four main steps:

1. Defining Business Metrics / Key Performance Indicators (KPIs)
2. Reporting
3. Analysis
4. Optimisation & Action

This process is very much focused on the analysis and understanding of what the underlying motivations for performing the analysis are. Hurst's process also stresses the importance of providing the right information to the right people, to allow them to perform as good an analysis as they can, which then can lead to favourable actions which can improve the search service for the end user.

This focus on analysis is similar to the '10/90 rule' created by Avinash Kaushik (one of the most prominent people in Web Analytics), which says that 10% of the money should be spent on technology (in our case, data) and the other 90% should be spent on people analysing the data. Data without good analysis is next to useless. (Kaushik, 2006)

2.2 Enterprise Search

A key understanding for this project is the difference between Enterprise Search and regular Web Search. Fagin et al. (2003) describe this difference as something that is rooted in the different social constructs in which they live. All modern Internet search services exploit the social connections in how information is presented on the web - the hyperlink - to present the most relevant results (e.g. Google's PageRank). Internet search is about finding the 'best' or 'most popular' answer to a query, the answer that has gained most connections among the vast number of answers available.

Intranet search on the other hand is about finding the 'right' answer within a much smaller collection of information that does not have the same underlying social structure as the Internet. Here the information is more tightly controlled and is made to serve the organisation which controls it - it is not up to the users to decide what the best answer is. However, Rosenfeld (2011) points out that from a users perspective relevance is relative. Even within a controlled environment, like an intranet, there is not always a single 'right' answer to a query. This becomes increasingly more true when one takes into account that the most common number of search terms in a query is about two (after comparing Stenmark (2005b) and Jansen, Spink and Saracevic (2000)), which naturally leads to more general and ambiguous queries.

Another interesting aspect of organisations and intranets is the language that emerges; internal search engines contain queries that relate to the organisations context (Wang, Berry and Yang, 2003). Certain words, expressions and abbreviations have a very specific meaning within the organisational context and this is something that must be taken into account when working with and maintaining enterprise search. Rosenfeld (2011) describes it as that there is a tonal pattern to the language that the users use. It is important that the language of the organisation matches the language of the users. To support this a contextualised language thesaurus should be set up, and preferred terms communicated to the content creators.

When it comes to session length, number of search terms and results viewed, there is no difference between intranet and Web search (Stenmark, 2005b).

2.3 Web Analytics

Web and Search Analytics share a number of goals and traits, but they are also different. This section is included to describe those similarities and differences.

Web analytics has emerged as an important tool for all who have a presence on the Internet, and there is a vast amount of books available on the subject

(e.g Jackson (2009), Kaushik (2010), Peterson (2004)). One reason for why the popularity of web analytics has become so great is that it is an easy, unobtrusive and cheap way of gathering data, compared to surveys and interviews; the loss of depth in the data gathered is made up for in volume.

At its core it is about understanding what the users are doing at a web site: which pages do they view, how do they navigate, which documents do they access, which services do they use, how much time do they spend on the site, how often they visit, how did they reach the site, where are they located, etc.

Compared to Search Analytics, Web Analytics has a much broader aim. Organisations can use the data to measure if users perform certain goals that have been set (conversion rate), or test different layouts (A/B testing) or just gain a better understanding of who is visiting the site (demographic research); Web Analytics Association (2008) provides a good overview of the basics. Web analytics is for analysing any relevant aspect across the whole site, whereas Search Analytics is only concerned with the subset of the site that deals with search.

2.4 Comparing Search Analytics to SEO and SEM

Just as with Web Analytics, there are some commonalities between Search Analytics and Search Engine Optimisation (SEO) and Search Engine Marketing (SEM). They are, however, not as strong as for Web Analytics. This is because SEO and SEM is usually discussed in the context of Web search and not site/enterprise search. They concern themselves with analysing a site and making changes to improve the organic ranking of the site in different Web search indexes to gain a competitive edge. This is not the case for Enterprise Search, where the organisation is in full control of the search index and can thus use other techniques to control the rankings.

For example, in contrast to Web search, where meta-data is scarcely used due to many sites using misrepresenting meta-data to lure in more visitors, Enterprise Search could use meta-data due to the controlled space in which it lives. The use of meta-data becomes even more crucial when considering that a lot of the content on intranets are not in searchable form (documents, database interfaces, etc.). Enterprise Search also allows for artificial results to be inserted into the index to appear for certain search terms (called key-matches), which have been selected by someone within the organisation because it perfectly answers that query. All in all, having this much control over the search service removes the need for SEO and SEM.

2.5 Zipf's Law

Contrary to what one may think, Search Analytics does not require extreme amounts of resources to make a significant difference. The reason is that, as with many other phenomena, search phrase usage follows Zipf's Law, or a Zipfian distribution. Simply described, Zipf's Law says that the most frequently used words will make up a large portion of the total number of queries (Rosenfeld, 2011); more exactly the most frequent search phrase will account for twice as much as the second, and the second twice as much as the third, etc.

Making sure that the most frequent queries (the "head" of the Zipf-curve) provide good results will result in an improved search experience for many users.

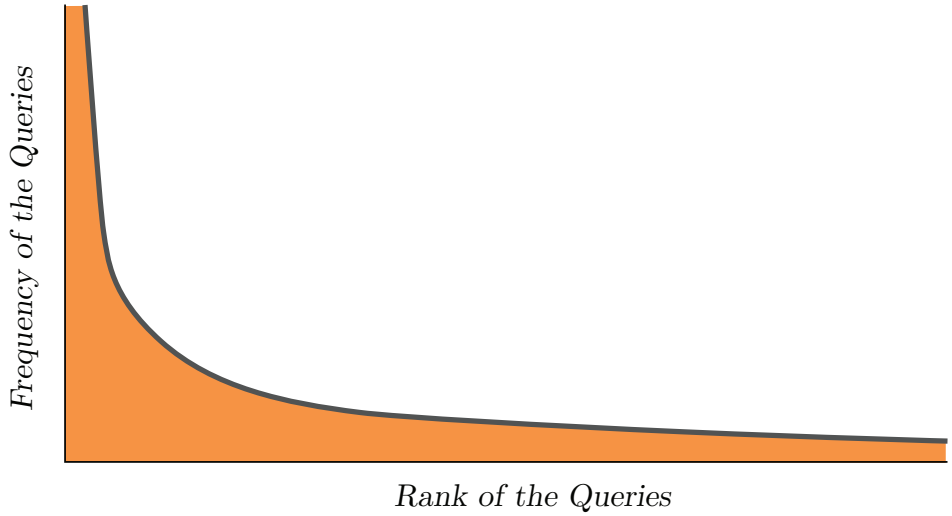


Figure 1: A Zipfian distribution for search queries.

2.6 Levels of Analysis

Jansen (2009) describes three different levels of analysis regarding Search Analytics: term, query and session.

2.6.1 Term

A term is defined as a character string separated by a delimiter (i.e. individual words). All terms are tracked and analysed. The most common analysis techniques are most frequently used terms, unique terms and term co-occurrence.

2.6.2 Query

Query analysis looks at the whole search string (i.e. zero or more terms). Here the frequency of the query and number of terms and/or search operators are tracked. The results of a query is also considered here. How many results were returned, which were viewed, the ranking of the results and time spent looking at the results.

A short note on semantics: many tools use the word 'keyword(s)' when referring to search terms or search queries. We find the use of this word for denoting both terms and queries to be unfortunate and a possible source of confusion - they are not the same thing! We have therefore opted to use 'term' to denote a single term and 'phrase' to describe a string of one or more terms. The term 'query' can mean both a single as well as several terms, no distinction is made then.

2.6.3 Session

A session encapsulates every query that a user has performed during a time period, but it is important to understand that there exists two ways to view what a session is. The first is how a user views a session: to her it encapsulates every action during the time it takes her to fulfil a single information need, to find the answer to her question. The second is how the search system views a session, which is usually limited to a specific time span or a timeout due to inactivity.

Currently, there is no way to combine these two definitions. Some of the difficulties that exist have been detailed by Stenmark (2005b, 2005c), e.g. a complicating issue is when the user initiates a new search for an unrelated information need during or right after the previous search - from the user's perspective these are two separate search sessions, but the system which is gathering data for the analytics will most likely record this as a single session. The reverse is true as well: if a substantial amount of time passes between two queries that belongs to, from the users perspective, the same information need, the system will not recognise them as belonging to the same session. This is a technical limitation which is yet to be solved.

Another type of behaviour called query reformulation (adding, removing or changing search terms) can make it harder to define session boundaries precisely - a query can change drastically due to additional information and learning from previous queries, but the aim remains the same.

2.7 Search Metrics

Search Analytics is quantitative research, and the analysis is dependent on the metrics that can be gathered from the search system. Ideally, search metrics should be derived from business goals and incorporated into the design of the search system.

There is no defined set of relevant search metrics (Spink et al., 2001), but some of the most common are listed in e.g. Graf, Macdonald and Ounis (2007), Jansen (2009), Lewis (2010) and Rosenfeld(2011). Below is a list of metrics from those sources, which the authors find relevant and valuable in the context of this thesis.

2.7.1 Search Term Co-occurence

Term co-occurrence is a method that finds the different combinations of terms that have been used together in a query; search queries containing term A also contained term B1 or B2, or B3, etc. This can be used to see how terms are used together and which term-pairs that are most popular. This can yield new and interesting patterns regarding information needs and vocabulary that might not emerge when just looking at single terms or full queries.

Buzydlowski, White and Lin (2002) use co-occurrence as a way to suggest related content; when one term is used in a search, results related to that terms co-occurrence terms are also displayed, providing the user access to potentially valuable information more quickly.

2.7.2 Bounce Rate

The number of users who do not progress further into the site than the search engine result page, after having performed a search - either their information need was not met or they found the information among the search results, in a page title or result snippet.

2.7.3 Query Length

Displays statistics regarding both length of the whole search phrase and the number of terms used. Can be used to decide how many characters wide the search box should be, and also to learn about how people search in general: do they write short queries consisting of only a few terms, or do they write long, natural speech-like queries. This information can be used to tweak the search engine and meta-data to better suit the users language.

2.7.4 Top Queries

The most common queries, in descending order of the number of searches made for each query. Easy access to the most common search queries enables the analyst to make the best use of her time. Making sure that those queries return good results will give the greatest return on investment and time spent, due to the fact that those queries are performed by many people.

Top queries has little to do with exact numbers; it is not whether a query has had 10 or 10 000 queries made, instead one should look at the percentage of the total number of queries that the query represents. This follows due to Zipf's Law (see section 2.5), which states that a rather small number of the query phrases make up for a substantial part of the queries made. From this follows also that one should display the number of searches for a query as both an exact number and as a percentage of the total number of searches. To take it one step further would be to display the cumulative percentage, the sum of this and all the queries above this position in the list, to better communicate how large a part of the queries that have been dealt with so far (Rosenfeld, 2011).

Tackling the Long-Tail of Search

Rosenfeld (2011) describes a method for analysing the long-tail of search, the queries that individually does not have that many occurrences but together make up a significant chunk of the total number of queries. Considering the size of the tail, it is not feasible to examine every query there. Instead, a random selection of 50-100 queries can be fetched and then analysed for any common patterns.

2.7.5 Selection Rate

The selection rate is a measure of the number of queries which led to a result link being clicked, a click being an indicator that the result link is at least interesting. A low ratio would be a sign of the users finding the search results poor.

2.7.6 Query Evolution

This is not a simple metric which is expressed in a number or sentence. It is a collection of all search phrases stemming from sessions with a common search phrase. The information presented displays how a search initiative has evolved, through query reformulation and/or adding/removing any search filters.

2.7.7 Click Path

Displays the navigation-path that the user took after leaving the search result page. Additionally, the metric could try to find aggregated occurrences of pages in all the navigation-paths.

Displaying this navigation-path to the analyst would provide the analyst with information that could be used to improve the ranking among the result links. It might be that it is not the first result link that appeared in the search results that is the page that best satisfies the users information need, but a result link that appears further down in the ranking or perhaps the search engine has not included that page in the results for this query. The first link, however, links to the correct page, which is why the users do not go back to the search result page or reformulates their query.

This complex metric aims to provide better insight into where the user finally found the information that satisfied her information need. The hyper-linked nature of the Web affords further navigation to other pages where the user might find the exact information she was looking for.

2.7.8 Average Number of Queries

To track how many queries that are made per session on average. Gives an indication to the need to, and how willing the user is, to reformulate their query (Rosenfeld, 2011).

2.7.9 Average Number of Result Pages Viewed

In relation to a query, display how many pages of result links that were viewed on average. If it is hard to find the correct result link (the information the user is looking for is not on the first result page) this number will be larger than one. A larger number would indicate that people can not easily find what they are looking for on the first result page, meaning that the search engines ranking is not as good as it should be for that search query.

As an average over all queries, this would give an indication to what type the users of the search service are; are they persistent and dig deep through the results, or do they give up easily and will not go past the first page?

2.7.10 Search Session Time

Displays the (average) time spent using the search service. The time measured is the session time.

2.7.11 Result Type

This metric is designed to display how many clicks an artificial result link has received. An artificial result link is a result link which has been added to the query manually (e.g. a key-match or a related link) or has had its ranking manually edited (e.g boosting). Contrasting this number to the regular, organic, search results will reveal how the users value the artificial link, and whether it should be there at all.

2.7.12 Content Type

Displays statistics regarding the amount that different types of content (web page, image, document, etc.) is accessed through the search service. Should be displayed in relation to the total amount of information of that type. This statistic can be both a summary of all queries on the site or coupled to a specific search phrase.

The ratio between accessed and available content will indicate which types of content are most sought after and can be used to guide future content creation.

2.7.13 Clicked Results

For a query, which result links on the search engine result page were clicked.

2.7.14 Top Result Links

Lists the most visited pages, regardless of the query which led to the page. This will give a fast and clear view of which pages that should receive some extra attention from the content owners.

2.7.15 Session Activity

Measures the amount of activity during a session. Activity can be defined in several ways: queries per session, result link clicks per session and as selecting

facets. The metric is used to indicate the level of user engagement.

2.7.16 User Tracking

By tracking users one can learn a lot. How often do they search, how do different user groups differ in behavior, what does new vs. returning users search for, what type of user uses certain functions, how many searches does a user do per month, etc.; user tracking can be viewed as a form of segmentation.

An important consideration regarding user tracking is that the data collected should be anonymised. Not only do many countries require it by law, but it is ethically sound to do so.

2.7.17 Number of (Unique) User and Searches

This metric is designed to show how much the search service is used. The numbers for unique users provide information regarding the size of the user base and the number of unique searches represents the amount of information needs. Non-unique numbers provide usage and performance numbers.

2.7.18 Entry Points of Search

Displays information regarding from which pages search was initialised.

2.7.19 Devices, Browsers and Plug-ins

Statistics over which devices, browsers and plug-ins that are used by or available to the users of the site. Provides information to the site owner about which devices, browsers and plug-ins that they should support, e.g. if a mobile-version of the site is needed or if they can use the Flash plug-in to present information without loosing a large set of users.

2.7.20 Queries with zero results

Displays the queries that returned zero results, i.e. there was no matching content available.

2.7.21 Term Frequency

To be able to see how frequent a specific search term is used when decoupled from any other terms in the search phrase. From this, different patterns may emerge than when whole search phrases are analysed.

2.8 Data Collection

Search metrics themselves are made up of different data fields that can be logged during a search session. Remember that it is the underlying design of the search service that determines which data fields, and therefore which metrics, that can be gathered. For examples of what a search log can look like, see Jansen (2006, p. 413) or Rosenfeld (2011, p. 31).

There are different ways to collect the data - either through JavaScript, which collects information from the users' browsers and sends it back to the logging server, or through mining search engine logs. The majority of publicly available Web and Search Analytics tools (see section 4.2 for a list) use JavaScript to gather information; only a few lines of JavaScript code needs to be added to each page that is to be tracked.

2.9 Related Analytics Research

Graf, Macdonald and Ounis (2007) describe four click-based metrics that they use to measure the performance of the search system. They measure the percentage of sessions resulting in clicks, the delta between query submission and first click, the frequency of query reformulation and the ratio of clicks and undesirable actions. While their result were not always great, just the idea of using different types of interactions, besides stating a query, as metrics is interesting and inspiring.

Azzopardi (2007) examines the users behaviour on the search results page, which he calls search and browse, and uses it to take an interesting approach to what information gathered during a session can be used for; he proposes that the sequence in which a user has accessed documents should form the basis for measuring the performance of the search system and hence the experience of the user.

Agichtein, Brill and Dumais (2006) shows how user behaviours can be used to affect the ranking of search results. They use implicit user feedback gathered from click and browser data to re-rank the search results. It was very useful for the top-results for a query, as well as for queries with poor original rankings of the results. This idea of leveraging user behaviour to rank results has been central to this project.

2.10 Information Visualisation

Visualisation used to mean the images that exist within our mind, conjured up while thinking. Today it has taken on a different meaning; Information visualisation is fundamentally about designing the ability to perceive, sort through and select information existing in the real world. Our minds are, however, still heavily involved in this complex task. Ware (2008) says "we should think about graphic design as cognitive tools, enhancing and extending our brains." (ibid, p. ix). Interacting with information entails solving some kind of cognitive problem - we want to extract meaning from the data. Finding meaning is about discovering patterns within the information, and patterns can often lead to a solution.

Tufte wrote "Clutter and confusion are failures of design, not attributes of information" (Tufte, 1990, p. 51). What he was getting at is that our world, and hence our information, is inherently complex, and working towards simplifying the world/information will not provide us with good results. It is the task of the designer to create a design that enables the user to navigate and understand this complexity; it is not the task of the designer to simplify or change the information, only to change how it is presented - with complexity and all.

2.11 Perception - Attention and Pop-out

When looking at information, what one is looking for heavily affects what one is actually seeing. The brain boosts the signals for red arrows if one is searching for just that; our attention directs what we are seeing. This has implications for information design. Understanding the tasks and goals of the user is necessary in order to visualise the information in such a way that it supports these goals.

Ware calls this 'visual task analysis'; goals and tasks are analysed, and from them a set of visual queries (how we obtain information in the world) that need to be supported are extracted. The design is then realised using shapes, textures, colours, motion, space, etc. in a way that supports this.

It is not only attention that can boost the ability to find something in a visualisation. There are several ways to make objects 'pop-out'. Generally, Ware says, it is about visual distinctness with regards to the surrounding environment. This is dependent on feature-level contrast between objects, which is based on differences in colour, size, orientation, motion and stereoscopic depth. For these differences to be noticed they must be sufficiently large. To be able to distinguish between many objects, it is recommended to use several features. It is important to note that what enables us to find an object distinct is not something that can be learned - it is a physiological feature of our brain.

Another way to support visual search is to use multi-scale structures; placing

smaller-scale structures within larger, visually recognisable structures will make search easier and faster if the hierarchy is understood.

2.12 Patterns

As for objects that pop-out, Ware (2008) describes several ways to design for distinctness when it comes to patterns that denote a relationship between objects. Generalising, one can say that patterns are distinguished by contours, regions and spatial layout.

There is also a semantic layer to patterned objects - e.g. similar patterns mean that the objects represent the same thing in some conceptual way, or that two objects connected with a thick line has a stronger relationship than those connected with a thin line. Making use of these semantic connections opens up a big, natural source of information.

2.13 Colour

Tufte (1990) lists many uses for colour: to label, to measure, to represent and to decorate. But how one is exactly to go about this has no easy answer - it is a very complex task. There are many ways to do it wrong, therefore Tufte proposes that the first principle when it comes to colouring information should be "Above all, do no harm" (Tufte 1990, p. 81).

One should have a clear goal when applying colour to a design, and understand the purpose of each colour used so that personal preference or taste does not dictate the design. Colour is a complex subject, but nonetheless there are many rules and guidelines to help. Most importantly, how we perceive colour is universal, not counting the different types of colour-blindness (Ware, 2008).

2.13.1 Contrast

Contrast is the difference between two colours. This might seem like an obvious statement, but the reason behind this is because of how our brain works. The brain perceives differences in colour, not absolute values. This is easily demonstrated with the simultaneous contrast phenomenon; the grey bars appear to have different shades when they in fact have the same. This has big implications when it comes to colour selection and composition. (Ware, 2008)

Figure 2: Illusion demonstrating the simultaneous contrast phenomenon.

2.13.2 Colour-Coding

The two most important factors when designing colour-coded information is visual distinctness and learnability. For a complex visualisation no more than twelve different colours should be used when colour-coding, and one should start with the unique hues and then use other easily identifiable colours. One must also consider background colour and other colours present in the visualisation to achieve the greatest pop-out effect. (Ware, 2008) When colour-coding regions, small areas should have strong colours with good luminance contrast against the background, which should use a low-saturated and muted colour.

2.14 Visual Noise

Tufte talks at length about visual noise, especially in the form of what he calls '1+1=3 or more'. This is the effect when two "heavy" (high contrast) lines react and produce a third line out of the negative space between them. This third line clutters the visualisation and steals the focus from what is really important. Tufte's solutions for this problem is first to, if possible, edit the presentation so that the negative space does not appear. If that is not possible then the contrast between the lines and the background must be lowered, which will produce a calming effect and reduce the clutter.

2.15 Related Visualisation Work

Oosthuizen, Wesson and Cilliers (2006) describe how to visualise a session-based analysis method called sequence analysis. The goal is to find patterns in the paths that the users have used to navigate within the site. They visualise this, and in consequence the whole structure of the site, using a radial tree graph, applying several simplifications to provide a less cluttered view of the structure.

Chen et al. (2007) does something very similar, also with radial trees. Their visualisation tools makes heavy use of colour, size and thickness to relate different data to the user. They also argue for what they call Visual Data Mining (somewhat similar to Keim et al.'s Visual Analytics), a process that uses existing data visualisations to generate new insights; an initial hypothesis is formulated, it is evaluated through visualisations that provide more information, a new hypothesis is formed which is then verified through visual data mining methods.

Both Keim et al. and Chen et al. advocate for making the visualisation and interaction with the data a part of the analytical process; a tight feedback-loop between human and computer, interaction and analysis. While this is an interesting approach, it is not directly incorporated into this project. Keim et al. recognise the limitations of fully automated analysis, and try to incorporate the human intelligence to assist and guide the analysis when needed. This fits well with Kaushik's 90/10-rule - people are needed to make sense of the analysis.

3 Methodology

This chapter describes the different methods that were used during the project.

3.1 Comparative Review of Existing Products

This comparison of existing tools for Web and Search Analytics was done to educate the authors about the current state of the tools available, and to enumerate and compare the functions they provide and how they have been visualised.

The aim was to discover both common (to see which the basic functions are) and unique (to see how others have tried to innovate) functions as well as the most common way of visualising information within this context.

The comparative review consisted of three main steps. First, finding a good amount of products that can be used for the purpose of Search Analytics. Second, analysing each product and determining which functions that are relevant to the field of Search Analytics. Last, these functions were organised in an affinity diagram to discover commonalities between the different products.

3.1.1 Affinity Diagrams

When a system domain is complex and poorly understood, it is hard to quantify its data, and even if it is done it can still be hard to understand how they are related to each other. Affinity diagrams is an excellent method to use in order to organise the data, to get an overall grasp of the problem and to understand the problem domain.

The Affinity Diagram was invented by Jiro Kawakita, as a way to organise ideas and data (Kawakita, 1982). It is therefore also called a KJ-diagram. Its basic principle is to write down ideas and data in short simple words on post-it notes. These notes are then grouped together according to how their content relate to each other. These groups then allows one to discover patterns within the content that has been analysed.

3.2 Interviews

Techniques such as shadowing were discussed but dismissed as being too time consuming, and therefore regular interviews were chosen instead. This proved to be a correct assessment as it proved difficult to even get time to do an interview. A, for the user, less time consuming method for gathering data, such as sending

out a questionnaire, was also considered as an alternative. However, it was discarded as it would mostly produce quantitative data and that of interest was user stories and other qualitative data. This is the reason why interviews was used to elicit data from people who use this kind of tools.

3.2.1 Semi-structured interview

The interviews were conducted according to the guidelines of a semi-structured interview; Maguire (2001) says a semi-structured interview is preferable when you yourself understand the broad issues of the problem at hand but do not know the detailed problems that the end-users have. This made a semi-structured interview a perfect fit as this was the situation that the authors found themselves in.

3.2.2 Mediating tool

A technique where the interviewer presents an artefact to the interviewee to facilitate discussion around it. This can help the interviewer to probe deeper regarding how the interviewee understands the artefact and how she uses it. It can also assist in discovering scenarios that otherwise would not be mentioned during the interview as the artefact can trigger memories in the interviewee (Karlsson, 1998).

3.3 Personas

Personas is a technique which tries to represent the end-user by creating a fictional character with characteristics and requirements corresponding to the actual users (Chang, Lim and Stolterman, 2008). It is a good method to use when it is difficult to have regular contact with the end-users. Using personas for different usage-scenarios allows the development team to have a shared understanding of the type of end-user that they are designing for.

3.4 6-3-5

Is a method designed to brainstorm creative ideas. To start, one formulates a problem or scenario that needs to be solved. Then the 6-3-5 method can be applied as follows. Six people gather around a table for the purpose of solving the problem or scenario. Each person draws, on a piece of paper, three ideas for solving the proposed problem or scenario. The piece of paper is then passed on in a rotating fashion. The ideas are then iterated over to improve on the original

designs. The iterations continue until everyone have the paper that they started out with. No iteration may be longer than five minutes (Rohrbach, 1969).

3.5 Prototyping

Two different types of prototypes were used. The first one was a low-fidelity low-resolution prototype, consisting of wireframes. This kind of prototype is good to use when one does not want to concentrate on the finer details of a design, but rather focus on the overall structure.

The second type was a high-fidelity, high-resolution prototype consisting of a detailed paper-prototype. This kind of prototype is appropriate to use when one wants to test an almost finished design, to see if the elements within the design works as the designer intended (Sharp, Preece and Rogers, 2007).

3.6 Heuristic Evaluation

A heuristic evaluation is a very popular and appreciated method for evaluating a user interface. This method involves an interface expert, who evaluates the interface based on a selected set of usability heuristics and principles (e.g Nielsen's Usability Heuristics (Nielsen, 1994)). Often, several independent evaluators are tasked with evaluating the interface, to generate enough data to discover any and all common problems. The evaluator inspects the interface several times, alone, to attain an unbiased result. Finally, the evaluators present their findings, and together with the development team they discuss solutions. (Conyer, 1995; Holzinger, 2005)

As this method can be used without the need of end-users, and only require a paper prototype, it can be considered to be a quick and dirty testing method - meaning that it takes a reasonable amount of time and resources to perform.

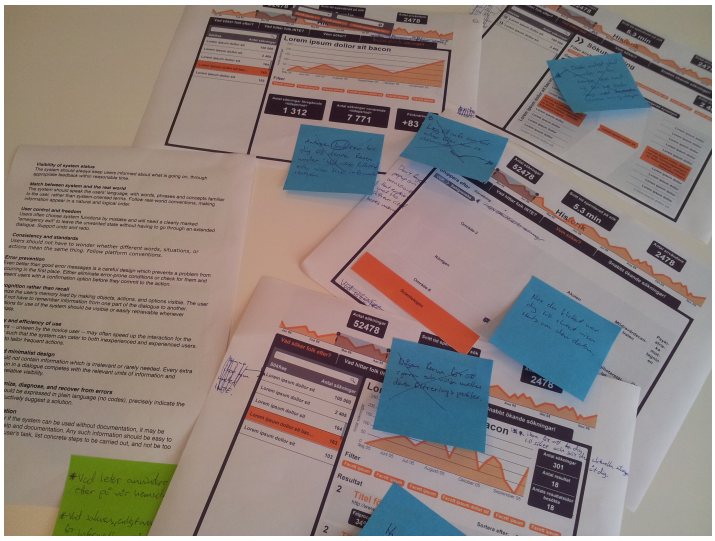


Figure 3: Picture of the authors heuristic evaluation.

3.7 Participatory Design

There are several forms of participatory design, and the one used here can be considered to be quite light as the participation of the user is not very extensive.

In this participatory design, a user-representative is invited to perform a task scenario, during which she is asked to 'think aloud'. Thinking aloud means that they should explain what they do, and why they do something. An observer should be present, prompting the user with questions about what they are thinking, their intent and expectations, etc. The evaluation session should be recorded on tape or video for later review. (Maguire, 2001)

An advantage of 'thinking aloud' is that you are getting the user's immediate feedback, without them filtering it by thinking it over and rationalising it. The task of analysing the test can be cumbersome but will yield qualitative data as it is possible to both study how the users interact with the system and how they react to it.

4 Process

In this chapter the authors describe the steps they performed during the project, and how different methods were utilised.

4.1 Literature Study

As the authors had, at the beginning of this project, little detail knowledge about the subject, a literature study of it and any related subjects was initiated. The following subjects were identified as being relevant to the project: Search Analytics, Web Analytics, Enterprise Search and Data Visualisation. Research papers, books, articles, blog-posts and presentations were studied and analysed.

This study gave the authors opportunity to gather knowledge and gain an understanding of the subject(s), which was integral to fulfil the goals of this thesis.

One of the more useful books, *Search Analytics for Your Site* by Louis Rosenfeld (2011), was not published until June of 2011, when the research phase of the project was over. The book, however, served as a good validator of the work that the authors had done so far since it came to many of the same conclusions.

4.2 Comparative Review of Existing Products

The products analysed were all tools that were free of charge and available to anyone. These types of tools were the only tools that could be analysed because the authors could not get access to any paid tools. The following tools were analysed:

- Piwik
- Site Meter
- Clicky
- Open Web Analytics
- iWebTrack
- Google Analytics

Most of these tools are Web Analytics tools. However, as Search Analytics is a subset of Web Analytics it was possible to find functions in each tool that dealt with Search Analytics. Each function was then evaluated by describing its purpose, listing the metrics that were presented in that function and how they were visualised. From this evaluation it was then possible to group the

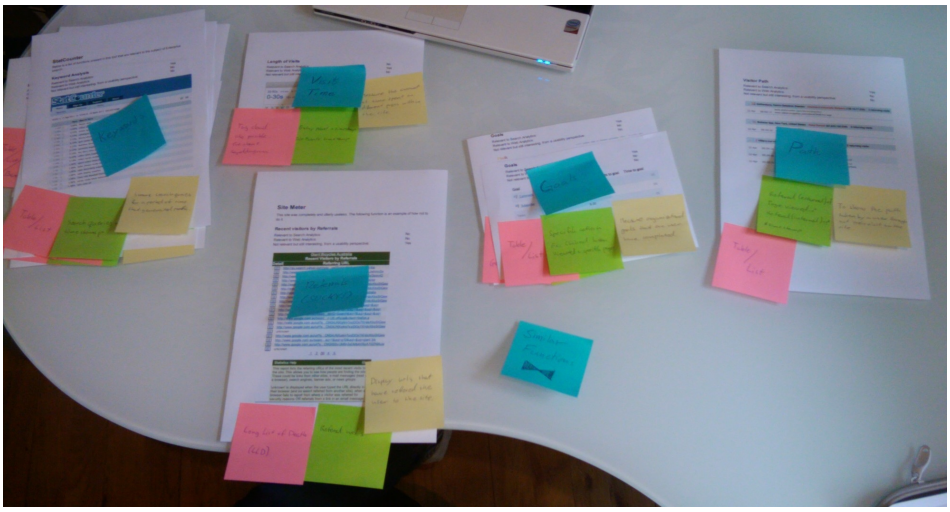


Figure 4: An affinity diagram of the comparative review of existing products.

functions and different patterns and similarities started to emerge. This was done by printing each function on paper, attaching coloured post-its to them and then grouping them. In this way it was simple to get a quick overview of the patterns and similarities that emerged from the different piles of paper.

When grouping on functions, five different categories emerged:

- keywords: what the authors would describe as 'search phrases', the queries the users performed on the site.
- visit time: how much time the users spent on the site for each session.
- referrals: which external search engine led them to the site, also including additional information, such as search query.
- goal tracking: how many users completed certain goals (download a file, buy a product, etc.)
- visitor path: how the user navigated through the site.

Of these, search phrases, visit time and referrals were the most common. It is important to note, however, that many of the tools only displayed visit time and referrals coupled to a search phrase. When taking this into account it is clear that search phrases is the single most popular tracked metric.

Furthermore, the search phrase-group could be divided into three sub-groups: query analysis, term analysis and referrals for search phrases. Of these, query analysis was the most common by far.

Grouping on visualisation schemes resulted in three groups emerging: table, table with graph and tag-cloud. The most common was the table (with or without a graph), while the tag-cloud must be considered to be quite rare.

This comparison shows that the tools and functions for performing Search Analytics available today are mostly focused on tracking search phrase frequencies. There are tools that provide more information, such as visit time, referrer and number of unique visitors, for each search phrase. More advanced functions, such as goal tracking and click-path, do exist as well, but are not very common.

When looking at how the data was visualised the authors were quite surprised of how little difference there was. Tables of text and quite basic and simple graphs were prevalent, and the only unorthodox (and only moderately so) visualisation was the tag-cloud.

When comparing the amount of information conveyed to the user for each function, clear differences were found. For search phrases there was sometimes more data than just the frequency of that search phrase displayed, e.g. data regarding visit time, actions performed / pages visited, referrer, goals completed / conversion rate and number of unique visitors for each search phrase. The amount of extra information varied from one to all of the above.

4.3 Interviews

The number of people that use Search Analytics tools today is quite low and it proved somewhat challenging to find interview subjects, but with the help of Findwise AB and other contacts seven interviewees were found.

The interview questions were structured in seven different categories. Each category of questions were designed to give an overarching perspective of that category. The categories are presented bellow.

4.3.1 About the Interviewee

The questions related to this category were designed to elicit information which was related to the interviewee's work situation, work experience and education. The purpose of this was to use these answers to create personas that could be used during the design and implementation phases.

4.3.2 Organisation

This category was constructed to try and elicit requirements which originate from the organisational goals regarding information in general and search in particular. The goal was to find problems that could be solved or assisted through Search Analytics.

4.3.3 Current Tools, Practices and Goals

As one of the more important categories, this was designed to shed light on to the situation in which the interviewee is currently placed when working with a search analytic tool - how they use their tools, what shortcomings they believe their tools inhabit and what it is that they are trying to solve with their tools.

4.3.4 Data visualisation

A short category whose purpose was to have the interviewee reflect upon the different data visualisations in their current tools, expressing pros and cons, likes and dislikes and the general look and feel.

4.3.5 Actionable Change

When an irregularity was discovered within the search service, was there any actions taken to correct it? If so, what where those actions and was there some sort of feedback if these actions helped to actually resolve the error?

4.3.6 Other Stakeholders

This category helped to discover if the interviewee shared their findings with other stakeholders within their organisation. What these people wanted the findings for and what, if any, challenges did this present to the interviewee.

4.3.7 Overall Experience

What was the interviewees overall experience and thoughts about Search Analytics in general and the tools used in particular, i.e. did they find it valuable or a waste of time, easy and pleasant to work with or a hassle to use?

4.3.8 Mediating Tool

During each interview the interviewee was asked to demonstrate a typical work task that they might preform by using the tools that they have. Using the interviewees own tools as a mediating tool allowed for thorough probing as it provided an excellent opportunity to discover problems that the interviewee might have with their current system.

4.4 Personas

This technique was used because it was difficult to keep regular contact with the end-users during the thesis work. The personas were used as a constant reminder of what kind of end-user the design was intended for, and were continuously consulted when making design decisions.

During the interviews there were a couple of categories specifically designed to elicit user stories that could be used to design personas. From these categories it was possible to elicit characteristics about each interviewee. The characteristics from all the interviews was then analysed by grouping them in different categories. From these grouped categories two types of personas started to emerge.

These two types were then realised by making a fictional person with these characteristics. Each person was then coupled together with a scenario, a picture and a name (see appendix B) to make them more lifelike and more memorable by attempting to construct an image of a person that actually exists. While making the personas an existing template for personas was used as a guide (Tortolano, 2009).

4.5 Metric Selection and Modification

One of the goals of this thesis was to find metrics that would be relevant for a Search Analytics application in the specific context of this thesis. To this end, a list of metrics was compiled from related research and literature, existing products and from the interviews conducted.

Some of the metrics here are modifications of existing metrics found in literature or in existing products, while others were created by the authors in response to findings from the interviews or to address situations that the authors identified during their work.

The usefulness of the metrics as presented here has not been scientifically evaluated but are grounded in research and the current practices of the field.

4.5.1 Error Rate

This is somewhat related to bounce rate (see section 2.7.2), which in its traditional meaning (number of users that only viewed one page and then exited) is not a very useful metric for Enterprise Search Analytics. The authors instead propose a metric formulated in the following way: 'the percentage of users that visited a page from a list of search results and then returned to the list of results within a set amount of time': the error rate.

This formulation's aim is to count the occurrences when the user could not find the answer to her question on the page, and as such returned to the list of results to find a better one. The terms pogosticking (Morville and Callender, 2010) and short/long click (Boykin, 2011) has been used to describe this kind of behaviour, and Graf, Macdonald and Ounis (2007) uses the name 'ratio of clicks and undesirable actions' for this (also including query reformulation into the undesirable actions).

As to how long the time threshold should be, recent findings (Liu, White and Dumais, 2010) suggest that the first 10 seconds is when many users decide whether a page is worth their time and that a user is still very likely to leave within 30 seconds if the page does not present any relevant information to them. Therefore the authors chose a 30 second threshold.

This metric was created to provide a better way to measure the value and usefulness of a search result link. We argue that simply counting the number of clicks on a search result link is not a good enough metric.

4.5.2 Unsatisfied Information Needs

This metric expands the idea of the zero results metric (see section 2.7.20). Instead of just displaying the searches that returned no results, this displays all searches that failed to fulfil the information need of the user. We believe this includes three different scenarios: the search returned no results, the user did not click on any of the result links, and all the result links had a high error rate (see section 4.5.1).

Each of these scenarios led to the same conclusion - the user did not find the information she was looking for. This is something that the analyst will have to investigate; should the site contain information regarding this subject, is the content available good enough, could the user have gotten all the information he needed from the search snippet?

4.5.3 Query Length

A small addition to the query length-metric: in the same way as the length of queries could be used to determine the width of the search box, it is important to consider the effect the current width of the search box has on the length of the queries entered; would a larger box lead to longer queries, can other design decisions influence the users' behaviour.

4.5.4 Query Evolution

This metric was originally designed to be used for when a user performed a search that returned zero results, and later expanded to include any searches that did not fulfil the users information need. The idea is then that some users will not simply give up in the face of this; instead they will try to reformulate their query in a way that they think will produce some valuable result. This can also be useful for discovering synonyms; seeing how the user continually refines their query, perhaps due to new information gathered during the search session (Azzopardi, 2007), provides an increased understanding of what the intent of the initial search query was. This scenario would be hard to examine without the query evolution.

However, it is not currently possible to determine the limit between two different information needs within one search session - due to the difficulties described in section 2.6.3. Instead, this has to be done via manual inspection by an analyst.

4.5.5 Clicked Results

The authors would argue that one of the most valuable pieces of information when working with Search Analytics is the result links that were visited after a query was made. The authors believe that this should be displayed in conjunction with the selection rate and the error rate-metric for each individual result link, to produce a useful and valuable result, i.e. only clicks on result links that the user found valuable. The whole point of a search engine is to provide access to information, so to know which pieces of information that the users selected is very valuable.

This becomes more apparent when you also take into account the fact that most users do not look further than the first page of results, and that some do not look past the first few result links presented (Stenmark, 2005b). This implies that the search engine's ranking of the result links has a large influence on the user's search experience.

Just as valuable as the results that were clicked are, the results that were not clicked - a result link not being clicked at all is a strong indicator that there is

something wrong with that particular result. Whether it may be that the title or snippet communicates the wrong information, or if the meta-data or content does not match the query, it must be looked into and addressed.

4.5.6 Number of (Unique) Users and Searches

To provide, in our opinion, a more truthful reading of the number of searches, it is important to not count repeat queries during the same session. Two identical queries during the same session is most likely a part of the same information need, and should as such only be counted once.

4.5.7 Search Session Time

This metric is somewhat more ambiguous than for example Average Number of Result Pages Viewed, as a large session time not necessarily means that the user could not find what he was looking for. This is because of the difficulty in defining a session, as described in section 2.6.3. Session time is instead meant to be used more as a way to gauge if enough resources are allocated to maintaining and improving the search service in relation to how much time the users spend using it.

4.5.8 Entry Points for Search

Two categories of pages should present themselves here. Popular pages which have many visitors will because of the high visitor count also have a high (natural) number of search initialisations. The other type is pages that contain unsatisfactory content (information and links). This will lead the user to turn to the search function to find information. To understand what information people were looking for, one could display the most popular search phrases for each of those pages.

4.5.9 Search Facets

Measuring the usage of search facets used can provide information regarding how the users interact with the search service. The popularity can be used to improve the positioning of facets to provide the fastest and best experience for the user - the most popular facet should be the easiest and fastest to reach. Usage can also be of interest, if it is lower than expected then an investigation into why that is should be launched. It might be a user interface issue where the users cannot find , or do not even see, the facets, the naming of the facets is not in line with how

the users think, or the users are not familiar with facets and do not understand what they are for.

4.5.10 Abbreviations and Contextual Language

Match search terms against an organisationally contextualised language thesaurus and highlight any terms that are not present in the thesaurus. The analyst will then have to investigate whether the term actually means something or if it is a misspelling or just gibberish.

4.5.11 Spelling Correction/Auto-Complete

Track how often spelling suggestions and auto-complete is used - do the users find it useful (high usage) or does it fail to provide the right corrections or completions (low usage)? If the data is coupled to a search phrase then more detailed information can be displayed, such as most popular alternative and statistics for this particular search phrase.

4.6 Tasks Identified

Just as identifying any relevant metrics, tasks relevant to Search Analytics was an important part of this project. The line between a metric and a task can sometimes be a bit fuzzy and it can be difficult to determine where to put more complex entities. The authors have tried to make a division by placing actions which perform a more general task with a goal in mind in the tasks-section, and more data-centric entries in the metrics-section. There is, however, some unavoidable overlap (as in the case of Unsatisfied Information Needs).

4.6.1 Understanding Information Needs

The most asked for capability was to be able to see what people are using the search service to look for, which also must be considered to be the most basic and fundamental task of any Search Analytics application.

The authors wish to make an addition to this task, that was not as prominent in the interviews. To really make good and actionable changes to the search service, it is not enough to know simply what people are searching for, one must also understand which pages people find valuable when looking at the results of a query.

In short: what are people looking for, and what they are finding.

4.6.2 Analysing Unsatisfied Information Needs

Just as for understanding what people are looking for, this is also a well known and common task when looking at related literature or analysing the interviews. This task is about finding queries which generated zero or only unsatisfying results and then analysing why that is and what action should be taken in response.

To assist in the analysis, the authors propose that the query evolutions for any unsatisfied query should be displayed. This will help the analyst to better understand the information need of the users.

4.6.3 Creating and Sharing Reports

Stenmark (2005a) points out that the manager of an intranet often is not the one providing the content, creating a disconnect between the content creators and the user. This was also apparent from the interviews. The authors therefore wish to support the communication between the managers and the content creators by allowing the managers to share Search Analytics reports with individual content creators. This could help the content creators to better understand how their work is received and also to make better use of valuable meta-data which can be used to improve the search experience.

The report could be in the form of a URL that opens the Search Analytics tool with the appropriate settings applied where the stakeholder himself can work with the data, or a document presenting the information in static form.

4.6.4 Segmenting and Comparing Data

Within an organisation an analyst is only responsible for the subset of queries that are performed under her area of responsibility. Naturally, there is then a need to be able to segment out that subset from the data.

This is but one example of segmenting. Also relevant is the ability to segment the user base on various properties (location, role, etc.) to better understand the needs of different groups.

Being able to compare results over time can reveal what effects the Search Analytics work has on the search service. Seasonal or other time based patterns might also emerge from this kind of analysis.

Again, the size of the time and the properties segmented on should be chosen depending on the particular organisation and its context (Rosenfeld, 2011).

4.6.5 Grouping Searches

This type of analysis did not appear in the interviews, but can be found in some literature, e.g. (Rosenfeld, 2011).

The analyst may group different searches together to better track what is going on within one specific topic or area. These groups could be user defined or created automatically from e.g. synonym rings, similar to Rosenfeld (2011), or through clustering (Andréen, 2010).

4.6.6 Tracking Changes

This task is something that the authors have not seen explicitly defined, but very much should be a part of any analytics application. By tracking different metrics after a change to the search service (e.g. ranking, key-match, synonym, meta-data) has been made and by presenting them with a clear visualisation of 'before' and 'after' the change. This allows the analysts to check if the change had the desired effect or not and to learn which type of changes have the biggest impact on user behaviour.

4.6.7 Analysing Trending Queries

When a new query, or a query which usually has very low usage numbers, suddenly jumps a lot in usage over a shorter period of time, it is added to the list of trending queries. This being queries that are new in some form, there is a risk that there is not any content that matches this query or that the content is outdated, which is why it is flagged to the analyst to investigate.

This type of task was not found in any literature, but came up in the interviews.

4.7 Prototyping

After having decided on a target user group and a set of tasks and metrics to realise, a process of prototyping different concepts of the design, both in detail and more general in scope, was begun.

The first activity was simply sketching by hand on paper, trying out different ideas for visualising a specific metric or function. At this stage, no idea was too crazy or weird - experimentation was encouraged. The process was a simplified version of the 6-3-5 method, adapted to just two people, where ideas for a topic would be sketched and then handed over to the other person to react to, reflect on and improve upon.

After having agreed upon a general layout of the application, we moved up one step in detail by creating wireframes. This was done in a collaborative fashion with two people on one computer, designing and discussing intermittently. The nature of wireframing allowed the focus to rest on task-flows and general layout and preventing as of yet undesired detail to clutter the design. See appendix C for images of the wireframes.

It was decided that more detailed sketches than what the wireframes could provide were needed. Therefore, the wireframes were translated into more detailed versions, with the chosen colour scheme and more fine details.

When creating the digital sketches (regardless of resolution) there were often discussions about how to design something, and while wireframing is a rather quick and easy way to work, nothing is faster than drawing by hand when something needs explaining or a redesign. A combination of pen and paper and whiteboard was used to support this.

As the sketches approached completion a more structured evaluation in the form of a heuristic evaluation was conducted, where issues regarding language, highlighting of some visual components and consistency within the design were discovered.

In the later part of the sketching process, work had also begun on fetching the needed data from the database. As more and more focus was directed on to the impending implementation of the design, parts of the design had to be altered because of how the data behaved. The biggest change that had to be made was the redesign of the search evolution. The naive assumptions of how the data would look were, although reasonable, wrong; there were a lot fewer common search phrases than expected, which did not fit the initial design. An emergency design session was initiated where the problem and the goal of the design was laid out and addressed.

4.7.1 General Layout

The decision to use a graph to select the time interval and to use a two-pane design to select a query came quite early in the process. What came later was the concept of using tabs for the different tasks. This was done to create a better flow, as one is likely to focus on solving one type of problem at a time. The naming of the tabs is also very deliberate; at first the names were very short and, in a sense, abstract, but after consulting the personas this was changed to be more descriptive to better suit the novice users.

4.7.2 Search Evolution

As mentioned before, the search evolution design has seen some design changes throughout the project, probably the most of all. The goal remained close to the same, to present how people reformulate their search queries when they did not find any information to satisfy their information need. The first concept was a directed graph where the nodes were queries and the edges represented a part of a search session. Common queries were highlighted by increasing the size of the node. After testing the design with more data it became apparent that this design would have resulted in a very busy and cluttered visualisation; the number of nodes and edges simply became too great to be manageable.

4.7.3 Colour Scheme

Considering that the project was conducted together with Findwise AB, it was decided that their colour scheme should be used. They did not have any requirements for how the colours were to be used. The colour scheme used is the following:

Figure 5: The colour scheme used.

This colour scheme worked well as all the colours are viewable by people with colour blindness and there is a good contrast between the colours. The colour orange has continually throughout the design been used to mark components that are important and/or can be interacted with in some way.

The very light grey, almost white, colour has been used as a background colour for most of the components. The darkest grey has been used as a framing colour for the main section that holds all the statistics data. It is also the colour of different artefacts within the design that are interesting but not as important as those that are orange.

The rest of the colours are used to get a good mix and contrast between all the elements within the design.

4.7.4 The Case for Rounded Corners

During the interviews, it became clear that some of the tools that were in use out at the client organisations were considered to be overwhelming and intimidating by novice users. This prompted the authors to try and add an inviting characteristic into their design. Rounded corners have the ability to lighten the cognitive load on the human brain. Thus geometric figures with rounded corners will be 'easier on the eyes' than geometric objects with jagged edges. Rounded

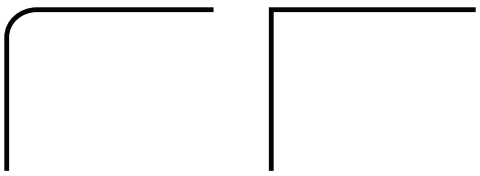


Figure 6: Round versus straight corner.

corners also have the ability to draw the attention of the eyes towards the centre of the geometric figure that they are a part of, while a jagged corner will draw the attention of the eyes outside of the geometric figure. Having rounded corners will make it easier for the user to focus their eyes on different parts of the design, and it will feel easier to look at (UX Movement, 2011).

4.8 User tests

Two user tests were conducted during the project. The first one was conducted with employees at Findwise AB, who can be viewed as expert users. The second test consisted of a group of users who can be considered to be novice users. Both user tests were conducted to reveal flaws in the proposed design. The method used during these user tests was a form of Participatory Design 3.7.

4.8.1 User Tests With Low-Fidelity Prototypes

During these tests a high resolution low-fidelity prototype (paper prototype) was used for the tests a couple of scenarios was constructed. Each scenario was designed to test different parts of the GUI. The tests were conducted in a closed room where only the subject and a moderator were present. The moderator explained how a typical participatory evaluation is conducted and then guided the subject through the test by asking the subject to perform what was described in each scenario. If needed, the moderator prompted the subject to talk out loud. Each test was recorded with video and audio to be analysed at a later time.

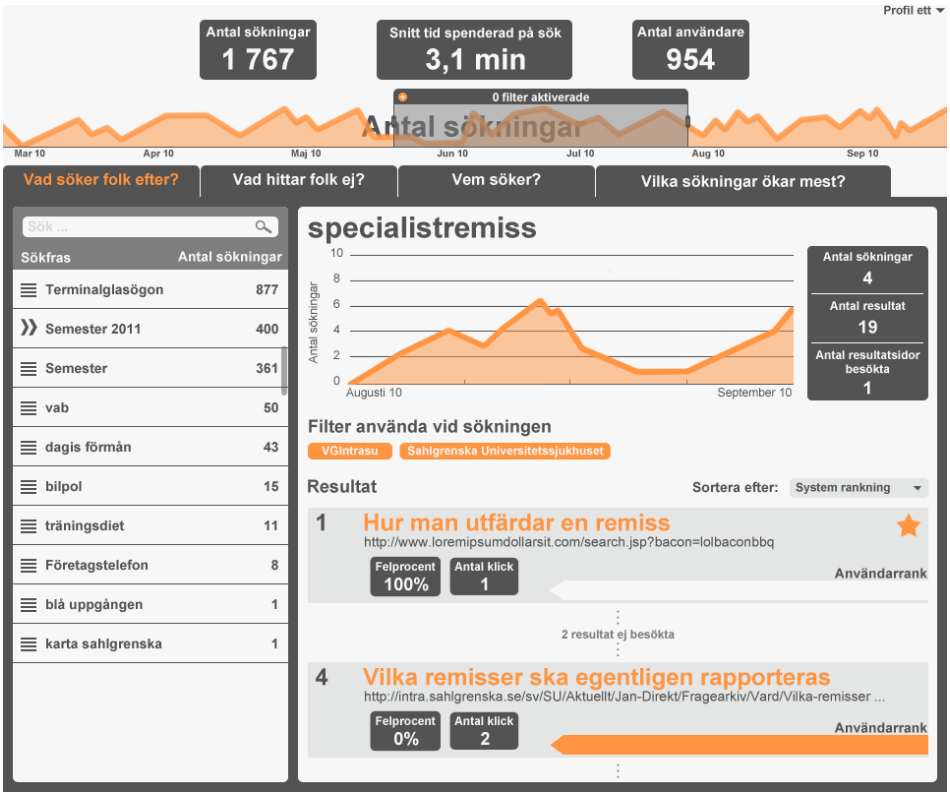


Figure 7: The paper prototype. For larger version see appendix D

Problems revealed

The user tests revealed that how the user was supposed to add filters to control the data presented was unclear. The users had a hard time to actually find the components within the design that controlled this. It was also revealed that the users disliked the disconnect between the date range of the query-specific search history graph and the date range that the user had selected. It was unclear what the icons in the list of search queries were supposed to represent. A small detail regarding the date formatting in the data set selection graph was discovered. Some users interpreted 'Okt 10' as the 10th of October, when it actually represented October 2010.

4.8.2 User Tests With High-Fidelity Prototypes

The same technique, participatory design, was used for these tests. Potential real end-users, that one day might actually be using the system at hand, were used. They are to be considered as novice users with limited experience of the type of

tool that was being tested. Before the test began, the moderator performed a demonstration of how a participatory design test with think-aloud works. The goal of this short demonstration was to allow the test participant to understand how they should act during the test. The tests were conducted under similar conditions as the low-fidelity prototypes. As this test used a high-fidelity prototype, a web application, both the screen and the test subject were recorded on video to capture the maximum amount of information which later could be analysed.

See appendix F for list of findings.

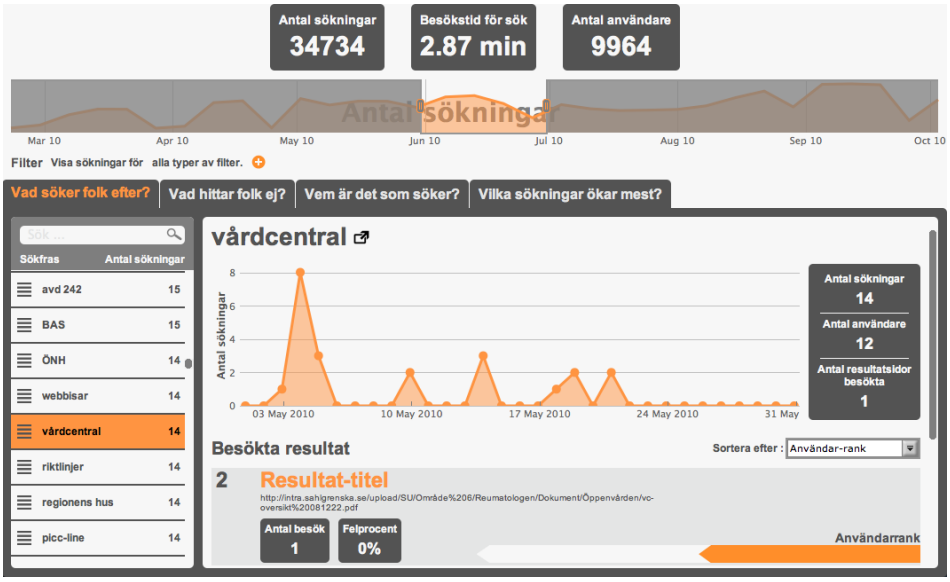


Figure 8: The web application prototype. For other, larger versions see appendix E

Problems revealed

The user tests revealed two main problems: language and understanding. The language used in the application contained both good and bad parts; the good was regarding how the titles of the tabs were phrased, clear and easy to understand and constructed in the form of a question; the bad was that other parts of the application did not follow the same guidelines, specifically pointed to was the drop-down selection menu for sorting the result links, which the users had a very hard time to understand. The phrasing of the tabs might have to be looked at again too, as for example the answer to the 'What Are People Looking For?'-tab (see appendix E) was found by just looking at the left pane with the search phrases, while it was not clear what the information in the right pane was for as the question had been answered already.

The second, big problem with the concept implementation was the lack of information explaining what the different metrics represented. The test users, being novices and possessing little domain knowledge, did not understand what the labels of the different metrics meant. The authors believe that a lot of the metrics could have been explained using a simple technique like tool-tips, which was not included due to technical problems with the implementation that was not solved in time.

A more specific problem was that the view representing the search evolutions (see appendix E) was not graspable by the users. The concept of the search evolution was new to them, and was not communicated well through the design. This lead to the users not knowing what they should expect from, or do with this function and thus they did not understand the design.

A smaller problem encountered was that when the application is first loaded the first tab is automatically selected. This was however not clear to the users. This resulted in that they sometimes got confused when nothing seemed to happen when they selected the already selected tab.

4.9 Report Creation

The data source that was available to the authors consisted of data collected over a period of about ten months, stored in a SQL database server.

To extract metrics from the raw data, a Java-application was used together with several different SQL-statements to create reports containing the processed metrics. The reports created were over fixed time spans (week, month and year), and serialised and stored onto disk when done. The Java-application used parts of a framework created by Findwise AB for creating the reports, but a lot of work had to be put into creating new implementations of existing super-classes and interfaces, and the SQL-statements were created from scratch.

The report creation will not be discussed in more detail as the implementation of it is not the focus of this thesis.

4.10 Implementation Tools

4.10.1 SmartClient

A JavaScript framework called SmartClient[1] had been selected by Findwise AB for another project, and was therefore also chosen to create the concept implemen-

[1] http://www.smartclient.com/

tation. SmartClient uses JavaScript to create the entire GUI, which dynamically generates cross-browser compatible HTML. SmartClient also supplies sending and receiving data using only JavaScript for some databases, or one can use their Java server framework to implement custom data connections.

The appearance of the implementation is controlled using a combination of JavaScript properties set in SmartClient and traditional CSS.

4.10.2 Highcharts

Highcharts[2], a JavaScript library, was used to create line charts for the concept implementation. Small modifications were made to hook in additional functionality together with the rest of the application.

4.10.3 InfoVis

The tree-map visualisation used in the implementation was created using a part of the InfoVis[3] JavaScript visualisation library.

[2]http://www.highcharts.com
[3]http://thejit.org

5 Results

Presented here are the three deliverables that constituted the goal of the thesis.

5.1 User Descriptions

This section will summarise the information gathered during the interviews (see appendix A), divided up into five different sections and also presented as two personas, each representing one of the two primary user groups identified.

A total of seven interviews were made, lasting between 35 minutes and one hour each. Four of the interviewees had in some capacity the overarching responsibility for the content of a website and worked as web-editors. The remaining three had different positions: one had an administrative and management role, one was a project leader and developer and one worked as a consultant.

Note: not all interviewees had worked specifically with Search Analytics, but if not then they had worked with Web Analytics, which share many traits with Search Analytics, and expressed an interest in what Search Analytics may offer to them. Many of their thoughts and experiences where therefore deemed transferable to the subject of Search Analytics.

5.1.1 Work Related

The general theme here was that the user had very little time allocated to perform the Search Analytics tasks. The answers ranged from a few hours a week (the consultant) to once a year (one web-editor), but most had enough time to perform the tasks on a monthly basis. This was due to them having many other responsibilities and tasks to perform in their regular work.

Another commonality that was discovered was that they were the only ones in their department tasked with doing this. There was, however, sometimes a loose network of people from other departments who performed the same tasks to whom one could turn for help or advice.

5.1.2 Stakeholders

Furthermore, it also became clear that the interviewees, who were responsible for looking at the data, were not the only, or at all, content contributors. There were other stakeholders within the organisation who were interested in hearing about how the site was used. However, the other stakeholders were content with receiv-

ing information from the Search Analytics-responsible rather than looking it up themselves . Also, there were often demands from management-level supervisors to get a quarterly or yearly report summarising different metrics regarding their site.

5.1.3 Domain Knowledge

The amount of domain knowledge possessed by the interviewees can be divided into two clear groups. The first group has quite a lot of domain knowledge, they have used Search Analytics for several years, sometimes also in their spare time for personal projects. They find the subject interesting and the information gathered to be of great value, and they are comfortable working with the tools available. The second group has relatively low domain knowledge, and do not spend more time than what they are required to working with Search Analytics. They found the tools somewhat intimidating and overwhelming, and had a hard time to define exactly what it is that they should do with the information.

5.1.4 Organisational Goals

When asking the question of why they have been tasked with using the Web/Search Analytics tool, there was a clear distinction between the organisations who had clear key performance indicators (KPIs), or a specified set of metrics set up to work and measure against, and the other organisations which simply had the somewhat broad high-level goal of 'making the search better' or 'better understand the users and what they are looking for'. We find it troubling that many of the organisation had such loosely and ill-defined goals regarding Search Analytics. Hurst's (2010) process has defined the business metrics or KPIs as the first step, which forms the basis for the rest of the work. This is an area where there is room for improvement.

5.1.5 Attitude

When asked about how they feel about Search Analytics - do they find it valuable, useful, interesting, etc. - there was a general consensus that it is a very interesting and useful subject and that they would like to have more time to spend on it. This was requested because they either considered it to be difficult and time consuming, or that more data and time was needed to be able to perform a more thorough analysis to leverage the great potential of the subject.

5.1.6 Personas

With the data gathered from the interviews, and a healthy dose of make-believe, two personas were created, following the method described in section 3.3 The first persona, Alice, was created to represent a web-editor with limited knowledge of Search Analytics, and little time to spend on it each month. The second persona, Robin, was created to represent a technically competent consultant with great knowledge and big interest in Search Analytics, dealing with several different clients. Please see appendix B for the complete descriptions of the personas.

5.2 Metrics

Here a list of the metrics that the authors believe are relevant to the subject of Search Analytics, presented together with a short description.

Average Number of Queries -
Displays the average number queries that is made per session; indicates how often users reformulate their query.

Average Number of Result Pages Viewed -
Displays how many pages of result links that were viewed on average; a large number means the top results are not good.

Search Session Time -
Displays how much time a user spends on search per session; used to measure the usage of the search service.

Session Activity -
Measures the amount of activity during a session (queries, clicks, etc.); used to measure user engagement.

Number of (Unique) User and Searches -
Shows how much the search service is used.

Error Rate -
The percentage of users who did not find the information they were looking for at this result link page.

Unsatisfied Information Needs -
The queries where the users' information need was not met (no results, no clicked results or only results with a high error rate).

Clicked Results -
The result links that were clicked for a query; gives an understanding of what the users found interesting.

Selection Rate -
A measure of the number of queries which led to a result link being clicked; a low ratio means poor search results were returned.

Top Queries -
Displays the queries making up the head of the Zipf curve; used to highlight the queries with the greatest return on investment potential.

Query Evolution -
Displays how a search initiative has evolved during a session, through query reformulation and/or adding/removing any search filters; used to gain an understanding of the user's true goal.

Click Path -
Displays the navigation-path that the user took after leaving the search result page; used to see where the user found the answer to their query

Result Type -
Used to measure the popularity of key-matches versus organic result links.

Content Type -
Displays the popularity of different types of content (images, text documents, etc.).

Top Result Links -
Lists the most visited pages, regardless of the query which led to the page; these pages should receive extra attention.

User Tracking -
Track other metrics related to specific users; provides better understanding of the users' behaviour.

Devices, Browsers and Plug-ins -
Displays information regarding which devices, browsers and plug-ins that are available to the users of the site; gives an understanding of which techniques that can be used.

Entry Points of Search -
Displays information regarding from which pages search was initialised.

Search Facets -
Displays the popularity of the facets; can be used to enhance the GUI.

Spelling Correction/Auto-Complete -
Displays the usage of these tools; used to gauge their perceived usefulness by the users.

Abbreviations and Contextual Language -
Displays terms and phrases that could be a part of an organisational contextual language; used for finding synonyms.

Search Term Co-occurrence -
Displays the different combinations of search terms; used to find patterns regarding the users' information needs.

Query Length -
Displays information of the average character length and number of terms in a search query; used to gain understanding of the users' behaviour.

5.3 Tasks

5.3.1 Understanding Information Needs

Understanding what the users at a site are looking for is the fundamental task of Search Analytics. A natural extension of this is to look at which pages that the users then have visited, and other types of metrics that can provide some information regarding their behaviour and, ultimately, determine if their information need was fulfilled.

5.3.2 Analysing Unsatisfied Information Needs

To discover gaps in the site content or mismatches in terminology between the site and its users, it is useful to look at searches that did not satisfy the users information need. The analyst then decides whether the query represents something that aligns with the organisations goals and thus should exist on the site, or if it is something that the organisation currently does not do but should add to the site, or if there already is content regarding this, analyse why it is not found, adjusting meta-data or the language of the site.

5.3.3 Creating and Sharing Reports

This task deals with forwarding relevant information about the sites usage to relevant parties. The analyst wishes to share a report detailing certain metrics for a specific time period to some other stakeholder(s). One should be able to create reports at different level of detail, to suit different types of stakeholders.

5.3.4 Segmenting and Comparing Data

To gain deeper and better understanding, the analyst wants to segment the data in different ways. Segmentation is very organisation-specific; depending on the context that the organisation works within, different metrics to segment by will emerge.

Another important task is to compare data across time spans, month-on-month, year-on-year, etc., to see how things have changed and progressed over time. Segmenting by hour can reveal micro-patterns, and larger segments can reveal seasonal patterns in what kind of information is requested.

5.3.5 Grouping Searches

To track what is going on within one specific topic or subject area, the analyst may group different searches together. These groups could be manually defined or created automatically through different techniques (e.g. clustering or synonym lists).

5.3.6 Tracking Changes

To understand how changes to the search service has effected the user experience and behaviour. Can be used at both a general and a detailed level, and may also be used in conjunction with other tasks, such as grouping and segmenting, to focus the tracking on desired areas.

5.3.7 Analysing Trending Queries

Finding and investigating the state of the content related to the currently trending queries, to keep the search service up to date and continuously valuable and useful to the user.

5.4 Concept Implementation Design

5.4.1 Intended User

The intended user of this concept implementation was based on the user description in section 5.1.

5.4.2 Selected Tasks

The implementation was limited due to time constraints to the three tasks identified as the most basic and important to the users and organisations out of the tasks listed in section 5.3. The tasks selected were: *Understanding Information Needs*, *Analysing Unsatisfied Information Needs*, and *Segmenting Searches*.

5.4.3 General Layout and Flow

The application has two main areas: on the top is the input area where the user selects a time period and any filters needed, and on the bottom is the task area

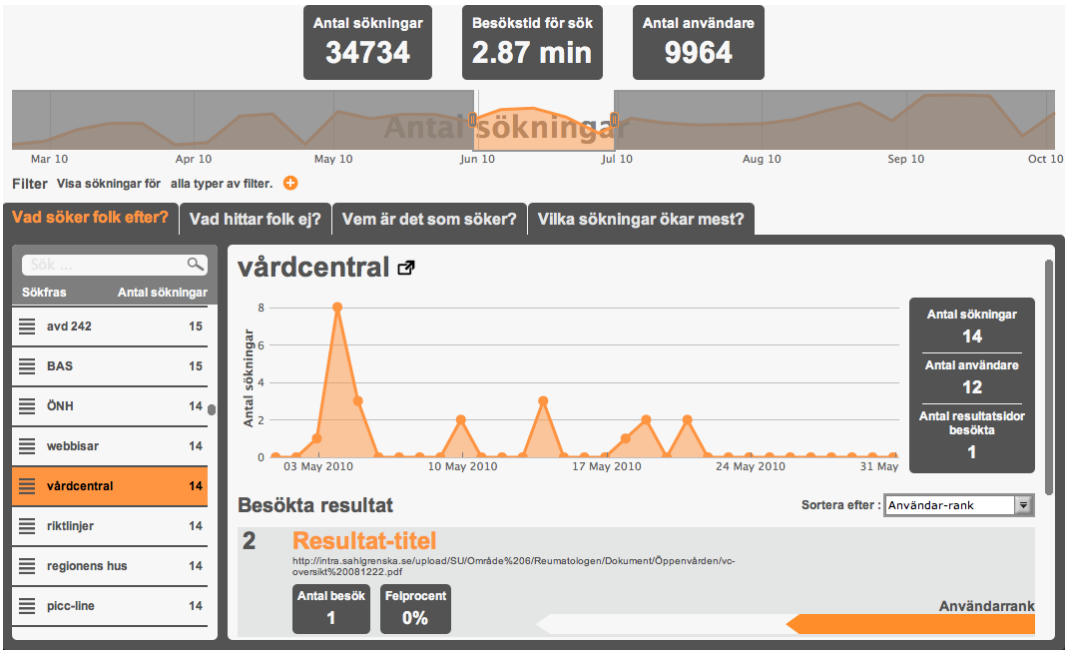


Figure 9: Overview of the concept implementation.

where the Search Analytics data is displayed and the different task are performed.

The standard flow of the application is then:

1. Select a time period
2. Select any filter needed
3. Select a task

Figure 10: Flow of the general layout.

Step 1 and 2 work independently of each other, a change in one of them does not have any effect in the other. The data displayed in step 3 is however dependent on the selections made in step 1 and 2.

The intended flow follows a natural top-to-bottom order, but the order of in which the steps are performed is not critical; one could perform them in any order.

5.4.4 Task Specific Layouts and Flow

The task area is a tab-set with a tab for each task. Each task has an individual flow and layout.

Understanding Information Needs

This task uses a two-pane layout for selecting and viewing information. The flow of this task is:

1. Select a query in the left pane to inspect by either
   (a) selecting a query from the list
   (b) searching for a query
2. Analyse the query data in the right pane by any of
   (a) inspect the search history graph
   (b) inspect the query statistics
   (c) inspect the result links
   (d) sort the result links if needed

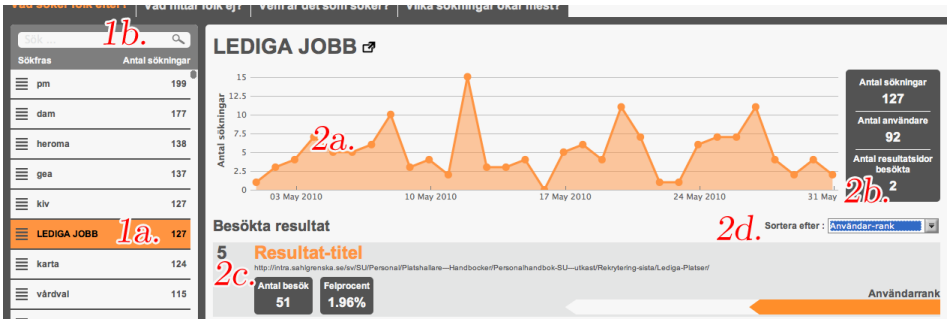


Figure 11: Flow of Understanding Information Needs

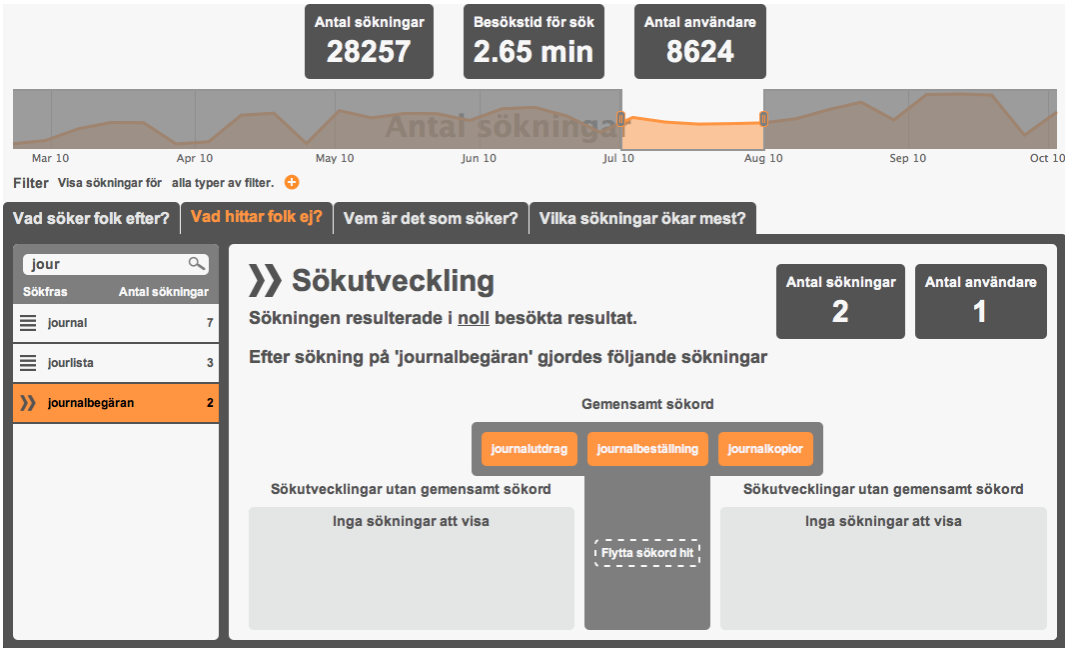


Figure 12: Overview of Unsatisfied Information Needs concept.

This task uses a two-pane layout for selecting and viewing information. This task also branches into two sub-task depending on what type of query is selected. The first sub-task is the same as Step 2 from *Understanding Information Needs*, while the other is a new task. The flow of this task is:

1. Select a query in the left pane to inspect by either
   (a) selecting a query from the list
   (b) searching for a query
2. Analyse the query data in the right pane
   (a) if the query had no (clicked) result links
      i. inspect the query reformulations made where there is no common search phrase
      ii. if any, select a common search phrase
      iii. inspect the query reformulations
   (b) if the query had clicked result links, do any of
      i. inspect the search history graph
      ii. inspect the query statistics
      iii. inspect the result links
      iv. sort the result links if needed

Figure 13: Flow of Unsatisfied Information Needs

Segmenting Searches

This task uses a single tree-map visualisation to illustrate how much different segments within the organisation uses the search service to find information from another segment. The flow for this task is:

1. Select whether to segment by department or role.
2. Gauge the amount of usage after the size of the segment
3. Click a segment to reveal more levels in the hierarchy (if any)

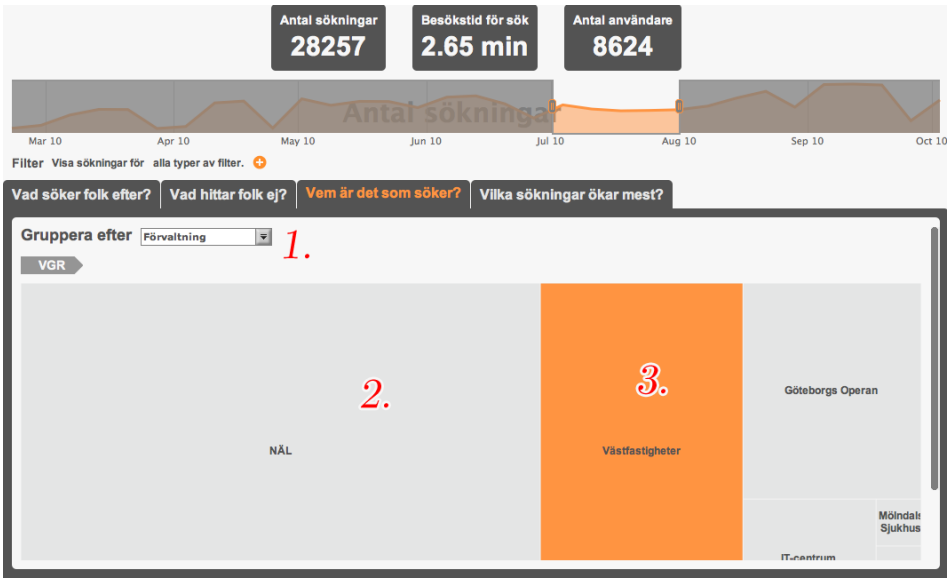


Figure 14: Visualisation of different segments.

5.4.5 Description of GUI Elements

This section will describe the individual GUI elements in more detail.

Result Link Entry

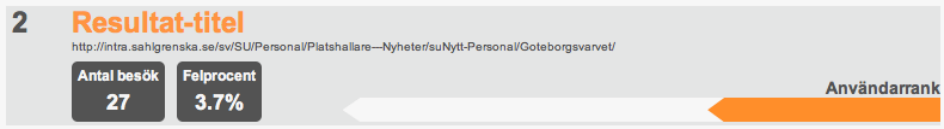


Figure 15: Result link entry.

The result link component is used to display all the data associated with a clicked result link. This is one of the most important components in the application, and is therefore given relatively much space.

The component is designed to mimic the look of a search result from a search service, with title, URL, etc. with additional information from the search service, such as ranking and whether it is an organic or artificial link. Ideally, all information associated to the link should be available in the component; examples of information would be meta-data, the search snippet and a description of how the search service has calculated the rank.

User Rank

Figure 16: User rank graphical component.

The user rank bar, which is a part of a result link entry, is used to present the percentage of good clicks a result link accounts for among all the result links for that particular query. There are two reasons for not simply visualising it as a number. Firstly, a pop-out effect that would draw the users attention was desired. This is achieved through the arrow-shape of the bar and the colour choice. The second reason was to facilitate easy comparison between different result links, which the horizontal alignment of the bar accomplishes this.

Figure 17: Two statistics boxes.

Statistics Box

This component is simply for displaying a numeric statistics with an explanatory label. The design of the component is very restrictive in use of colour, as there will often be many of them in view. While the box may vary in size, the other attributes stay the same across the whole application; it should be easy to recognise this component, and understand that it is meant to communicate some form of statistics.

Time Span Selector

Figure 18: Combined range selector and area graph.

The application is built on the concept of working over different time spans. This component can be used to quickly change the selected time span as well as the size of it. This is done via dragging either the whole selected span, or by moving the start or end bars.

The background of the selector is made up of a graph of how the number of searches has varied over time. There is also a hint-text overlaid behind the graph, giving the users first a hint of how to use the selector and then a description of what the graph represents.

Filters

The graphical components to add a filter is first limited to a simple icon. When clicked, it reveals two combo-boxes (drop-down-boxes with the ability to search) which is used to select the appropriate filter. These combo-boxes uses in-line validation on the input. A filter is added by clicking the button next to the

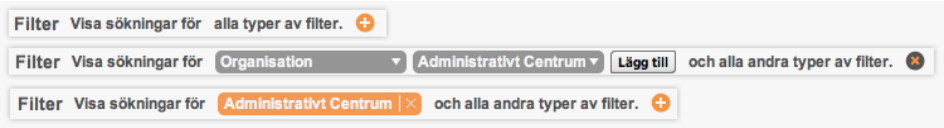


Figure 19: The three different states of the filter component.

combo-boxes, or cancelled by pressing the cancel-icon. When added, the combo-boxes are replaced with a component containing a label with the inputs from the combo-boxes. The filter is directly viewable and can be easily removed by pressing its remove-icon.

Query Pane

When the task involves displaying information associated with a specific query, this component is used to list the queries available to choose from.

Each entry in the list has three components: an icon representing the type of information that will be displayed, a text-label of the query phrase, and a label with the number of searches made with this query phrase. The background of the entry is white if not selected, orange when selected, and a lighter shade of orange when hovered over. There is also a tool-tip that appears when hovering over an entry for an extended period of time, which contains information about the number of results the query generated.

At the top of the query pane there is a search field which can be used to search within the list of queries.

Figure 20: The list of performed search queries.

Analysing Unsatisfied Information Needs

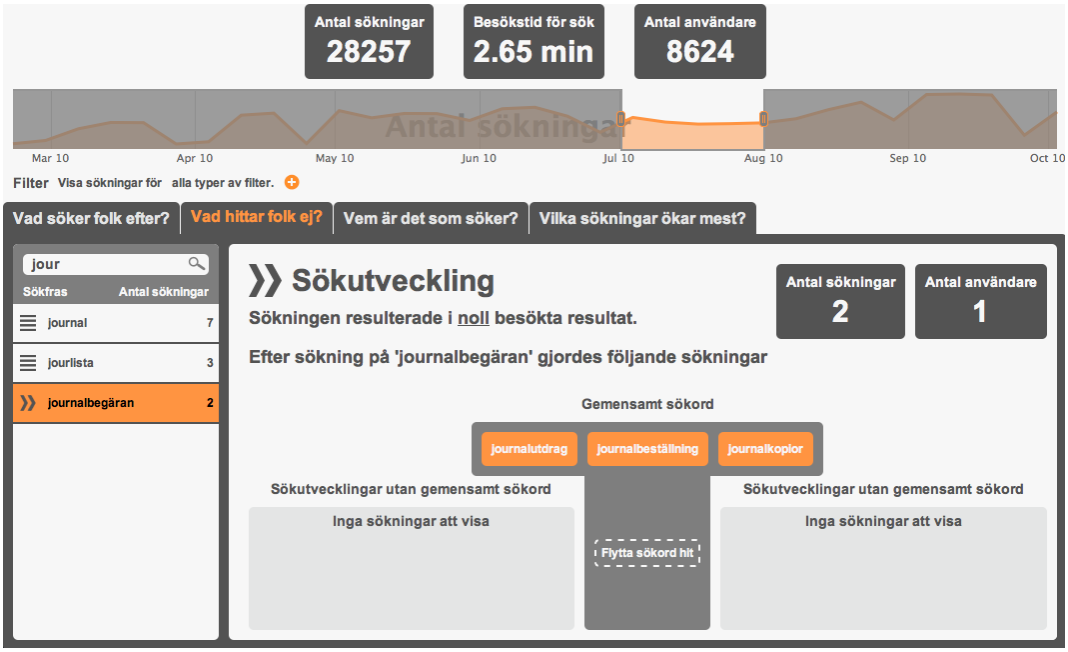


Figure 21: Overview of the concept implementation. More images can be found in appendix E

This is a compound component, consisting of two lists detailing search sessions

and an area for selecting a common search phrase (if available). A quick recap of the purpose: when a query resulted in no results or no results visited by the user, show the query reformulations that the users have made to get some results that they find valuable. Any common search phrases among the query reformulations are found and displayed to the user.

The initial state of the component is to show any search evolutions which had no search phrase in common. This is communicated through the title displayed above each of the lists. The search evolution is split between the lists to conform to the behaviour of when a common search phrase is selected. To select a search phrase, the user performs a click-and-drag action on the search phrase component, moving it to the drop area below. The formulation of the label in the drop area, *Move search phrase here*, and the out-lined look of a box is used to indicate that click-and-drag is to be used to move the component there. After doing that, the lists now show the search phrases entered before (the list to the left) and after (the list to the right) the common search phrase, to allow the analyst to better understand the users goals. Clicking an entry in one of the lists will highlight it in orange in both lists, which makes it easier to follow the whole query evolution. The same is true for scrolling; scrolling in one list also scrolls the other.

Understanding Information Needs

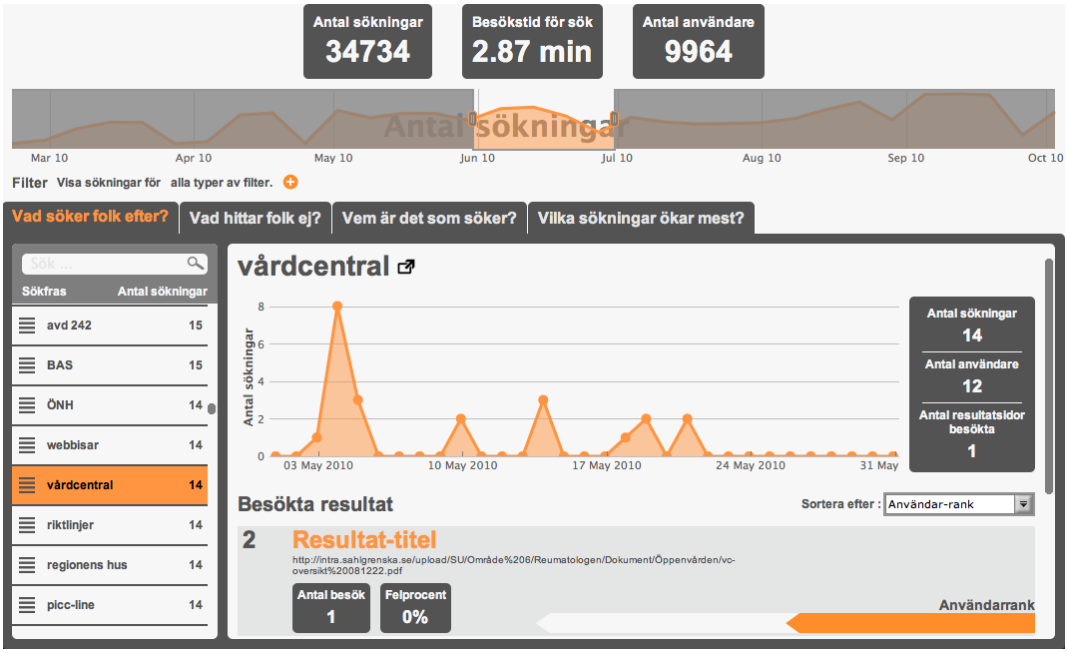


Figure 22: Overview of the concept implementation. More images can be found in appendix E

Another compound component. This one consists of a title, a graph, a set of statistics boxes and a list of result links.

The title is placed at the top, and also works as a link to the query in the search service. This is indicated through the icon placed to the right of the title, and through underlining the title when hovered.

The graph displays the usage of the selected search phrase for the time period selected. This allows the analyst to find any interesting patterns regarding usage. The resolution of the graph depends on the size of the time period selected; a week provides a graph-resolution of one point per day, as does a month, while a year gives a resolution of one point per week.

The list of result links has a drop-down menu for sorting the list in a number of ways, each chosen to better understand something. The default sorting is according to the ranking that the search service provides which can be used to see if the top results are good and relevant. Sorting by user rank brings forward the most valuable (according to the users) result. Sorting by number of clicks reveals what the users think is the most valuable result, before they have visited it, and the error rate shows which results that the users did not find valuable. Each of the different sort-options have the same name as the labels for the data on the result link.

Segmenting Searches

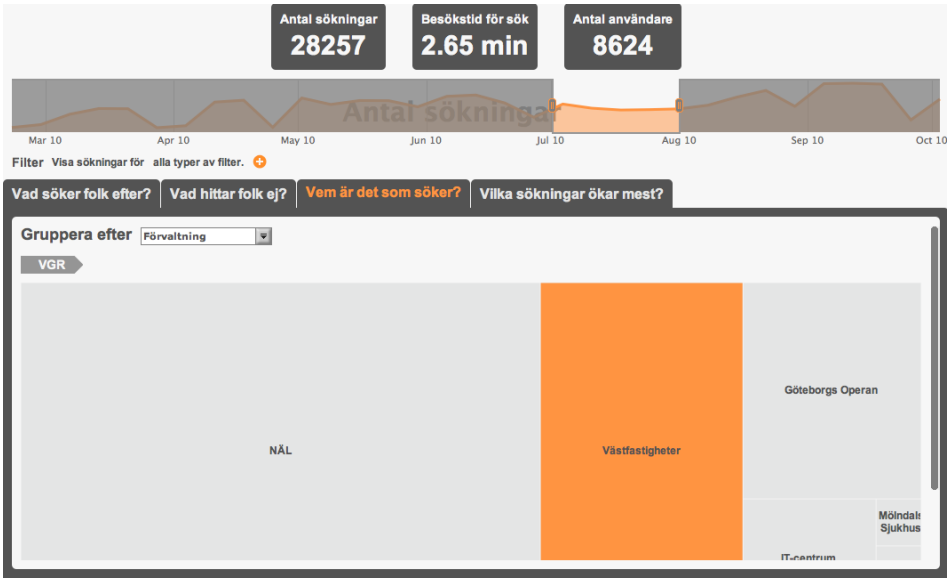


Figure 23: Overview of the concept implementation. More images can be found in appendix E

A compound component with four parts, where a tree-map graph is the central part, designed to visualise how much different groupings use the search service.

The first component is a drop-down menu with which the user can select on which attribute the tree-map should group the data. Below that is the bread-crumb trail of the hierarchy that the user has progressed through in the tree-map. Each section is clickable and takes the user back to that level. At the bottom, below the tree-map, is a contextual-link which takes the user to the What are people looking for-tab, and applies the necessary filters. The different nodes/parts of the tree-map graph represents different groupings, and the size of the node represents the amount of queries that group makes. Clicking a node reveals any sub-groups, again with size representing the number of queries made.

6 Discussion

6.1 Validity of the User Descriptions

Considering the size and distribution of the interview group, the authors would hesitate to draw broad and general conclusions. However, we do feel that the results obtained do provide a good and truthful image of that particular user group when it comes to their working-situation and their level of domain-knowledge.

One of the initial main goals of the interviews was to capture requirements regarding organisational needs and requests for additional metrics and functions. This yielded a few good responses, but far from as many as hoped. We attribute this to the relatively newness of the subject of Search Analytics, and the shortness of the interviews; had it been possible to conduct deeper and more probing interviews with more people at different positions within the organisation, coupled with deeper studies of the organisations goals and practices, a better understanding of their needs could have been attained, and more informed questions could have been asked. This need for extensive interviews was, however, not included in the project time plan and was therefore not pursued.

Late in the project, when the concept implementation was finished, a demo session with a customer of Findwise AB was held. The demo session generated an interesting discussion where many ideas and comments were voiced. The authors believe that a workshop with several stakeholders could be a good next step, where the concept implementation can be used as a mediating tool to catalyse discussion.

6.2 Tasks and Metrics

The goal with the task and metrics lists was to compile lists that would only contain entries that had been deemed to be relevant to the users and organisations after having studied their needs and requirements that had been gathered from the interviews. However, as we felt that the results did not exploit the wide range of possibilities that Search Analytics can provide, we elected to expand the list with additional tasks and metrics found during the literature study and the comparative review process.

The list is even after this expansion, however, not to be considered as a complete or full representation of all of the types of tasks and metrics which are related to Search Analytics. Such an undertaking would require more time than this project would allow. Also worth considering is that, following Hurst's process for Search Analytics, the relevant tasks and metrics is dependent on the organisation that is to use them. One size does most likely not fit all; it will be too big for

some, too small for some and not the right thing for some. The idea of creating a "definitive" list of tasks and metrics is therefore perhaps not achievable and not worth pursuing.

6.3 The Concept Implementation

Time constraints required us to select a small sub-set of the task and metrics to implement. This has undoubtedly had an effect on the design and layout of the application, and it is uncertain whether the current design would allow for more data and new functions to be added and still be usable. It is certainly possible to expand the number of tasks that can be performed by simply adding another tab, or add more metrics to the content of a tab, but this requires the task to fit into the current structure, which may or may not work.

6.3.1 SmartClient, HighCharts and InfoVis

Using SmartClient to implement the concept design had both pros and cons. It was relatively easy to learn the basics of SmartClient, and we got a first version running with dummy data quite quickly. However, to get everything looking exactly as wanted was sometimes a bit cumbersome due to how SmartClient works, as the visuals were controlled by three separate files: in-line in the JavaScript code, a CSS-file and a "skin"-file which defined how SmartClient visualised some built-in components. There was also some problems related to integrating SmartClient and the other JavaScript libraries used, which required less than elegant solutions due to bugs in SmartClient.

6.3.2 Regarding the Second User Test

There were several problems revealed during the second user test, detailed in section 4.8.2, but it should be noted that due to time constraints no changes were made to the concept implementation after the second user test.

6.4 Process

If the authors look for any learnings from this project, one of the major ones has to do with time estimation. The time allocated for some of the phases of the project was not adequate, especially concerning the idea-generation and paper prototyping phases. This led to a stressful period where it felt like the time was running out, and resulted in less than ideal prototyping phase where too few design alternatives were created and evaluated.

The difficulty of finding interview subjects was also something that the authors had not anticipated beforehand, most likely due to inexperience in dealing with real users and their limited time schedules.

The need to allot more time to creating reports from the raw data than what was initially wanted and planned for reduced the time left for other tasks, which impacted the quality of the design somewhat, and the quality of the implementation more so. The scope of the report creation was underestimated, due to bad planning and led to it taking longer than what perhaps would have been necessary.

6.5 Future Work

This thesis has dealt with the very basics of requirements gathering and realisation regarding Search Analytics. The authors have a few recommendations for where future work is needed.

6.5.1 Methodology for Requirements Elicitation

To gain a better understanding of the needs of potential users and organisations, a detailed methodology for identifying and analysing the needs of relevant stakeholders should be created. This is needed because knowledge regarding Search Analytics is not very widespread, and to facilitate real, actionable change with positive results many different stakeholders need to be involved in the design process. This would also be a good position from which the need for educating the users in the field of Search Analytics could be determined. If the level of knowledge is low, then a strategy for working with Search Analytics should be developed together with the target organisation.

6.5.2 Expanded Data Collection

The authors also believe that there is room for improvement in the area of data collection as well - the more the data, the (potentially) better analysis. Viewing Search Analytics as something that should only concern itself with what goes on within the confines of the search service is, in our view, limiting the potential of Search Analytics. We believe that the goal of Search Analytics should be to understand how search has been used to fulfil an information need. This will most likely require information about actions that took place outside of the search service. In addition, more information from the search service could be captured as well, e.g. meta-data and snippet for each result link.

6.5.3 Integration with Related Tools

Considering the concept implementation, there are many aspects that would benefit from more work. Besides improving the existing design from the feedback gathered from the user tests, designing concepts of the identified tasks that were not implemented here would be a good start. As it stands now, the implementation has only the most bare minimum of functions that can be expected of a Search Analytics application. The number of functions available can, in our view, provide value to the users and their organisations, but to create a really useful application more functions need to be implemented, and the application needs to be integrated with any other applications which are used to actually modify the search service. That aspect was deliberately left out of the scope of this project, due to it being a large project in itself, but it is something that needs to be looked at carefully.

7 References

Agichtein, E., Brill, E. and Dumais, S. (2006) Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of SIGIR 2006*; 6-11 August 2006, Seattle. p. 19-26.

Andréen, E. (2010) *Assessing Findability.* Master of Science Thesis from Royal Institute of Technology School of Computer Science and Communication, Stockholm, Sweden.

Azzopardi, L. (2007) Position Paper: Towards Evaluating the User Experience of Interactive Information Access Systems. In *SIGRI WISI '07 Workshop*; 27 July 2007, Amsterdam. pp. 60-64

Boykin, J. (2011) Google Panda Update - Short Clicks and Long Clicks / Pogosticking. *Internet Marketing Ninjas.* `http://www.internetmarketingninjas.com/blog/google/short-clicks-long-clicks-pogosticking/` (2011-09-26)

Buzydlowski J., White, H. D., Lin, X. (2002) Term co-occurrence analysis as an interface for digital libraries. In *Visual Interfaces to Digital Libraries.* ed. C.C. K. Börner, pp. 133-144. Heidelberg: Springer Verlag.

Chang, Y., Lim, Y., and Stolterman, E. (2008) Personas: From Theory to Practices. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*; 20-22 October 2008, Lund. p. 439-442.

Chen J., Zheng T., Thorne W., Zaiane O. R., Goebel R. (2007). Visual data mining of web navigational data. In *Proceedings of the 11th International Conference on Information Visualization (IV'07)*, 4-6 July 2007, Zurich. p. 649-656.

comScore, Inc. (2010) comScore Reports Global Search Market Growth of 46 Percent in 2009. *comScore.* `http://www.comscore.com/Press_Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_Percent_in_2009` (2011-10-05)

Conyer, M. (1995) User and usability testing-How it should be undertaken? *Australian Journal of Educational Technology*, vol. 11, no. 2, pp. 38-51.

EMC Corporation (2011) New Study Forecasts Explosive Growth Of The Digital Universe; Spotlights Worldwide Phenomenon Of "Digital Shadow". *EMC.* `http://www.emc.com/about/news/press/2008/20080311-01.htm` (2011-10-05)

Fagin, R., Kumar, R., McCurley, K. S., Novak, J., Sivakumar, D., Tomlin, J. A. and Williamson, D. P. (2003). Searching the workplace web. In *Proceedings of the 12th international conference on World Wide Web (WWW '03)*; 20-24 May

2003, Budapest. p. 366-375.

Graf, E., Macdonald, C. and Ounis, I. (2007) Clickthrough based measures of search engine performance. In *SIGRI WISI '07 Workshop*; 27 July 2007, Amsterdam. pp. 63-67

Holzinger, A. (2005), Usability Engineering Methods for Software Developers. *Communications of the ACM - Interaction design and children*, vol. 48, no. 1, pp. 71-74.

Hurst, M.(2010) Search ROI's missing element: Search Analytics. *Journal of Digital Asset Management*, vol. 6, pp. 327-331.

Jackson, S. (2009) *Cult of Analytics.* Burlington, MA, USA: Butterworth-Heinemann.

Jansen, B. J., Spink, A. and Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*, vol. 36, no. 2, p. 207-227.

Jansen, B. J (2006) Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research*, vol. 28, pp. 407-432.

Jansen, B. J (2009) Understanding User-Web Interactions via Web Analytics. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 1, no. 1, pp. 1-102.

Karlsson M., Kaulio M., Hampf J., Sperling L. (1998) Eliciting Customer Requirements: Product representations as mediating objects in focus group interviews. Presented at *5th International Product Development Management Conference*; 25-26 May 1998, Como, Italy.

Kaushik, A. (2006) The 10 / 90 Rule for Magnificent Web Analytics Success. *Occam's Razor by Avinash Kaushik.* `http://www.kaushik.net/avinash/2006/05/the-10-90-rule-for-magnificient-web-analytics-success.html` (2011-09-26)

Kaushik, A. (2010) *Web Analytics 2.0.* Indianapolis: Wiley Publishing, Inc.

Kawakita, J. (1982) *The original KJ-method.* Tokyo: Kawakita Re-search Institute.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J. and Melançon, G. (2008) Visual Analytics: Definition, Process, and Challenges Information Visualization. In *Information Visualization.* ed. Kerren, A., Stasko, J.T., Fekete, J.-D., North, C.J., vol. 4950, pp. 154-175. Heidelberg: Springer Verlag.

Lewis, A. A. (2010) *Enterprise Users and Web Search Behavior*. Masters Thesis from University of Tennessee, Knoxville.

Liu, C., White, R. W. and Dumais, S. (2010) Understanding web browsing behaviors through weibull analysis of dwell time. In *Proceedings of SIGIR 2010*; 19-23 July 2010, Geneva. p. 379-386.

Maguire, M. (2001) Methods to support human-centred design. *International Journal of Human-Computer Studies*, vol. 55, no. 4, pp. 587-634.

Morville, P. and Callendar, J. (2010) *Search Patterns*. Sebastopol: O'Reilly Media, Inc.

Nielsen, J. (1994) *Usability Engineering*. San Francisco: Morgan Kaufmann.

Nielsen, J. (2002) Intranet Usability: The Trillion-Dollar Question. *useit.com: Jakob Nielsen's Website*. `http://www.useit.com/alertbox/intranet-usability-1st-study.html` (2011-04-26)

Nielsen, J. (2006) Data Visualization of Web Stats: Logarithmic Charts and the Drooping Tail. *useit.com: Jakob Nielsen's Website*. `http://www.useit.com/alertbox/visualizing-traffic-analysis.html` (2011-09-26)

Oosthuizen, C., Wesson, J., Cilliers, C. (2006) Visual Web Mining of Organizational Web Sites. In *Proceedings of the International Conference Information Visualization (IV'06)*; 5-7 July 2006, London. p. 395-401.

Peterson, E. T (2004) *Web Analytics Demystified*. Portland: Celilo Group Media.

Rohrbach, B. (1969) Kreativ nach Regeln - Methode 635, eine neue Technik zum Lösen von Problemen. *Absatzwirtschaft*, vol. 12, p.73-75.

Sharp, H., Preece, J. and Rogers, Y. (2007) *Interaction Design: Beyond Human-Computer Interaction*. Second edition. Chichester, UK: John Wiley & Sons Ltd.

Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. (2001) Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, vol. 52, no. 3, p. 226-234.

Stenmark, D. (2005a) How Intranets Differ From The Web: Organizational Culture's Effect on Technology. In *Proceedings of ECIS2005*, 26-28 May 2005, Regensburg.

Stenmark, D. (2005b). Searching the intranet: Corporate users and their queries. In *Proceedings of ASIS&T 2005*, October 28-November 2, Charlotte, North Carolina.

Stenmark, D. (2005c) One week with a corporate search engine: A time-based analysis of intranet information seeking. In *Proceedings of the Eleventh Americas Conference on Information Systems*; 11-14 August 2005, Omaha. p. 2306-2316

Tortolano, G. (2009) Persona Template. *Super User Friendly.* `https://sites.google.com/site/superuserfriendly/templates/persona-template`. (2011-10-07).

Tufte, E. R. (1990) *Envisioning Information.* Cheshire: Graphics Press.

UX Movement (2011) Why Rounded Corners are Easier on the Eyes. *UX Momvement.*`http://uxmovement.com/thinking/why-rounded-corners-are-easier-on-the-eyes/` (2011-10-06)

Wang, P., Berry, M. and Yang, Y. (2003) Mining longitudinal web queries: trends and patterns. *Journal of the American Society for Information Science and Technology*, vol. 54, no. 8, pp. 743-758.

Ware, C. (2008) *Visual Thinking for Design.* Burlington: Morgan Kaufmann (imprint of Elsevier).

Web Analytics Association (2008) Web Analytics Definitions – Draft for Public Comment 9/22/2008. *Web Analytics Association.*`http://www.webanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitions.pdf` (2011-05-12)

8 Appendices

A Interview Transcriptions

Anonymized Interviews

Interviewee 1

Date: 2011-05-04

Who is she?

Has studied informatics and health care administration. Has worked for the organisation for the last three years, and has spent all of her time working within the web services team with support for the CMS and with web and search analytics.

What are the conditions under which she works?

Spends about 60% of her time on web and search analytics tasks. A lot of time is spent on incorporating more departments into the search index and search infrastructure, and on support for web and search analytics and on search engine optimizations. The need for this has continuously increased over the last few years.

How does the company work with Search Analytics?

It is a very large organisation which have employees working with a wide variety of subjects, which in turn mean that there are very different needs across the organisation. This has led to that each branch of the organisation is responsible for looking at their own data (often interested in web analytics-data).

What are their goals with Search Analytics?

The company has a vision of their search tools being “One window towards the information”. What this means is that they want to have one search tool for the entire organisation instead of every branch has their own. This is to make it easier for their users, as they won’t have to remember where to search for different types of information.
They provide search for users both within and outside the organisation.
They want to become independent when it comes to data collection, and avoid the collected data to be stored on someone else’s server where it could be used by others than themselves.
However these weren’t outspoken goals that the organisation had set, it was rather more a vision.

How do they serve these goals today?

Since they don’t have any clear goals within the organisation, it is hard to define how well they serve these goals today.
She expressed difficulty in serving such a broad user-base; there will always be compromises between different departments and their needs.

Which tasks does she perform?

She responds to requests for access to the statistical data - creates accounts and give access to different users so that they can use the different tools available. She also assists users in

finding and understanding the data that they need; explaining what different metrics mean, where to find certain data.
She does look at the analytics data for the search, mainly zero-hits. She does not, however, address the problem herself, that is delegated to the department of information. She provides the tools, others are responsible for the content.
To understand what different abbreviations mean she conducts a small investigation, talking to colleagues and searching the intranet to see whom might be responsible for this.

Who else work with search analytics within the organisation?

A few colleagues are available for support, but mostly her responsibility. They all have different areas of responsibility/expertise.
Her job is not to look at and analyse the data - that is up to the 'information master' of each department. Some are more interested and knowledgeable than others; there exist just a few experts that really understand the tools, but the majority are novices. Interest has increased with time, requests for more detailed information have appeared. Most of them don't have that much time to spend on analytics - their job is mainly to produce content and creating the web pages - and only do it when they have some time to spare, or need to do it for a report. But they do see the value of having the analytics data.
Google Analytics is slightly overwhelming for many, and this leads to them not going past the basic level of functionality when using the program. But there are examples of people asking for help in learning and understanding the tool - requesting a list of the 'Top 5 Features'.

The Info-Master

The job of looking at the data is delegated to the info-masters of each departments. It is their job to compile and analyse the data, and then present it to whomever it may concern.
They are responsible for running and maintaining their site, and also the main responsibility for looking at the web/search analytics and improve their site. It is also they that can add goal/ action tracking / conversion rate criteria to the site, but it is not done today as far as she knows.

The Most Valuable Data

Had a hard time to answer - "everything is important". Concluded that search terms that generate zero-results are the most valuable data. Can point out content that the site is missing, and keywords that should lead to content but for some reason does not.

Thoughts about the data presentation of today

The internal GUI used today has a very poor visualization. It only shows the data in table format. Many of the users that look at the statistics want to be able to create reports from the data. Today they use two other systems besides their own custom made tool (made by Findwise) to view the data: Google Analytics and AWStats. AWStats will soon be deprecated and will be replaced by Piwik. They no longer want to use AWStats as it's hard to understand its data visualizations (bar charts and tables).
Overall when they look at the statistics they are more interested in seeing trends rather than exact numbers. Largely because search statistics isn't viewed as an exact representation of how the search is used. Trends over longer amounts of time (month, quarter, year) is often used.

What would she like to be able to do that she can't do today?

She would like to be able to have more time to meet the users as to collect more qualitative data.
However this is hard to accomplish today. Other than that, there is nothing that she specifically feel that she is missing.

Requests for features

People has asked for more metrics to track (no details given) and more ways to compile and present data in the form of a report - customizable for each department.
Worth noting is that they track a lot of information today, but just a small part of it is displayed today (this is for their custom internal tool).

Tracking Changes

It is up to each department to track any changes that they make; she is not aware of anyone doing it today. She herself does not really track any changes that are made after she has handed off a case to the information department. She think that she probably should be better at doing this, but the continuous day-to-day work leaves little room for it.

Key Matches

Any user is able to add key matches (sponsored results) to any search.

Overall feeling

Analytics has a huge potential for improving the web experience - but she feels that there is not enough time to do as much as she would want to.

Interviewee 2

2011-05-16

Who is she?

Works as a consultant contracted by other companies to perform certain tasks, among them search analytics.

What are the conditions under which she works?

She doesn't work with this full time. The amount of time she can dedicate to search analytics is very much dependent on what other tasks she must perform at the same time - some periods she might only have time one or two times a month, others several times per week. Every quarter they have a follow up where they compare the current quarter with the previous one. All the data that she works with are tables in an excel file.

How does the company work with Search Analytics?

One of her clients have been working actively with it since 2007, which she considers to be fairly long.

What are their goals with Search Analytics?

Their goal is to better understand their users and what they are looking for. This is to lead to better business values regarding external users, and higher efficiency for internal users.
One client also has a KPI stating that the top 100 search terms are to be “managed”, meaning that the search result for these terms are to be reviewed and their ranking adjusted id needed, that was tracked and measured. This is the only thing that management is interested in.
Search has become very important to how the client enables its users to reconnect to data.
Her two biggest clients are rather large organisations, so there are selected groups that are prioritized; everyone cannot have the best experience, compromise is needed.

How do they serve these goals today?

They work with “boosted” keywords. In other words keywords which are weighted and they can manually increase the weight of a keyword. They put a lot of effort into examining the keywords that has zero-hits.
Currently they work on two levels. First is the general level, which is department or group based - here they can see how the search is used overall. The second level is more detailed where “top” or “heavy” users are studied, in part to assist them since they use the service much and in part to gather information that could be applicable to the general user as well.

Which tasks does she perform?

She looks at the top 100-500 search terms used and see if they are managed or not. If they are not, she conducts a small investigation to find the best links. This can include contacting the department that is in charge of that area, if they need to change/add new data to their site.
She also look at dead links and qualitative aspects of the information available.

Who else work with search analytics within the organization?

The Global Content Manager receives reports regarding the KPI, but nothing else. Content managers may receive instructions to add or change content based on the search analytics data, but they don’t request the information themselves.

What would she like to be able to do that she can’t do today?

They are in the process of creating a new process where they are to manage a set percentage of the queries instead of the top 100 or so. This is to capture more of the “long-tail” of the searches. There are also plans to delegate the responsibility to manage search terms throughout the organisation, but it is hard to find people with time to do this.
There is also a desire to capture more qualitative information, to better know if they users found what they were looking for and were satisfied with the quality of the information.
Another request is to be able to dig deeper into the data and slice it up into even smaller pieces. In other words be able to group on for example ‘role’, which would require stricter guidelines and rules for setting such information.
She would like to be more proactive and perform changes before it’s to late. Today they often do the changes after it’s noticed that a change was needed. Coupled with this is the ability to see what impact a change have/had.
The change in the most frequently used keywords can change rapidly in extreme cases. One such extreme case was the swine flu epidemic of 2010. During a short period of time the frequency of words regarding swine flu increase rapidly. However, no information was available for these keywords, and if the statistics isn’t checked regularly it could cause a problem since many searches would return zero results and people not finding the information they needed which could cause unnecessary discomfort.

Thoughts about the data presentation of today

The tool she is using today is simply an Excel-sheet of tables with information. She recognizes that it is not ideal to work with, and that it is pretty cumbersome, but it is enough to do what she has to do.
However, a good graphical interface so that she would no longer have to work in Excel would be highly welcomed. The possibility to create custom reports that makes the data more abstract is also welcomed. She believes this would make it more pleasant to work with and it would be easier to communicate the data and its value to others.

Time span

Currently they compare this quarters data with the previous quarter. There have been instances of looking at yearly statistics to gather data for future projects.

Tracking changes

Changes are not tracked today, although they should be. The problem is lack of data and tools. Not even they number of clicks on the different result-pages are tracked.

Overall feeling

There is value in search analytics, but they should do much more than what they do today.

Interviewee 3

2011-05-16

Who is he and what does he do?

He sees himself as a service provider - providing search services for different parts of the company. His responsibilities are to make sure that the product works, manage the team that implements the search solutions and he is also responsible for managing the funding for the projects.

How does the company work with Search Analytics?

They have not used search analytics much before, they were not satisfied with the tools available. It was too hard to get good reports from the previous systems (Google Search Appliance and Microsoft SharePoint Server Search).
The company has many different clients that they serve with search solutions.
One important IT-resource within the company is the teamplace-site. It has been incorporated into the general enterprise search. The problem with the teamplace is that there are no strict rules about meta-tagging the data that is added there - right now it is the users that set the rules.

What are their goals with Search Analytics?

The companies focus on search have increased significantly the past years. They are currently in the implementation phase of a new search solution (Microsoft FAST Search Server) where

they will be able to log searches and actually have good data to analyze.
The ultimate end-goal is the somewhat broad “make the search better”. More detailed goals are harder to define since they don’t actively work with search analytics yet.
They also want to gather more qualitative information regarding the search service and the documents therein.

How do they serve these goals today?

They do not really work with search analytics actively today, but are in the process of implementing a new search service with analytics capability.
In their current stage of implementation, one of the big additions to the enterprise search is the gathering of several information source into one search service, and the ability to filter between the sources when searching.

Requested Functions

While they do not use search analytics at the moment, he mentioned what they have planned to implement first. This included top-searches, zero-hit searches, time spent on a page, exit-page and clicked results. To filter on time-span is required. Some of these things have been requested, and some are things that he think others would find useful.
While they currently do not work a lot with search analytics there is an interest in what search analytics can provide. To better understand who is searching and what they are searching for is of great interest. Also, to supply reports tailored to different needs within the organization is wanted.
The most sought after feature is to be able to group searches from different parts of the organization (grouping on role, team, department, company, location, language, web-site, etc.). Since the needs may differ greatly between the companies within the Volvo group, it would be very beneficial to be able to look at searches from only one company or section within the group. This is not possible today, but very much wanted. There is also a fear that some organizations will use the enterprise search function a lot more than other parts of the organization, which could skew the statistics in a bad way, marginalizing the smaller organizations needs.
He also talked about separating some more detailed information from the regular enterprise search among more unstructured data. In this case it was product development-related information that should be indexed in a separate search service. This is in part to make it easier for the users to find the right information, and in part (we assume) to protect valuable information. Each company has a business management system, which is not currently tied into the enterprise search, which is not good. To find that information you must know where it exists.
He also see auto-completion as a valuable asset in search, so tracking that can be valuable as well.
To sum up, they want to start slow with the basics, and get them right. When that is done they can move on and see what further functionality/statistics they need.

Who else work with search analytics within the organization?

The Info-Master

Person responsible for adding content to the sites and managing the meta-data of the site. They do not currently work with search analytics, and have a lot of other things on their plate. Is considered one of the main target-groups for this search service.

Management

The requirements for management level personnel is not known today, does not even know which management level this data would be for.

Thoughts about the data presentation of today

He recognizes that tables are the most common way to present data, and think that other visualizations would require third-party implementation. They have considered using Adobe Omniture to work and visualize the data.

Key-matches

They have a key-matching system today, but no way to see how much the promoted links are used or how well they match the users information request. More qualitative information is wanted.

Interviewee 4

Date: 2011-05-27

Who is he and what does he do?

He has been working as an information master for about one year, and has also worked with web-related tasks for a long time. Is responsible to coordinate the questions that regard their websites, both internal and external. Communicates the strategies about communication to the informants. All in all responsible for their websites.
The amount of time he spends on web- and search-analytics is quite small. Each month he uses the tools from time to time, but not on any regular basis. He consider him self to be a novice in the subject and when it comes to the usage of these tools.

What are the conditions under which he works?

He works part time (50%) with looking at statistics regarding their websites. This involves both web analytics and search analytics. However looking at search analytics isn't something he does. He did once but didn't see the benefits with it so he haven't looked at it since.

Who else work with search analytics within the organization?

It is a big organisation, which have many people working with web and search statistics, but at his location he is pretty much the only one working with it.
There are also web-masters and content editors that can access the data as well but it is unclear if they do; he might forward information to them though.

How does the company work with Search Analytics?

The organizations goal with web and search statistics is to make sure that the investments in content creation and curation has a good pay-off in terms of users (internal and external) finding and accessing the content. This is to motivate the investments of time and money in the content from a budgetary perspective.

What does he look at today?

He wishes to get an understanding about how the content on their sites is being used by the visitors. Today he uses what links on a page are the most used to get this understanding. Basically he wants a ranking system for all links on a specific page. The number of unique users is also something that is looked at.
He's interested in seeing what information it is that their users are looking for and if they are finding it. When he looks at this data he often look at them in time-chunks of one month. He says this is because otherwise there isn't enough interesting information to look at.

What he thinks of today's tools

The current systems that they use (AWStats and Google Analytics) make it to complicated to elicit the information that he wants. The system themselves are too complicated for him. He only has time (and reason) to access the statistics about once a month, and this is not enough time to learn the tools.
He also thinks that those tools are very much targeted at commercial organizations, tracking conversions and such, and not a great fit for a public service organization such as theirs.
As mentioned he looks a great deal at web analytics and not search analytics. However he has had a look at search analytics, top ranked search queries, but he couldn't see how he was supposed to use that information and hence he didn't view the information as useful.

Tracking Changes

This is not something that they do today - he does not know if they *can* do it, the tools are too complicated for him to figure it out.

What he wants but can't do today

See pre-generated standard reports that combine all the different key factors (such as visitor time, ranking of links, etc) into one easy and understandable report. He doesn't want to make these reports, or collect the data needed for them, him self but rather have them presented to him automatically in the statistics system. If someone else had prepared these reports for him so that he only needs to look at them and from them get an understanding and then take the required action the would find it very helpful.
He has trouble formulating his needs, due to his inexperience with the area - it appears he knows too little to understand what is possible.
His main request is to make it easier to use and understand. He believes it is a very interesting and useful way to gain insight into certain questions, but also very complicated to find answers to those questions.

Interviewee 5

Date: 2011-05-31

Who is she and what does she do?

She has worked as an info-master and web-responsible for 10 years. Of these 10 years she have been working with looking at web and search statistics for 7 years.
As a info-master she is responsible for the content on their website. She sometimes informs others of what information that needs to be changed, or she removes sites that no one are using.
She works with web/search analytics once a month.

What are their goals with search analytics?

The main area of interest is “what do people want, what is it that they feel is missing?”, although she did not believe statistics can answer those questions. She also expressed interest in knowing what people does not view, pages with low view-counts. Another goal is to engage more people to use the intranet in general.
All in all, there appeared not to be any other goals set by the organization, and the interest in analytics seem to be pretty small. She also expressed doubt in the usefulness of search/ web analytics to provide information that can be used to improve things - because it only displayed “how things used to be and how things stand right now”.

What task does she perform?

Her main focus is looking at which links people use on the site (through a statistics-overlay in Google Analytics) and who is visiting the site (departments within the company, external contractors/customers). She also looks at search queries and modifies meta-data based on that information, but could not explain more in detail.
Overall, she feels that she can do what she needs to do today, but at the same time she admits not knowing everything that *is* possible to do. She does look at a lot of statistics, but seem to not really know what to do with it.

Who else works with or uses the information?

No one else works with the information. However her superiors are interested in seeing statistics about how their website is used. So she have to be able to deliverer these kinds of reports. These reports are very simple, just a summary of the statistics presented in a power point slide. The tools do not support her very well in doing this.

What does she think of the tools they have today?

She feels that the tools of today are hard to work with. She thinks they are too advanced and hard to understand. The lingo of the tools are hard to understand. The tools are not very nice to look at, they’re not pleasant on the eyes. They are to “mathematical”, strict looking, too serious and have an overall look that is overwhelming.
When asking questions about the visualizations in Google Analytics she responded by saying that they are boring looking and it is not immediately possible to understand what they mean.
One tool that she uses today is AWStats. She thinks that it looks like an outdated piece of software. However they have a simple interface that is kinda minimalistic that she feels is easier to understand because it looks cleaner than i.e Google Analytics which she also uses.
One good thing that she can do with Google Analytics is to have it send her a report of the statistics to her each month, reminding her to look at it.

What would she like to be able to do?

Visualizations that are easy to understand is something that she wants. She would like a tool that tells her what to look at and alerts here to look at a certain thing that might need attention.
Segmentation of the data is useful here. She thinks it’s interesting to see what a certain department searches for on their website. Basically she wants to know what a certain customer group is searching for on her site.
Having an intelligent tool that tells here what to-do/look at is something that she would really want. This would lighten the burden that she seems to associate with looking at statistics. In her

own words, “Slipper tänka själv, det är ju skönt” (translation: “I don’t have to think, that’s nice”). She would really like to have a tool that has a much simpler language, more descriptive and easier to understand what things mean. The tools should provide a more exciting, cleaner and simpler look.
When demonstrating the use of Google Analytics, she said that she would have to click through many different pages in the tool to find what she was looking for, since she could not remember where specific types of data were located. She expressed, both verbally and through her actions, that she did not read many of the names of different functions.

Closing comments

It takes a lot of work to gather the right information, but the information is useful - make it easier!

Interviewee 6

Date: 2011-06-08

Who is he?

Works as a web-editor, mainly with the external-facing website. Has been working with this for the past three years.
Has an overarching responsibility for the content of the front page and county-common sections of the website. Plans and orders material from free-lance journalists to provide content to the site. Also does some work with requirements and analysis regarding new IT-solutions together with the IT-department.
Has some experience in using web analytics (both at work and privately), but has not spent a lot of time with search analytics.

Others

The co-workers within the department are interested to learn how their site is used, and find it a bit exciting as well. A yearly report is to be presented to the department head, summarizing the data for the year. He could collect a lot of the visualizations for the report from the tools, but he thinks it looks better if he creates them himself.

Tools

When he first joined the organization they used AWStats for analysing web statistics, but he believes it to be not that intuitive, fun to use or good-looking. Then worked to introduce Google Analytics into the organization.
Google Analytics is thought to be better looking. The dashboard help him to prioritize what to look at, and the names of the functions available are easy to understand. He thinks the filtering is great, and like that you can really dig deep down into the data by just moving forward.
He recognizes that there is probably more data available in Google Analytics that what they need or actually looks at, and that they could probably work just as well with a simpler tool. But then again, you don’t *have* to dig deep, he argues. He sees the fact that they don’t utilize all the potentiality within Google Analytics as a bigger problem than it being too complicated.
Filtering is very important to him. He wants his personal filters already applied when he opens the tool, and then use different profiles to view different data.

Goals with Analytics

Before and in the beginning when Google Analytics was introduced, there were no set goals for what the statistics were to be used for. But late last year they created a short list of things that are to be tracked. It was important that it would not take to much time and work to perform the data extraction and analysis, because they all have other tasks to do at work. Therefore, the list ended up being quite short: check the usage of the front page and the different sections of “prioritized” content. This was initially planned to be done on a monthly basis, but was later moved to a quarterly and finally yearly basis. This was due to the time requirement to perform an analysis, and the low number of staff available.
The data that is being used is: number of users, devices used to access, plugin-supported, traffic to specific pages or subject areas, searches from external search engines.
The goal is to understand what the users are reading, and if the investment in different content was justified and to see developments year-over-year.. This can then be used in prioritizing different areas (both subject and internal/external), understanding how much content is needed, finding content that needs to be promoted in some way, justify adapting the site to other devices (mobile, tablet, etc.). Not all of this is realised yet, but they are working towards it.

Search Analytics

They have not looked much at all on the search statistics for the site-search. It’s mostly because of them not understanding the tool available for viewing that data. They were hoping to look at it in Google Analytics, but it was not possible.
They really want to do more work with search analytics, but there are other things that need to be done before that.

Time span

While they have decided to work with analytics on a yearly basis, there is still need to be able to see data regarding individual months and weeks, and to compare them to other time periods. It is useful to be able to know why the data differs month-over-month, etc.

Missing Data

He does not feel that there is anything missing today, Google Analytics provides him with everything he needs.

SEO

There is not much work done with specific meta-data, but there are guidelines in how to improve the ranking in external search engines (which words to use, etc.) - standard SEO work.

Finishing Words

When asked to sum up his approach to web/search analytics he gave his view on how he thinks it should be adapted to suit the organization as a whole, because it is important for both the more central organizational units as well as for the editors further out in the organization.
To suit everyone, it should be initially quite simple with clear checklists to follow. When they feel comfortable with that, then there should be ways to dig deeper into the data.

Interviewee 7

Date: 2011-06-08

Who is he?

Works as a web-editor (webbredaktör),managing both internal and external content and their sites. Has about seven or eight years of experience with web and search analytics, both from his previous job and his current job at (which he started about two years ago).
He works with statistics about once a month, for about four hours. This time constraint is set by himself but is limited because he has other tasks to perform as well. Sometimes when there is a special event (e.g. press releases and such) he will follow up on that more quickly and frequently.

Others

There is an interest from the content-owners/creators within the department to know how many views their content had and how they got there. He regularly posts a news item on the intranet with a short text summary and a few “printed” (in pdf) reports from Google Analytics containing the most popular information (top pages, top search words, etc.).

Goals with Analytics

The goal is to understand which pages people view, which pages that are not viewed, what people search for (in external search engines), track special events and their sites; summing up: what works, and what does not.
Examines ranking in external search engines to see understand if the page can be optimized for especially relevant search phrases. The internal search engine can use those improvements as well. He also uses referrals to track who links to them, and works to encourage and maintain those referring site in any way he can.

Tools

He likes Google Analytics, both the visual and the logical layout. Greatly appreciates the ability to filter data and the fact that it is personalizable, that *his* data is presented immediately to him in the way he has decided. He also has a good understanding that the statistics does not tell the whole story, that cookie-based tracking is not representative for everyone, and this leads him to focus more on ratios than exact numbers.

Finishing Words

He believes it to be a quite unique type of analysis, allowing different views than what regular media provides. But it is also under-utilized, people don’t track these things enough! He suggests that there need to be ‘tip-of-the-week’-styled suggestions to the user regarding what they can do with the data available.
He also thinks that the best information can be gathered when several related sources of data is cross-referenced and matched, and this is quite demanding to do today.

He also describes a need for more regular statistics (pages visited, time spent, etc.), the content creators are interested in these kinds of things to understand how people use their content.

B Personas

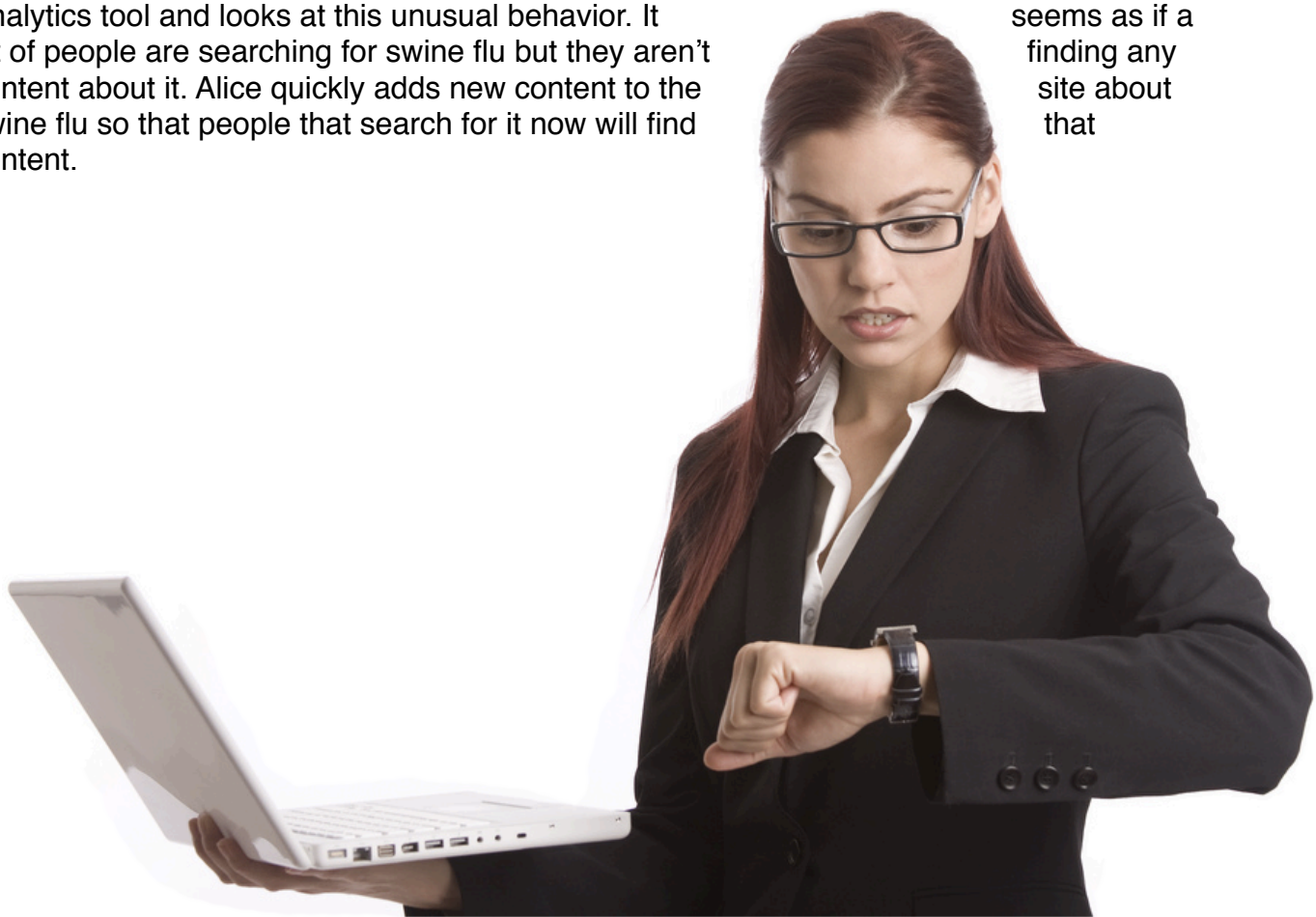
Alice

- 40 years-old
- Family, married with children
- Novice regarding working with SA and understanding the goals of SA
- Lives and works in an urban area
- Many different work tasks each day, amount of time for SA varies depending on workload
- Medium computer experience
- Completed higher education
- Wears glasses
- Works in a large organization with a wide variety of users and goals
- Appreciates new, better tools
- Mandatory use of tool
- Works with adding and editing content on company website
- Reports to upper management
- Full time employee
- Have used similar tools previously, but they were not satisfactory

Alice is a married woman with two children living in an urban area. She enjoys working with her colleges and spending time with her family. She is curious about the world around her and likes to explore. She has an academic background, however not within the field of computer science or other computer related subjects.

Scenario
After performing her other tasks for this day, Alice has a look at the search statistics. She wants to see what people have been searching for on the company website, and if they have found what they were looking for.

All of a sudden she receives a notification about a report from Robin. Robin has noticed an unusual behavior that he wants Alice to have a look at. Alice starts the Search Analytics tool and looks at this unusual behavior. It seems as if a lot of people are searching for swine flu but they aren't finding any content about it. Alice quickly adds new content to the site about swine flu so that people that search for it now will find that content.



Robin

- 35 years-old
- Lives in city urban area
- Expert when it comes to computers.
- Has a good understanding of how the back end of a web site works
- Understands of goals with SA well
- Has completed a higher education with a technical focus
- Works for several companies as a consultant
- He is a very adaptable person who knows the possibilities within the subject
- The use of the SA application is required in his job
- Primarily works with analyzing search statistics, and can perform advanced analysis
- Informs other about potential problem areas/subjects, does not make any changes to the content himself

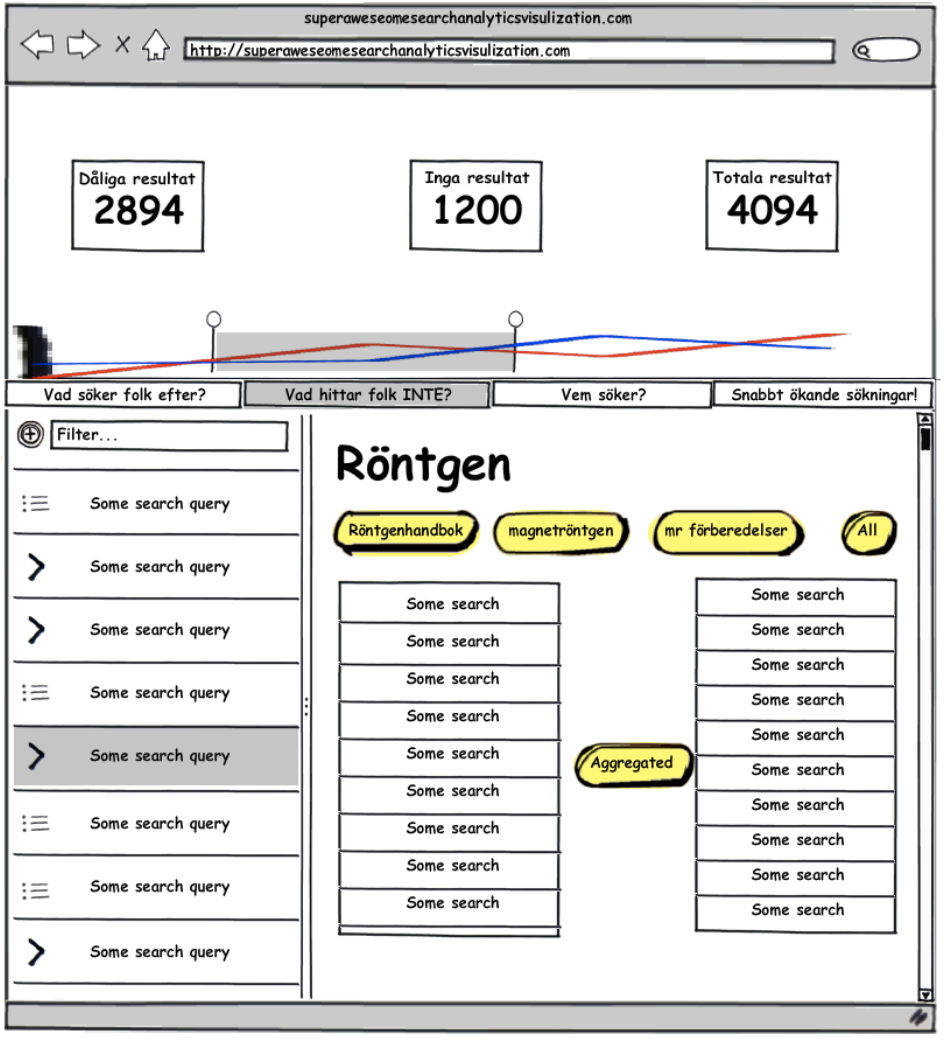
Robin is a single guy who enjoys hanging out with friends, loves sports and watching a good film from time to time. His dream is to start a successful company and get rich. Takes pride in doing a god job.

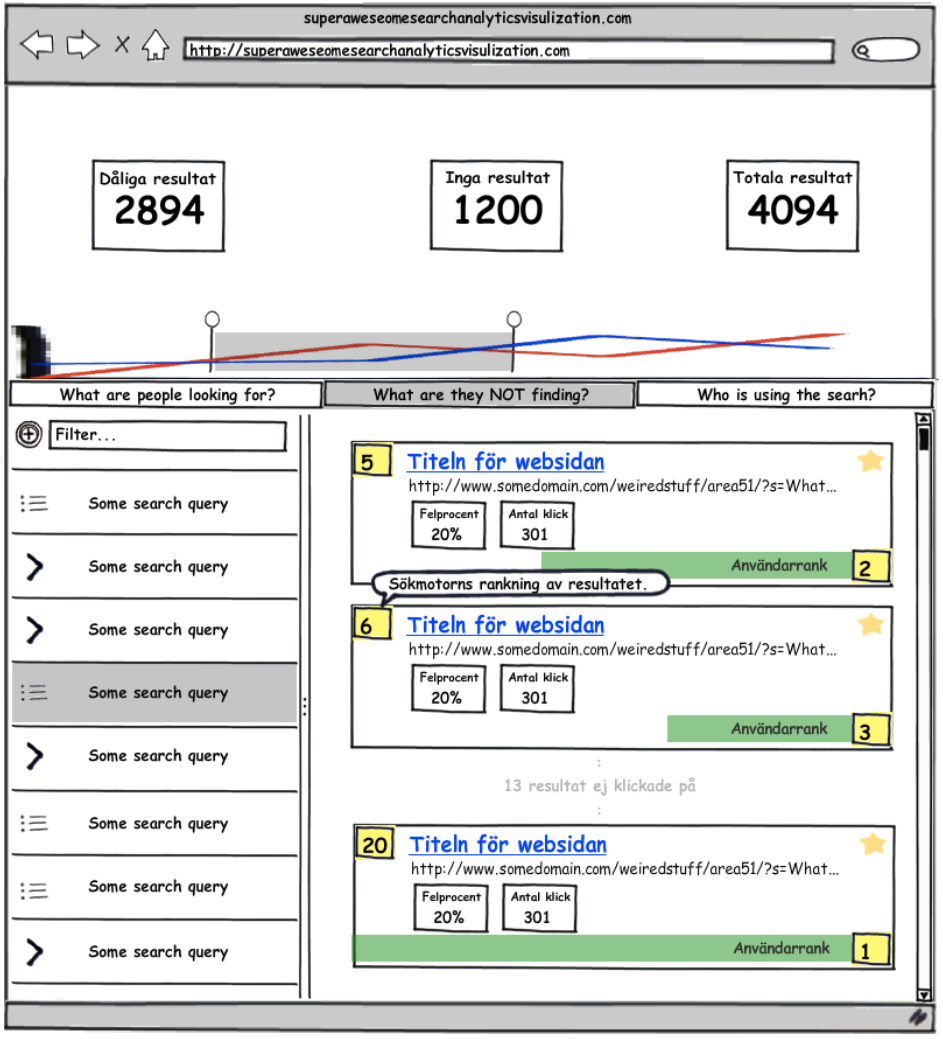
Scenario

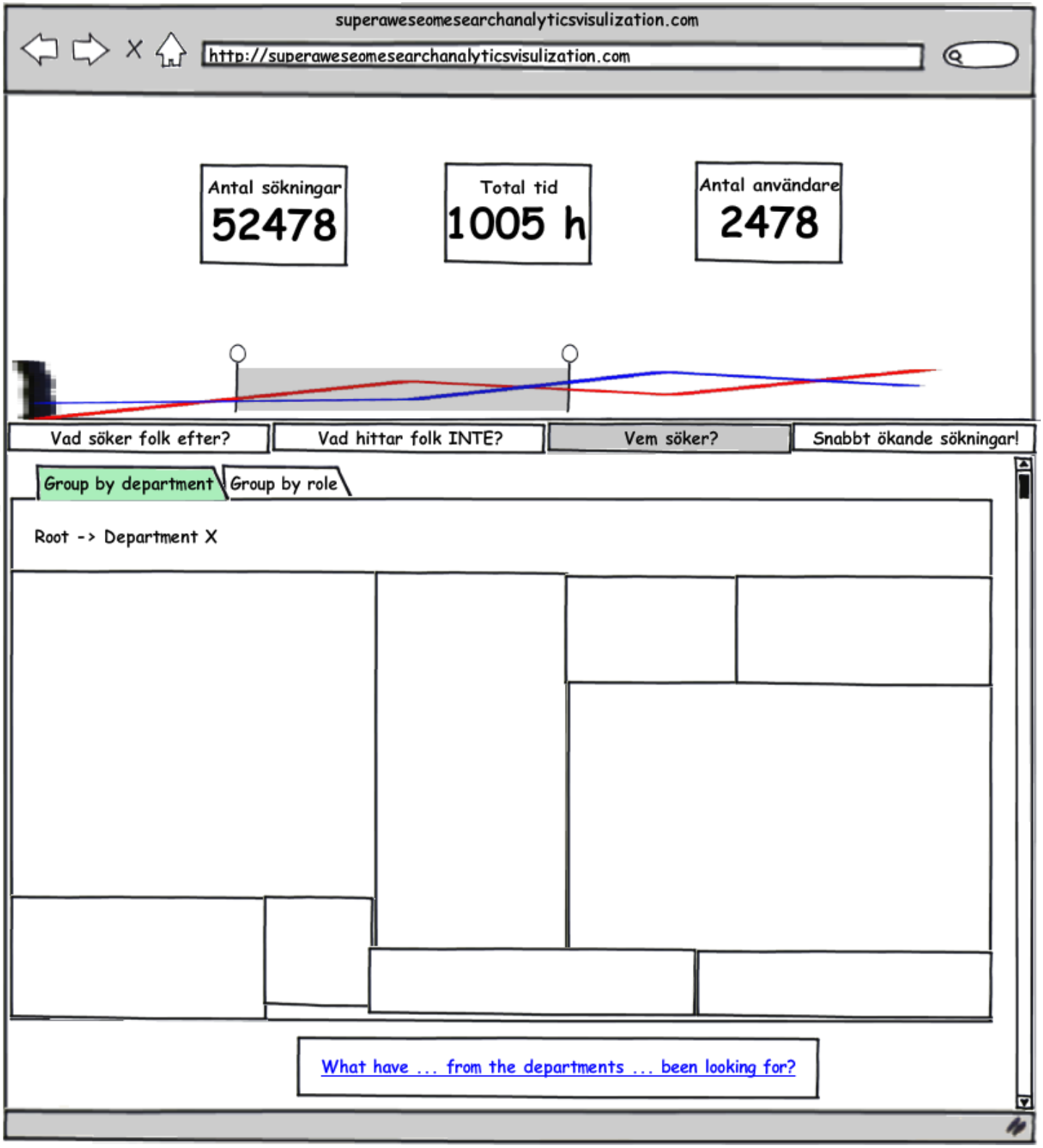
Robin is working on analyzing the last quarters search logs to create reports that he can hand over to the different client companies. He has several KPI's to measure and compare with the previous quarter that he then have to report to each company. After he has performed his analysis and formed a conclusion, he has to make his findings and the data that they are based on understandable for the intended receiver. His regular job is, among other things, to supervise the usage of the search service and report any issues regarding the sites content which may need to be addressed. After any changes have been made, he keeps that subject under surveillance to check that the changes had the desired effect.

C Wireframes


superaweseomesearchanalyticsvisulization.com
http://superaweseomesearchanalyticsvisulization.com
Lägg till filter
3 filter aktiverade
Antal sökningar
52478
Spenderad tid
15min
Antal användare
2478
Vad söker folk efter?
Vad hittar folk INTE?
Vem söker?
Snabbt ökande sökningar!
Sök
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
En sökfråga
Söktermen
Antal sökningar
301
AvgResultPagesV
2
Antal resultat
4
Facetter
En facett
en Annan facett
en väldigt lång facett
...
Facetter
Facett ipsum
Facett ipsum
Facett ipsum dollar sit bacon
Facett ipsum
Resultat
Sortera efter
5
Titeln för websidan
http://www.somedomain.com/weiredstuff/area51/?s=What...
Felprocent
20%
Antal klick
301
Användarrank
2
13 resultat ej klickade på
20
Titeln för websidan
http://www.somedomain.com/weiredstuff/area51/?s=What...
Felprocent
20%
Antal klick
301
Användarrank
1


superaweseomesearchanalyticsvisulization.com
http://superaweseomesearchanalyticsvisulization.com
Dåliga resultat
2894
Inga resultat
1200
Totala resultat
4094
Vad söker folk efter?
Vad hittar folk INTE?
Vem söker?
Snabbt ökande sökningar!
Filter...
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
Röntgen
Röntgenhandbok
magnetröntgen
mr förberedelser
All
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Aggregated
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search
Some search


superaweseomesearchanalyticsvisulization.com
http://superaweseomesearchanalyticsvisulization.com
Dåliga resultat
2894
Inga resultat
1200
Totala resultat
4094
What are people looking for?
What are they NOT finding?
Who is using the searh?
Filter...
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
Some search query
5
Titeln för websidan
http://www.somedomain.com/weiredstuff/area51/?s=What...
Felprocent
20%
Antal klick
301
Användarrank
2
Sökmotorns rankning av resultatet.
6
Titeln för websidan
http://www.somedomain.com/weiredstuff/area51/?s=What...
Felprocent
20%
Antal klick
301
Användarrank
3
13 resultat ej klickade på
20
Titeln för websidan
http://www.somedomain.com/weiredstuff/area51/?s=What...
Felprocent
20%
Antal klick
301
Användarrank
1


superaweseomesearchanalyticsvisulization.com
http://superaweseomesearchanalyticsvisulization.com
Antal sökningar
52478
Total tid
1005 h
Antal användare
2478
Vad söker folk efter?
Vad hittar folk INTE?
Vem söker?
Snabbt ökande sökningar!
Group by department
Group by role
Root -> Department X
What have ... from the departments ... been looking for?

D Paper Prototype

Profil ett

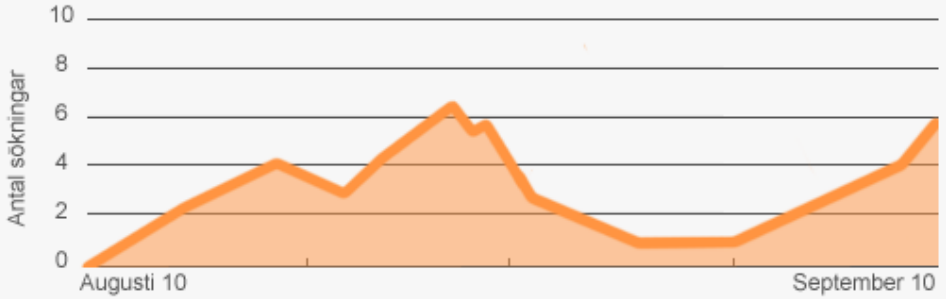
Antal sökningar	Snitt tid spenderad på sök	Antal användare
1 767	3,1 min	954

Vad söker folk efter? | Vad hittar folk ej? | Vem söker? | Vilka sökningar ökar mest?

Sök ...

Sökfras	Antal sökningar
Terminalglasögon	877
Semester 2011	400
Semester	361
vab	50
dagis förmån	43
bilpol	15
träningsdiet	11
Företagstelefon	8
blå uppgången	1
karta sahlgrenska	1

specialistremiss



Antal sökningar	Antal resultat	Antal resultatsidor besökta
4	19	1

Filter använda vid sökningen

VGIntrasu | Sahlgrenska Universitetssjukhuset

Resultat

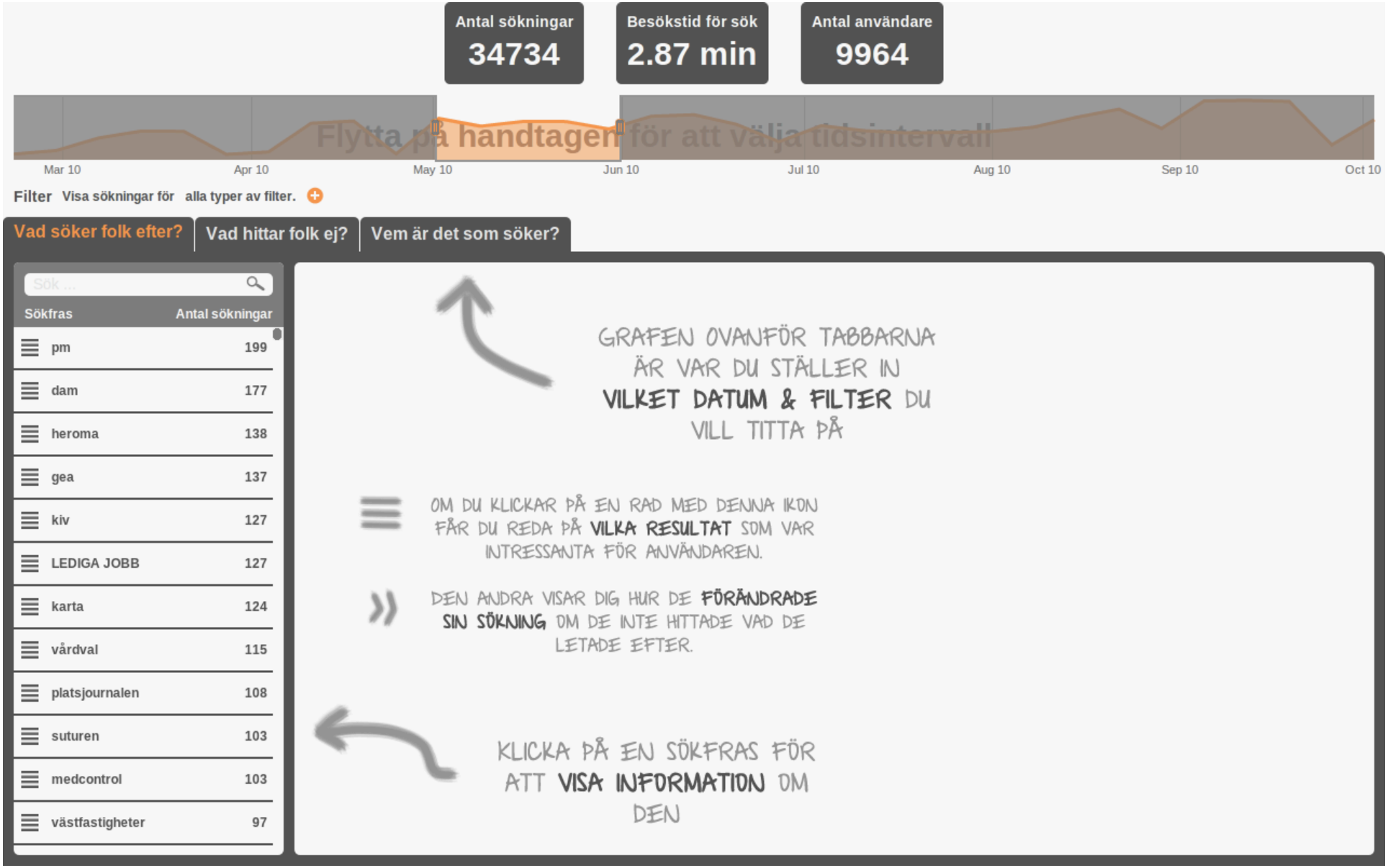
Sortera efter: System rankning

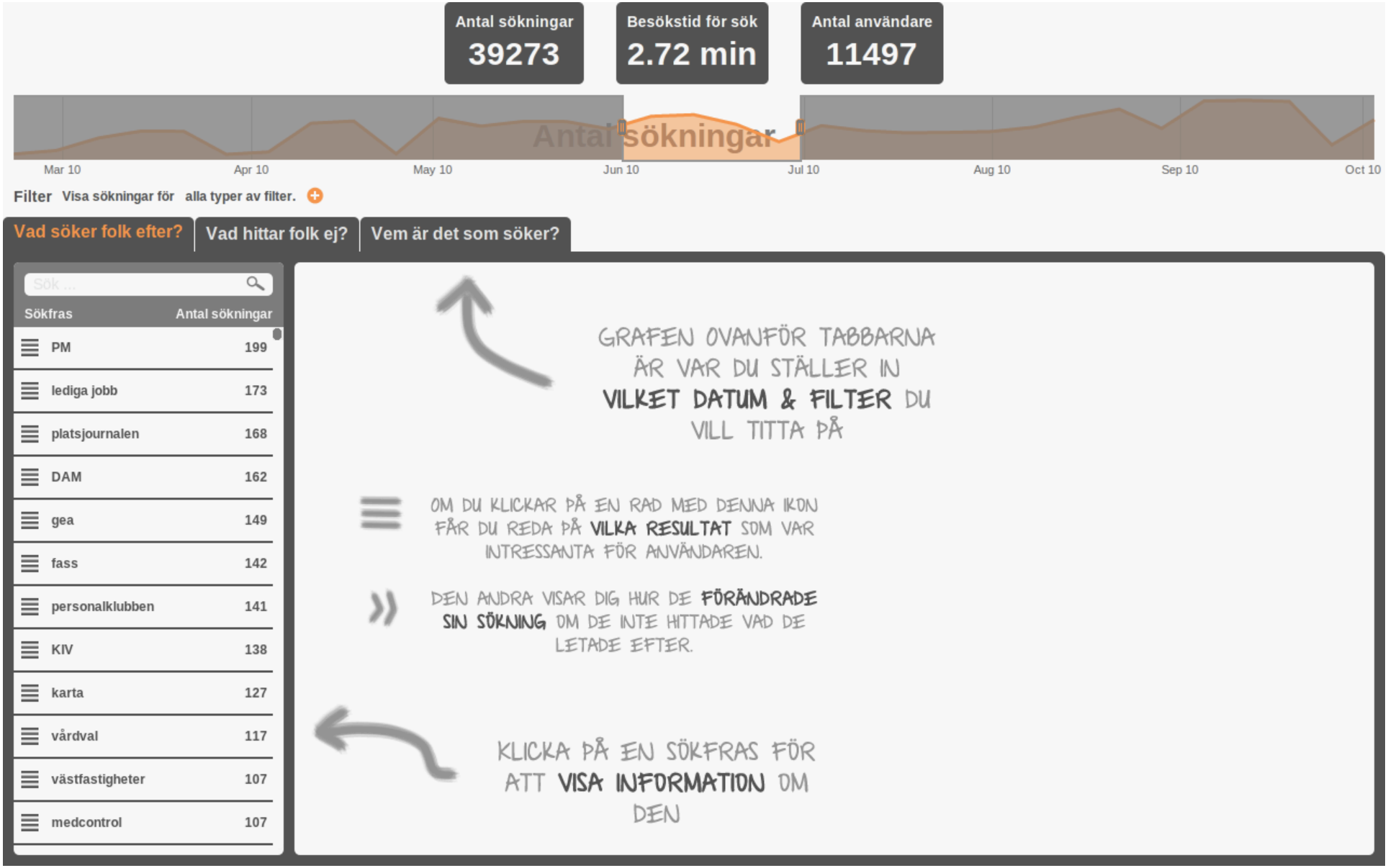
1 **Hur man utfärdar en remiss**
http://www.loremipsumdollarsit.com/search.jsp?bacon=lolbaconbbq
Felprocent 100% | Antal klick 1 | Användarrank

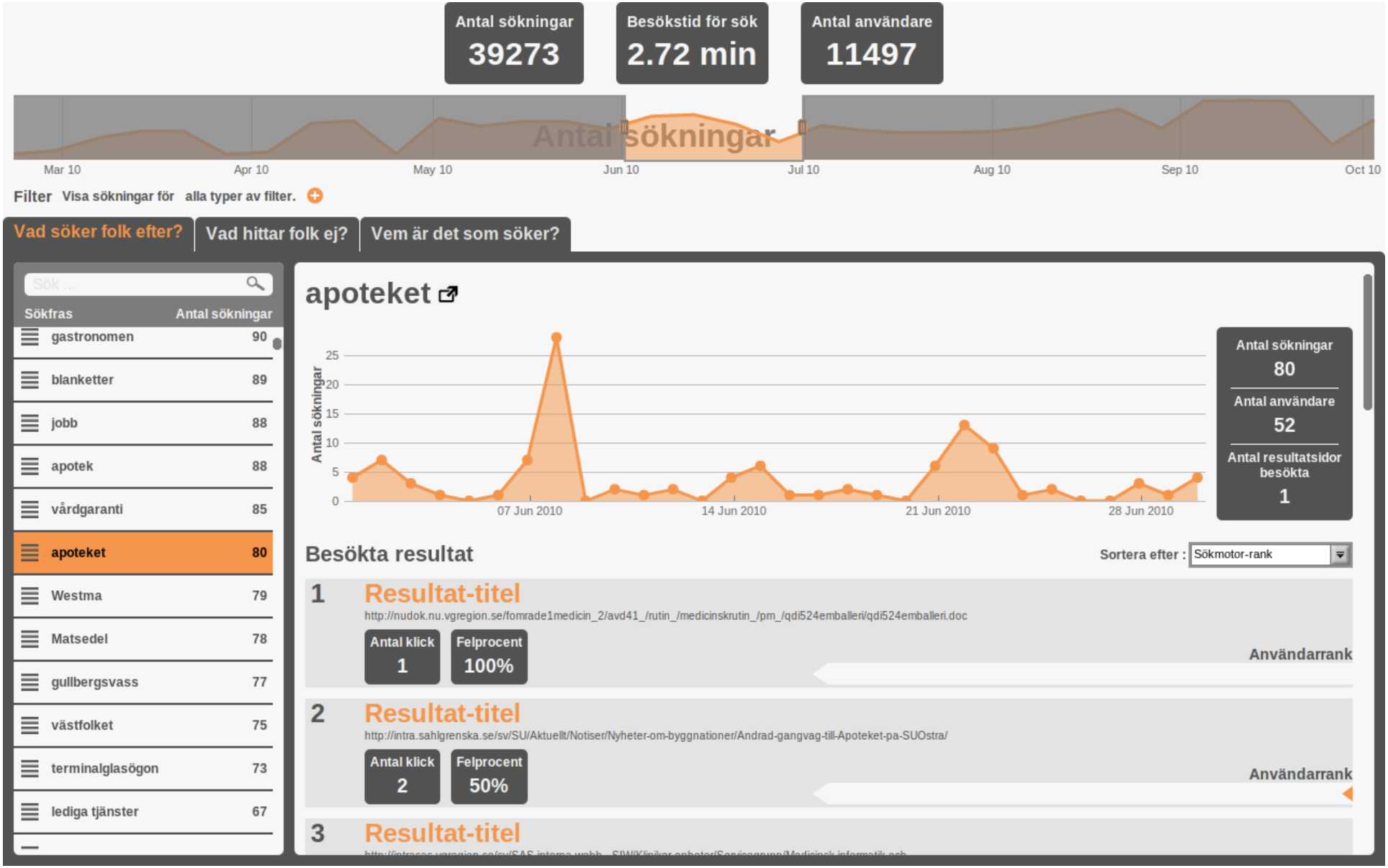
2 resultat ej besökta

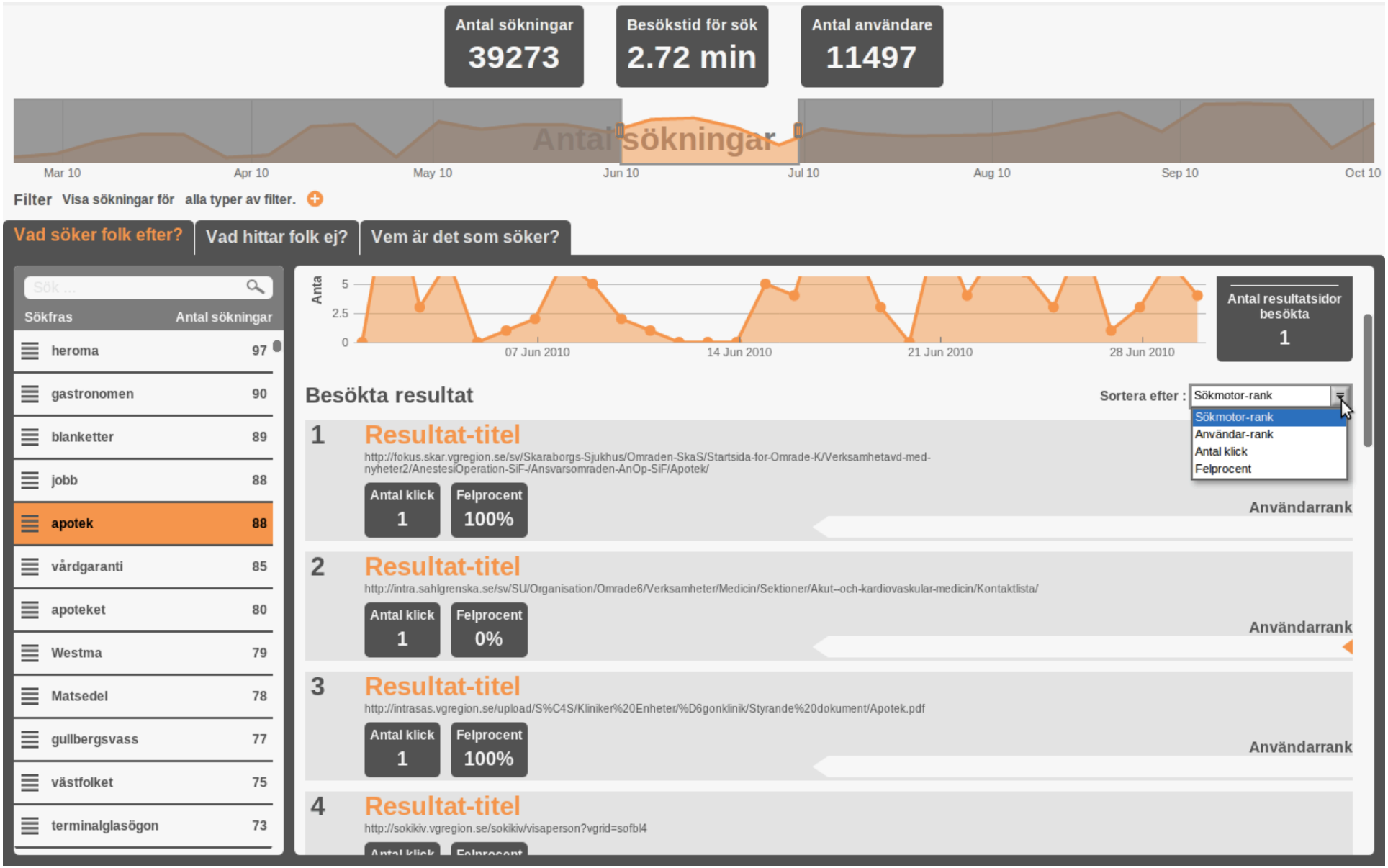
4 **Vilka remisser ska egentligen rapporteras**
http://intra.sahlgrenska.se/sv/SU/Aktuellt/Jan-Direkt/Fragearkiv/Vard/Vilka-remisser ...
Felprocent 0% | Antal klick 2 | Användarrank

E Implementation

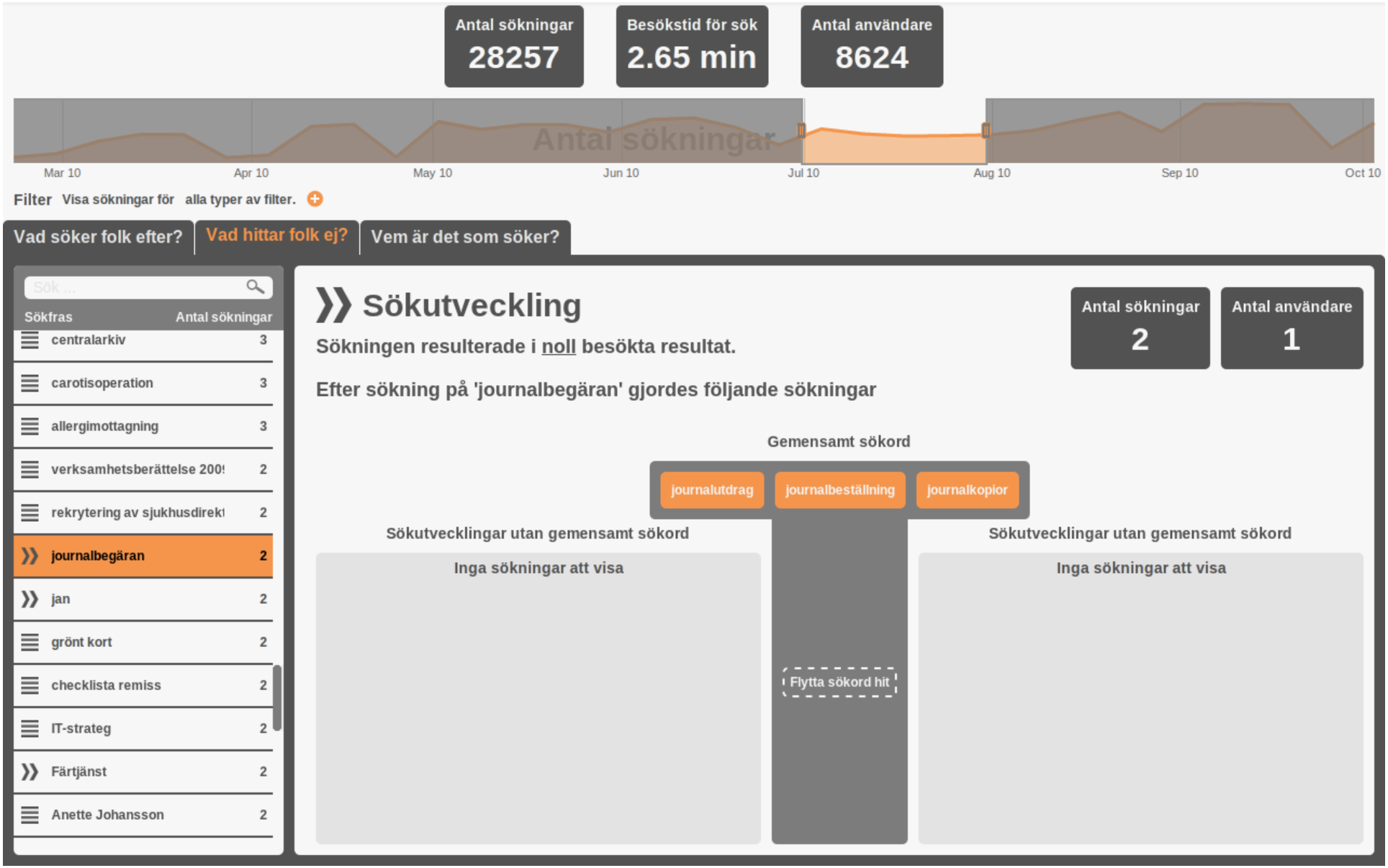

Antal sökningar
34734
Besökstid för sök
2.87 min
Antal användare
9964
Flytta på handtagen för att välja tidsintervall
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter Visa sökningar för alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
pm 199
dam 177
heroma 138
gea 137
kiv 127
LEDIGA JOBB 127
karta 124
vårdval 115
platsjournalen 108
suturen 103
medcontrol 103
västfastigheter 97
GRAFEN OVANFÖR TABBARNA ÄR VAR DU STÄLLER IN VILKET DATUM & FILTER DU VILL TITTA PÅ
OM DU KLICKAR PÅ EN RAD MED DENNA IKON FÅR DU REDA PÅ VILKA RESULTAT SOM VAR INTRESSANTA FÖR ANVÄNDAREN.
DEN ANDRA VISAR DIG HUR DE FÖRÄNDRADE SIN SÖKNING OM DE INTE HITTADE VAD DE LETADE EFTER.
KLICKA PÅ EN SÖKFRAS FÖR ATT VISA INFORMATION OM DEN


Antal sökningar
39273
Besökstid för sök
2.72 min
Antal användare
11497
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter Visa sökningar för alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
PM 199
lediga jobb 173
platsjournalen 168
DAM 162
gea 149
fass 142
personalklubben 141
KIV 138
karta 127
vårdval 117
västfastigheter 107
medcontrol 107
GRAFEN OVANFÖR TABBARNA ÄR VAR DU STÄLLER IN VILKET DATUM & FILTER DU VILL TITTA PÅ
OM DU KLICKAR PÅ EN RAD MED DENNA IKON FÅR DU REDA PÅ VILKA RESULTAT SOM VAR INTRESSANTA FÖR ANVÄNDAREN.
DEN ANDRA VISAR DIG HUR DE FÖRÄNDRADE SIN SÖKNING OM DE INTE HITTADE VAD DE LETADE EFTER.
KLICKA PÅ EN SÖKFRAS FÖR ATT VISA INFORMATION OM DEN


Antal sökningar
39273
Besökstid för sök
2.72 min
Antal användare
11497
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter Visa sökningar för alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
gastronomen 90
blanketter 89
jobb 88
apotek 88
vårdgaranti 85
apoteket 80
Westma 79
Matsedel 78
gullbergsvass 77
västfolket 75
terminalglasögon 73
lediga tjänster 67
apoteket
Antal sökningar
25
20
15
10
5
0
07 Jun 2010
14 Jun 2010
21 Jun 2010
28 Jun 2010
Antal sökningar
80
Antal användare
52
Antal resultatsidor besökta
1
Besökta resultat
Sortera efter : Sökmotor-rank
1 Resultat-titel
http://nudok.nu.vgregion.se/fomrade1medicin_2/avd41_/rutin_/medicinskrutin_/pm_/qdi524emballeri/qdi524emballeri.doc
Antal klick 1
Felprocent 100%
Användarrank
2 Resultat-titel
http://intra.sahlgrenska.se/sv/SU/Aktuellt/Notiser/Nyheter-om-byggnationer/Andrad-gangvag-till-Apoteket-pa-SUOstra/
Antal klick 2
Felprocent 50%
Användarrank
3 Resultat-titel

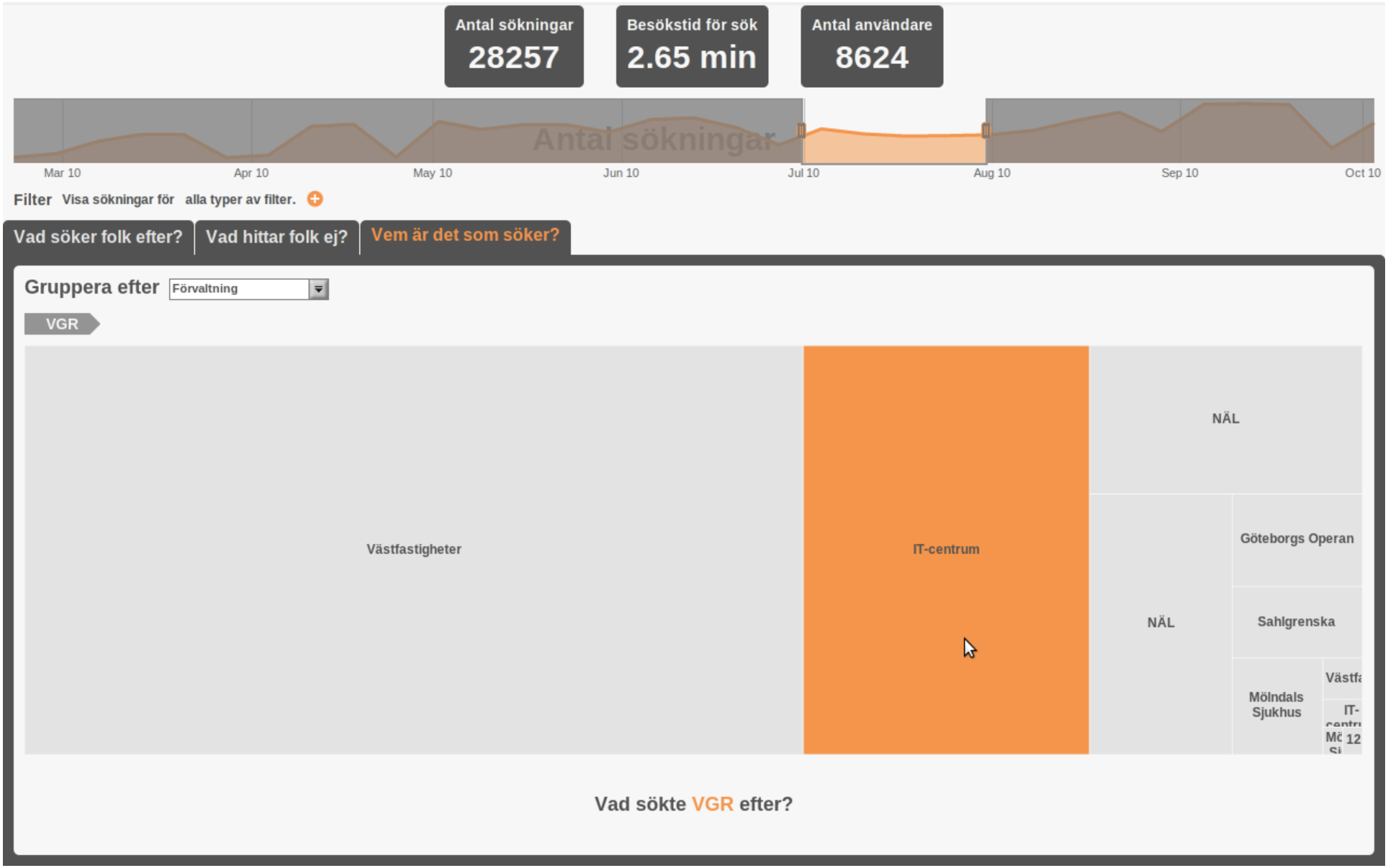

Antal sökningar
39273
Besökstid för sök
2.72 min
Antal användare
11497
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter Visa sökningar för alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
heroma 97
gastronomen 90
blanketter 89
jobb 88
apotek 88
vårdgaranti 85
apoteket 80
Westma 79
Matsedel 78
gullbergsvass 77
västfolket 75
terminalglasögon 73
Anta
5
2.5
0
07 Jun 2010
14 Jun 2010
21 Jun 2010
28 Jun 2010
Antal resultatsidor besökta
1
Besökta resultat
Sortera efter :
Sökmotor-rank
Sökmotor-rank
Användar-rank
Antal klick
Felprocent
1 Resultat-titel
http://fokus.skar.vgregion.se/sv/Skaraborgs-Sjukhus/Omraden-SkaS/Startsida-for-Omrade-K/Verksamhetavd-med-nyheter2/AnestesiOperation-SiF-/Ansvarsomraden-AnOp-SiF/Apotek/
Antal klick 1
Felprocent 100%
Användarrank
2 Resultat-titel
http://intra.sahlgrenska.se/sv/SU/Organisation/Omrade6/Verksamheter/Medicin/Sektioner/Akut--och-kardiovaskular-medicin/Kontaktlista/
Antal klick 1
Felprocent 0%
Användarrank
3 Resultat-titel
http://intrasas.vgregion.se/upload/S%C4S/Kliniker%20Enheter/%D6gonklinik/Styrande%20dokument/Apotek.pdf
Antal klick 1
Felprocent 100%
Användarrank
4 Resultat-titel
http://sokikiv.vgregion.se/sokikiv/visaperson?vgrid=sofbl4


Antal sökningar
39273
Besökstid för sök
2.72 min
Antal användare
11497
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter Visa sökningar för alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
heroma 97
gastronomen 90
blanketter 89
jobb 88
apotek 88
vårdgaranti 85
apoteket 80
Westma 79
Matsedel 78
gullbergsvass 77
västfolket 75
terminalglasögon 73
Anta
5
2.5
0
07 Jun 2010
14 Jun 2010
21 Jun 2010
28 Jun 2010
Antal resultatsidor besökta
1
Besökta resultat
Sortera efter :
Användar-rank
6
Resultat-titel
http://intra.sahlgrenska.se/sv/SU/Organisation/Omrade-1/Verksamheter/Medicin/Enheter/Avdelning-324/Dokument/Externa-kontakter/Apotek-/
Antal klick
13
Felprocent
7.69%
Användarrank
5
Resultat-titel
http://intra.sahlgrenska.se/sv/SU/Organisation/Omrade6/Verksamheter/Medicin/Sektioner/Hematologi-och-koagulation/Om-sektionen/Ovrigt/apotek/
Antal klick
12
Felprocent
8.33%
Användarrank
8
Resultat-titel
http://fokus.skar.vgregion.se/sv/Skaraborgs-Sjukhus/Nyheter-Startsidan/Fokus-Nyhetsarkiv/Helgstangt-pa-Apotek-Kaman-pa-KSS/
Antal klick
6
Felprocent
0%
Användarrank
2
Resultat-titel
http://intra.sahlgrenska.se/sv/SU/Organisation/Omrade6/Verksamheter/Medicin/Sektioner/Akut--och-kardiovaskular-medicin/Kontaktlista/


Antal sökningar
28257
Besökstid för sök
2.65 min
Antal användare
8624
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter
Visa sökningar för
alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
centralarkiv 3
carotisoperation 3
allergimottagning 3
verksamhetsberättelse 200 2
rekrytering av sjukhusdirekt 2
journalbegäran 2
jan 2
grönt kort 2
checklista remiss 2
IT-strateg 2
Färtjänst 2
Anette Johansson 2
Sökutveckling
Sökningen resulterade i noll besökta resultat.
Efter sökning på 'journalbegäran' gjordes följande sökningar
Antal sökningar
2
Antal användare
1
Gemensamt sökord
journalutdrag
journalbeställning
journalkopior
Sökutvecklingar utan gemensamt sökord
Inga sökningar att visa
Flytta sökord hit
Sökutvecklingar utan gemensamt sökord
Inga sökningar att visa


Antal sökningar
28257
Besökstid för sök
2.65 min
Antal användare
8624
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter
Visa sökningar för
alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
centralarkiv 3
carotisoperation 3
allergimottagning 3
verksamhetsberättelse 200 2
rekrytering av sjukhusdirekt 2
journalbegäran 2
jan 2
grönt kort 2
checklista remiss 2
IT-strateg 2
Färtjänst 2
Anette Johansson 2
Sökutveckling
Sökningen resulterade i noll besökta resultat.
Efter sökning på 'journalbegäran' gjordes följande sökningar
Antal sökningar
2
Antal användare
1
Gemensamt sökord
journalbeställning
journalkopior
Sökutveckling innan journalutdrag
journal, journalbeställning, journalkopior, journautdrag
journalkopior
journalutdrag
Sökutveckling efter journalutdrag
journalhandling, journal annat sjukhus, journalförfrågan, begäran, journal begäran, andra sjukhus, andra sjukhus journal
journalbeställning


Antal sökningar
28257
Besökstid för sök
2.65 min
Antal användare
8624
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter
Visa sökningar för
alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Sök ...
Sökfras
Antal sökningar
centralarkiv 3
carotisoperation 3
allergimottagning 3
verksamhetsberättelse 200 2
rekrytering av sjukhusdirekt 2
journalbegäran 2
jan 2
grönt kort 2
checklista remiss 2
IT-strateg 2
Färtjänst 2
Anette Johansson 2
Sökutveckling
Sökningen resulterade i noll besökta resultat.
Efter sökning på 'journalbegäran' gjordes följande sökningar
Antal sökningar
2
Antal användare
1
Gemensamt sökord
journalbeställning
journalkopior
Sökutveckling innan journalutdrag
journal, journalbeställning, journalkopior, journautdrag
journalkopior
journalutdrag
Sökutveckling efter journalutdrag
journalhandling, journal annat sjukhus, journalförfrågan, begäran, journal begäran, andra sjukhus, andra sjukhus journal
journalbeställning


Antal sökningar
28257
Besökstid för sök
2.65 min
Antal användare
8624
Antal sökningar
Mar 10
Apr 10
May 10
Jun 10
Jul 10
Aug 10
Sep 10
Oct 10
Filter
Visa sökningar för
alla typer av filter.
Vad söker folk efter?
Vad hittar folk ej?
Vem är det som söker?
Gruppera efter
Förvaltning
VGR
Västfastigheter
IT-centrum
NÄL
NÄL
Göteborgs Operan
Sahlgrenska
Mölndals Sjukhus
Västfa
IT-
12
Vad sökte VGR efter?

F Summary of User Test Two

User Test Analysis

User Test 2

Friday, September 23, 2011

Task 1

- Does not understand the wording “tabbarna” in the help text.
- Could easily identify the time-graph
    - immediately grasped the idea of dragging and resizing the selected time span
- Did not understand that the first tab was already selected
- Understands the ranking of the queries
    - but does not understand that the list of queries continues below and is scrollable
- When trying to figure out if people are finding what they are looking for, instead of clicking a query in the search pane he moves to the ‘not finding’-tab
    - the ‘looking for’-tab appears to not connect with that question that well - it does not lead to the understanding of the final use of the tab
- understands the mini-graph at once
- error rate is assumed to be connected to visit time, probably due to previous experience
- sorting
    - search engine rank is thought to be connected to external search engines
    - does not understand the user rank, thinks it is something about the users who performed the query
- The URL was initially thought to be some source of the users (referral?)
    - was later identified as the “landing page”
- gave the error rate a lot of thought, concluded that that it is dwell time
- thinks that the whole result link-section is one visit for a user
- really likes the wording of the tab-titles
- number of visits is finally understood
- does not understand the search engine rank label
- thinks the related/sponsored star is for indicating that the use found something good

Task 2

- finds the search function quickly
    - presses the search icon to perform the search, even though search as you type is activated
    - later presses enter instead
- has a hard time grasping what results mean in the tool-tip - uses “hit” to describe the

correct term to him
- uses the error rate to define the “best” result
    - wonders why the error rate sorts in the “wrong” order
    - when prompted comes to the right conclusion
- slowly begins to understand that the user rank is for describing the best results
    - but is still focused solely on the error rate and does not understand what it means

Task 3

- understands that it is the ‘not finding’-tab which contains the right information
- does not understand the different icons
    - the arrow is described as “go there”
    - the other as “text”
- has a very hard time to grasp the meaning of the search evolution
    - “they got a result, but did not proceed”
    - “after the searched for a they went to b...?”
- bad-results
    - likes the graph, looks like it is fun when things react
- understands the difference between the two types (no results / some results)
- thinks the ‘not finding’ search pane is interesting and useful
- search evolution
    - at first does not understand what ‘move search word here” means. tries to drag word from search pane to the drop zone. then tried to drag the agg-word, and success!
    - thinks the search evolutions are synonyms for the search phrase, that the list is before and after the indexation of the phrase
    - when there is no agg-word he thinks the words are pages (???), but is very unsure overall
    - does not understand the underlying point of the search evolution

Task 4

- explains that he sees himself as a novice user
    - does not understand the term ‘search engine rank’
    - likes the wording of the tabs
- wants the number of visits to say unique explicitly
- wants more information regarding the different metrics overall
    - needs to understand what he sees to use it (user rank especially)
- help-picture
    - the arrows steals the focus
    - does not link the icons in the picture to the icons in the search pane
- does not find the average session time-metric very useful

- does not grasp the meaning of the ‘result pages viewed’
    - thinks it represents a result link
- again, explain the metrics more and better

User test 2

Monday, 26 September, 2011

Task 1

- At the start she doesn't understand that the "correct" tab is already selected
- She completely understands the list showing what people have searched for.
- The explanatory text wasn't the first thing she noticed. Instead it was the tabs and the list showing what people have searched for.
- She can easily find what people haven't found but it's much harder to figure out what people have found.
- She don't believe that "Vad söker folk efter?" can answer the question if people have found what they were looking for.
- The explanatory text feels to "snäll" so she doesn't look at it, specifically pointing to the font-style. The arrows catch her attention more than the text.
- On the result page she immediately understands the boxes "Antal besök" and "Felprocent"
- At the start she thinks that the URL in a search result represents WHO has performed the query?
- She doesn't understand what relation "Felprocent" has to "Antal besök"
- Doesn't understand what the user rank should be used for. However, she does hint that it has a relation to "Antal besök"
- Doesn't understand what the star is used for.
- Doesn't understand what 'sökmotor-rank' is in the sort drop-down menu. She is unsure about the user rank but is correct with out knowing it.

Task 2

- Understands how to use the search in the list of what people have searched for. However, even if she sees the search-as-you-type she continues to write the entire word and then click on the magnifying glass to do a search. Didn't understand that the search-as-you-type did the search for her.
- She understands how to change the date. However she would like to be able to see exactly what dates she have selected.
- Doesn't quite grasp how to use user rank to see what results are the "good" ones.
- She understands that antal besök coupled with felprocent can give her the information of which results that are the "good" ones.

Task 3

- When she is on a search evolution with no common search phrases she tries to drag and drop rows from the search evolution lists and from the search phrase list of what people have searched for.
- Is confused over the fact that 8 users have done 14 searches.
- Does not understand the search evolution at all. Not even when presented with a good example of its function.
- When there was a common search phrase in the search evolution she understood to drag-and-drop the box to the drop area.

Task 4

- Has a hard time to understand what “number of result pages visited” means.

Task 5

- It is unclear that the number indicating the rank of the search engine is related to how the search engine does its ranking.
- She does understand that what change she will have to do depending on the result of a result list.