# Designing Government Portal Navigation Around Citizens' Needs

Rob Klaassen, Joyce Karreman, and Thea van der Geest

**Abstract.** Improving the usability of government portal sites requires a focus shift from system to user in both research and design. Empirical studies into user behavior are needed to support decisions on navigation, labeling and search systems. This paper presents such a study. Through scenario based interviews data were collected on citizens' information seeking needs and search strategies. Additionally, server logs files were analyzed. The results demonstrated the complexity of the search task from a user perspective, and provided suggestions for user friendly portal design. On the basis of the results it was recommended that portal sites' navigation systems should be context-rich, and labeling systems should be adapted to citizens' colloquial speech.

### 1 Introduction

Providing government information, services, products and transactions electronically has the potential benefit of accessibility for a wider audience, political and administrative transparency and improved service delivery [1] [2] [3]. Many governments and public bodies have created general portal websites to give users (such as citizens and businesses) access to their information base. Such portals are aimed at making the information available to all citizens through one access point<sup>1</sup>.

Research on providing access to large bodies of government information thus far has been predominantly sender-oriented. It has focused on methods for structuring the infobase and making individual items retrievable on the basis of precise specifications, or on the basis of taxonomies [4]. Although this work is important and fruitful, it puts the system at the center of the design process rather than the user (see also [5] [6] [7] [8]). This paper describes an effort to incorporate actual user behavior into the (re-)design of a government portal site.

The project was an initiative of ICTU, a Dutch ministerial agency that coordinates the portal site Overheid.nl ("Government.nl"). *Overheid.nl* is a portal that provides access to all online available government information, provided by a wide range of national, regional and local institutions and agencies. The website

<sup>&</sup>lt;sup>1</sup> General portals are to be distinguished from personal portals, that contain information that is specifically targeted at the user. (see [9])

M.A. Wimmer et al. (Eds.): EGOV 2006, LNCS 4084, pp. 162–173, 2006.

<sup>©</sup> Springer-Verlag Berlin Heidelberg 2006

was subject to a major reconstruction, and the research results were to be used as input in that process.

# 2 User Support Systems in Portal Web Sites

The main function of a portal is to give access to a wide collection of sources. The information contained in these sources is organized through some kind of structuring principle. However, the internal organization of the information base is not very relevant to end-users, as it can differ greatly from the way it is presented on screen. On screen, users are presented with different aids and means that should give them a grip on both the contents and structure of the information base. We call these user support systems. Three different systems can be distinguished [10]: navigation systems, search systems and labeling systems.

**Navigation System.** The navigation system is the complex of aids and means that enable users to find their way within the website. Some systems closely follow the information organization, resulting in for instance hierarchical lists of clickable topics. Other systems are more user-oriented, e.g. organized around frequently asked questions<sup>2</sup>, or scenarios derived from actual contexts of use. A navigating user is well supported when he or she:

- is able to construct a global representation of the websites contents;
- is aware of his/her location within the portal site;
- knows where he/she is coming from;
- knows where he/she can go from here.

To enable the users to orient themselves, the navigation system should provide appropriate feedback. This feedback can be presented in the interface in many different ways, for example as breadcrumbs (which display the followed link path on screen), or as a sitemap which offers an overview of the information available.

**Search System.** The search system is the complex of means and aids that enable users to search directly for relevant information within the portal. A search engine within the site (ideally) offers direct routes to the desired information, relieving the users from the chore of browsing through multiple pages. Search engines work on the basis of user input, often (combinations of) keywords. It therefore is critical that users' keywords match the terms used in the information or by the search system (cf. [4]).

Labeling System. The labeling system is the complex of labels (terms, names) that is used to identify parts of the site's contents and references. Names are used for buttons, menus, site maps and links. They must provide a good view of the content behind the label. Most importantly, they must be comprehensible, logical and clear for the user.

 $<sup>^2</sup>$  In principle, FAQ's are user-oriented because they are based on actual information needs of real users. In practice however, this principle is often neglected, and many FAQ's primarily represent the messages the site owner wants to send out.

Each of the three types of user support system leaves the designer with many options and choices. Decisions about how to realize the support system in the interface cannot be made without a thorough knowledge of users' information seeking needs and skills. Our empirical study was aimed at collecting data to support those design decisions about user-oriented navigation, search and labeling systems for the Overheid.nl portal.

# 3 Research Questions and Data Collection

The following three research questions guided our study:

- 1. Which search strategies do citizens use? (navigation system and search system)
- 2. How do citizens phrase their search questions? (labeling system and search system)
- 3. Which navigation system is best suited to citizens' needs? (navigation system)

We collected our data using three methods: 1) a comparative analysis of portal websites, 2) scenario-based interviews with citizens, and 3) an analysis of server log files. Due to time pressure in the portal design process, we decided to use the three methods in parallel.

*Comparative Analysis of Websites*: Thirty portal websites were analyzed and compared for their application of navigation systems.

Interviews: Thirty Dutch citizens who had recently used an online government information resource were selected randomly and interviewed. In the interviews respondents were asked to describe their most recent online government information search action in detail. Then twelve realistic scenarios were presented to them. Each scenario ended with a question. The respondents were invited to imagine how they would go about finding the necessary information to answer the question. The topics for the twelve scenarios were selected on the basis of the most frequently used search terms from the log files.

Log File Analysis: Two large server logs files were analyzed. One log contained the search terms that people had entered in the Overheid.nl portal's search engine. The second file contained the search terms that were used by people who had reached the portal site via a general search engine (like Google). Both log files contained the queries conducted from January until October 2005. The first log file contained 54.654 search terms and phrases. The second one the 100 terms and phrases most frequently used by the users who found Overheid.nl through a general search engine. For the study, both lists were combined into one comprehensive list. The accumulated list covered 787.144 actual queries of users. Table 1 contains an overview of the research questions and the contribution of the methods to the various questions.

<b>Research Question</b>	Method
Which search strategies do citizens use?	Interviews: to determine citizens' search strategies by analyzing accounts of their latest search action and accounts of scenario-based projected) search actions.
How do people phrase their search questions?	Interviews: to determine citizens' phrasing by analyzing the questions and search terms they produced for their latest search action and for the scenario-based (projected) search actions.
	Server logs: to determine patterns in the search terms used, by clustering the contents of the log file.
	Server logs + interviews: to assess search action success by entering citizens' search terms in Overheid.nl and Google, and analyzing the results.
Which navigation system is best suited to citizens' needs?	Site comparison: to gain an overview of frequently applied classifications in information organization.
	Server logs: to determine the match between logged search terms at one hand and frequently applied classifications at the other hand.

Table 1. Overview of research questions and methods

### 4 Results

#### 4.1 Navigation Systems and Their Match with Citizens' Needs

Thirty portal sites (twenty-one commercial ones and nine portals of governments) were analyzed for the classifications applied. Information can be organized in many different ways. For example: government information about unemployment could be organized in themes like Work and Income, but also around life events (Losing your job), or around specific groups in the larger population (information for the unemployed). Table 2 shows the six most frequently applied classifications.

To determine which classification system would fit the information needs of Overheid.nl visitors best, we used the server log data. The 350 most frequently used search terms from the server log file were analyzed and matched with four of the six most used classifications from Table 2: Thematical/hierarchical, Current issues/news, Organizations/Government agencies and Life events. The search terms were not classified by target groups, because they could not be attributed to target groups with sufficient reliability. The search terms were not classified by FAQ either, as the Overheid.nl website does not contain such a list. The last column in Table 2 shows the degree of coverage of the four classifications that were analyzed. Coverage was defined as the percentage of the first 350 terms that could be matched with each of the classification systems.

The classification on the basis of Current issues/news covered the smallest percentage of all search actions. Obviously, citizens did not go out to find information on topics just because they were current. The thematic/hierarchical appeared to be the most effective organizing principle for a navigation system on a portal site.

Classification	Applied in # of websites $(n=30)$	Coverage (%)
Thematical/hierarchical	28	85
Current issues/news	23	9
Groups within population	17	*)
Organizations / gvt. agencies	8	52
FAQ	3	*)
Life events	2	22

Table 2.	Classifications	found	in 30	portal	sites
----------	-----------------	-------	-------	--------	-------

#### 4.2 Citizens' Search Strategies

In order to document citizens' search strategies, we analyzed the accounts the respondents gave of their most recent search for government information. The results show that the citizens mostly have open and complex questions, which are framed within a very individual and personal context. Only a few respondents had searched for government information with a clearly defined, non-complex question like: "I needed a mutation form for the rent subsidy". Most respondents had less well-defined questions like: "I was looking for information from the CWI, UWV, the city of [E] and the Internal Revenue Service, because I wanted to apply for social security."

Most respondents (17 out of 30) used only online resources to find an answer to their question. All of them reported that their search had been successful. The other 13 respondents also used web sites, but in combination with other channels, like telephone, or visits to agencies.

In the second part of the interviews, respondents were asked to describe their projected search strategy for twelve scenarios. Table 3 shows the channels respondents chose as a starting point for their search. We distinguished between online and offline channels, and between government and non-government sources.

Table 3 shows that respondents considered to use online and offline channels, government and non-government sources to an almost equal extent. However, preferred channels varied greatly between scenarios. People appeared to select the channel they used on the basis of the type of question they were confronted with. For some scenarios most respondents knew immediately which government agency to contact. This was the case for questions concerning Passport renewal, Unemployment benefit, Tax return (young workers), Driving license renewal, Road taxation, and to a somewhat lesser extent also forTax return (mortgage deduction) and Rental subsidies.

In these cases, where citizens seemed to know their way, the preference for offline channels appeared to be stronger than in the less obvious cases.

The other scenarios posed more problems for the respondents. In these cases, respondents started their projected search mostly with non-government sources. For instance, the scenario about Child day-care subsidy inspired respondents to

Scenario	Online gvt <sup>1)</sup>	Offline gvt	Total gvt	Online other <sup>2)</sup>	Offline other	Total other
Driving licence	5	25	30	0	0	0
Passport renewal	9	18	27	1	2	3
Tax return (young workers)	16	10	26	1	3	4
Unemployment benefit	10	13	23	2	5	7
Road taxation	16	6	22	3	5	8
Tax return (mortgage deduction)	10	10	20	2	8	10
Rental support	11	6	17	6	7	13
School holidays	7	4	11	8	11	19
Minimum wages	6	3	9	17	4	21
Child day-care subsidy	4	4	8	7	15	22
Lifecycle saving plan	5	1	6	15	9	24
Physiotherapy	1	1	2	10	18	28
Mean (n=30)	8	8	17	6	7	13

Table 3. Preferred channels and sources for 12 scenarios

1) gvt = government

2) other = non governmental organization, or individual person.

contact a day care center, or their employers. It seemed, that in these scenarios our respondents were less certain that their strategy would lead to success.

We also looked at the strategies respondents followed when they thought to use online channels. Table 4 shows the preferred strategies for the twelve scenarios.

Scenario	Most popular search strategy 1)		
School holidays	Google		
Minimum wages	Google		
Child day care subsidy	Google		
Lifecycle savingplan	Google		
Rental support	Google	URL, site's SE	
Physiotherapie	Google	URL, site's SE	
Driving license renewal	URL, site's SE <sup>2)</sup>		
Tax return (young workers)	URL, site's SE		
Unemployment benefit	URL, site's SE		
Road taxation	URL, site's SE		
Tax return (mortgage deduction)	URL, site's SE	URL, browse site	
Passport renewal	URL, browse site	Google	

Table 4. Projected online search strategies

1) When two strategies were equally popular, both are mentioned. 2) SE = Search Engine

Table 4 shows that for the less clear scenarios, people reverted to a general search engine (always *Google*). For the other scenarios, people came up with an URL they would try. Those respondents reported they would continue their search action by using the site's search engine. Only a few respondents expressed a preference for browsing the target site. This finding contrasts with earlier user

studies for Overheid.nl; it might well be a test artifact. Our respondents did not actually carry out the searches; they only had to imagine what they would do once they got to the website. It might be hard to imagine that you would browse a site when you are not actually visiting it.

Many of our respondents' comments suggest that they relied heavily on the website's contents and interface to guide them further. For example: "I expect there will be some information", "I'm sure that new regulations are on the site. I also expect a heading 'allowances overview' there".

In conclusion, citizens appear to focus their search process on finding the right source: the institution or agency that they expect to be able to answer their question. Citizens show a slight preference for direct ways of communicating with these sources (i.e. by visiting or telephone). The more certain citizens are about their choice for a source to consult, the more apparent this preference becomes. When, however, citizens are uncertain about which institution they should turn to, they prefer to use general search engines like Google to find relevant web sites. Once citizens find a (possibly) relevant website, they expect the website to guide them. Further, they have high hopes of the effectiveness of -their search engine use.

### 4.3 Citizens' Phrasing of Search Questions

We have analyzed the way in which citizens translate their information needs into concrete search questions in two ways. First, we analyzed the log files for patterns in the registered search terms. Secondly we analyzed the search terms that respondents chose in the scenario based interviews. Thirdly, the most frequently mentioned terms were entered in two search engines: Google and the search engine of Overheid.nl. The results for each term in both search engines were evaluated.

Log File Analysis. The combined server log files contained 54.754 search terms. However, this accumulated list contains many synonyms and near-duplicates We clustered related terms on the basis of four criteria:

- 1. Literal or near literal copies of the search term
- 2. Literal or near literal copies of the search term, combined with other terms
- 3. Synonyms
- 4. Semantically related terms

This produced a much clearer picture of the information the users were trying to find. Single keywords do not reveal much about the sort of questions that users have when they enter them. Seeing these keywords in a context of related search terms made interpretation easier. Table 5 shows the first 11 clustered items, the number of search actions performed with the original term, the number of related terms and the number of search actions with the clustered terms. Each popular term in the log files had a high number of related terms that appeared further down in the list. For instance, "cao" (collective labour agreement) was related to 1095 other entries. "Diefstal" (theft) had the lowest number of related terms: 56. The average number of related terms was 409.

Original term	# search actions with the original	# clustered (related)	# search actions with the clustered
	term	terms	terms
Government	24.758	876	58.781
Theft	9.746	56	10.091
Constitution	8.901	151	11.639
Cao <sup>1)</sup>	8.277	1.095	21.174
Ministries	6.837	754	22.905
Saving	4.149	133	6.136
Vacancies	3.155	271	15.666
Passport	2.752	182	5.040
Rental support	2.276	706	12.641
Bpr <sup>2)</sup>	2.198	58	2.678
Civil lawbook	2.180	215	4.960
Total	75.229	4.497	171.711

 Table 5. Overview of formed clusters of related terms

1) CAO stands for Collective Labour Agreement

2) BPR stands for National Registry of Personal Information and Travel Documents

Clustering was found to be an effective and efficient method for determining the actual information needs of the portal's users. The most popular terms were single terms. They revealed little about the types of questions users had. Clustering brought context to these terms in the form of related multiple term queries. This is best illustrated by an example. The search term "passport", for instance was entered 2752 times. The cluster "passport" contained 183 related search terms, that were entered 5040 times in total. A closer examination of the passport cluster shows that the term passport actually may have stood for many different questions. Some users searched for information on the number of years a passport remains valid, others wanted to know how to renew their passport, yet others wanted to register their children on their passport; some wanted to know what to do when a passport gets damaged, or how to renew a passport from abroad etcetera. This type of information is very useful for web site designers. It can help them decide on the topics for (for instance) a portal site.

**Interviews.** From our interviews it became clear that respondents found it very hard to phrase usable search questions and search terms. Presented with the twelve scenarios hey seemed to regard the scenarios as the search questions, and they tended to copy search terms from the scenario text. Of course, in normal life citizens will not often search for information on the basis of a given scenario. However, it seems probable that they will still use terms and phrases that are closely connected to their own interpretation of the problem situation.

Ranking of first usable answer	Score (n=18) in Google	Score (n=18) in Overheid.nl
nr. 1	5	5
nr. 2 to 5	4	2
nr. 6 to 10	0	1
nr. 10 to 50	2	not analyzed
No usable answer within first 50 (Google) or first 10 hits	7	10

Table 6. Search results for 18 frequently mentioned search terms

Also, day-to-day language usage will dictate the search terms people will come up with. People will try any term they assume to be connected with their question. A good example was given by one of the respondents, when he reports one of his experiences: "I got a heavy tax assessment, and didn't understand how this could have happened. Had something to do with overhevelingstoeslag<sup>3</sup>". The respondent visited the Internal Revenue Service website, and tried entering the term in the site's search engine. The respondent: "It didn't help. I searched for about 5 to 10 minutes, and I didn't make any progress."

Our results show clearly that respondents lack the notion that in order to search successfully they have to translate their personal problem situation into phrases and terms that match those used by the chosen sources and channels (for instance a government portal website).

The respondents did not seem to have an adequate mental model of their search situation. They had no overview over the sources and channels they used, the information available within these sources, and the structure of these sources. Put simply: respondents only have their own context to bring to the search process. This context does not always correspond very well to the official jargon. For instance, one of the scenarios ended with the question: "How would you try to find out whether your daughter's low salary is legal?". While any civil servant would immediately connect this question to the term "minimum wage", less than half of our respondents made the same connection.

Effectiveness of Citizens' Search Terms in Two Search Engines. In our scenario based interviews we asked the respondents to name suitable search terms for each specific problem: terms that could be entered into a search engine. We tested all 18 search terms that were mentioned by more than 3 respondents, by entering them in Google and the current government portal website (overheid.nl). We were able to establish the quality of the search results ('the search engines' 'hits'), by determining if any of the 'hits' provided in the search results contained the information needed for solving the problem described in the scenarios. We analyzed the first 50 hits in Google. Since Overheid.nl is a government website it seemed fair to apply a stricter criterium. Therefore we decided to analyze only the first 10 hits in Overheid.nl.

 $<sup>^{3}</sup>$  'Overheveling stoeslag' is a technical term in Dutch Tax Law.

Many of the 18 search terms that were frequently mentioned by the respondents delivered no results. However, several topics were well covered. "passport renewal", "driving license renewal", "holidays", "rental support", and "lifecycle saving plan" lead directly to the right information. These topics are either standard government themes, or themes that have been in the news a lot in recent times. With the other, less clear cut scenarios people are bound to be much less successful.

# 5 Conclusions

This paper demonstrates that user oriented research in the field of government portal design is valuable in two ways. First, observing actual users carrying out search tasks is a powerful method for obtaining insight into users' search strategies, information needs and search skills. Second, these insights provide designers with concrete suggestions for developing user friendly government web portals.

**Citizens' Search Strategies.** Citizens' search questions often are complex rather than simple. They can be characterized as follows:

- The question is open ended.
- There is more than one 'good' answer or adequate solution.
- The search question originates from a certain (personal) context. The search results have to be evaluated permanently for their applicability in this context.
- The search question is one step in an already complicated process. Often separate steps have to be taken towards separate government agencies or institutions.
- Different information sources have to be combined to produce an adequate answer.

The respondents' search strategies in this study resembled what Choo, Detlor & Tornbull (2000) have described as 'situated action'. Situated action means that search questions are not static througout the search process. Each newly found snippet of information can lead to the question being altered, refined and developed, or even to completely new questions.

The complex nature of search questions and the fact that information needs develop over the course of the search process have important consequences for portal website design. They have to enable users to constantly evaluate the relevance and applicability of the information they find on their personal situations. And they must enable users to decide whether to continue the search or not, and if so in what direction. Our results match findings from earlier studies [11] [12].

Our results also support claims that users benefit when answers are presented together with extra information about the context (see [13][14]). Users may not only be interested in the exact answer, but also in related information. Merely presenting users with seemingly 'exact' answers may not be sufficient to meet users' needs. Presenting them with related information may trigger a new search question, that the user could not have conceived by him or herself.

**Phrasing of Search Questions.** Citizens approach online government services, expecting to be understood when they use colloquial speech. Most citizens find it hard to translate their personal situation, or question into terms that match government vocabulary. They tend to use single general terms. Only a minority uses multiple term search phrases.

A closer analysis of log files shows that many different questions can be hidden behind one general term. Clustering terms provides some insight into this.

Not only is it difficult for citizens to phrase their questions 'correctly', even if they know the right terms, it still may lead them nowhere. The search engines, the most popular navigation system with our respondents, too often deliver disappointing results.

Navigation Systems and Citizens' Needs. Our study confirms that citizens need a lot of support when they search for information online. Many systems and tools can be used to provide this support. However, we believe that the following criteria should always guide government portal design.

- 1. Navigation systems should be contextually rich. Instead of merely presenting ordered lists of links and documents, the user should have the opportunity to evaluate the relevance of each link or document before opening it. Providing necessary context can be done by adding short descriptions to each link provided.
- 2. Labeling systems should be adapted to citizens' colloquial speech. The system should not only recognize government jargon. The issue of complexity of formal government language is relevant to more contexts than just portal sites. However, in an online environment citizens have to find the right terms themselves, and the dominance of formal language becomes even more problematic.
- 3. For a limited number of important and frequently searched for themes the portal site should provide information pages written in colloquial language, on which all information related to that theme is gathered, ordered and provided with context information. This will give users a necessary overview of all possibly relevant information. Moreover, they will not have to browse through endless lists of search results, without any guarantee for success.

The information pages should be made accessible both from the portal homepage and the site's search engine. The search engine should direct anyone who enters a related search term to these information pages. The clustered terms could be used in the design process as an indicator of possible content.

### References

- 1. European Union (2005, november) Ministerial Declaration on E-government: Transforming public services. Retrieved February 14, 2006, from http://www.egov2005conference.gov.uk/proceedings/.
- 2. Beyond e-Government. Report for the UK Cabinet Office. Maclean (VA): Booz Allen Hamilton.

- 3. West, D.M. (2004). E-Government and the Transformation of Service Delivery and Citizen Attitudes. Public Administration Review 64, (1). 15-27
- Sacco, G.M. (2005). Guided Interactive Information Access for E-Citizens. In: M.A. Wimmer, R. Traunmller, Å. Grönlund, K.V. Andersen (Eds.). Electronic Government: 4th International Conference, EGOV 2005, Copenhagen, Denmark, August 22-26, 2005. Proceedings. Lecture Notes on Computer Science 3591. 261- 268.
- 5. Thomas, J.C., & Streib, G. (2003). The New Face of Government: Citizen-Initiated Contacts in the Era of E-Government. Journal of public administration: research and theory 13, (1). 83-102.
- Klaassen, R.F. (2004). Voorlichtingskundig Ontwerpen: De Totstandkoming van Postbus 51-Campagnes [Public Information Design: The Development of Postbus 51 Public Information Campaigns]. Assen, The Netherlands: Van Gorcum Publishers.
- Van der Geest, Th. (2004). Tax to the max: Designing Web Services for Ordinary People. Document Design 12, (2). 213-218
- 8. Deursen, A. van, Dijk, J. van, & Ebbers, W. Why e-Government Usage Lags Behind: Explaining the Gap Between Potential and Actual Usage of Electronic Public Services in The Netherlands (in press).
- 9. Tatnall, A. (2005). Web portals : the new gateways to Internet information and services. Hershey (PA). Idea Group.
- Rosenfeld, L, & Morville, P. (2002). Information architecture for the World Wide Web. Beijing: O'Reilly.
- 11. Albers, M.J. (2004). Communication of Complex Information; User goals and Information Needs for Dynamic Web Information. Mahwah (NJ): Lawrence Erlbaum Associates, Publishers.
- 12. Choo, C.W., Detlor, B. & Tornbull, D. (2000) Web work : information seeking and knowledge work on the World Wide Web. Information science and knowledge management; vol. 1. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- 13. Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., Jacquemin, C., Lin, C.-Y., Maiorano, S., Miller, G., Moldovan, D., Ogden, B., Prager, J., Riloff, E., Singhal, A., Shrihari, R., Strzalkowski, T., Voorhees, E. and Weishedel, R. (2000), Issues, tasks, and program structures to roadmap research in question & answering (q&a), NIST DUC Vision and Roadmap Documents.
- Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., and Karger, D. R. (2003), What makes a good answer? the role of context in question answering, Proceedings of the Ninth IFIP TC13 International Conference on Human- Computer Interaction, Zurich, Switzerland.