

Evaluating a Thesaurus Browser for an Audio-visual Archive

Véronique Malaisé¹, Lora Aroyo², Hennie Brugman³, Luit Gazendam⁴,
Annemieke de Jong⁵, Christian Negru², and Guus Schreiber¹

¹ Vrije Universiteit Amsterdam, The Netherlands
schreiber@cs.vu.nl, vmalaise@few.vu.nl

² Technische Universiteit Eindhoven, The Netherlands
l.m.aroyo@tue.nl, c.m.negru@student.tue.nl

³ Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Hennie.Brugman@mpi.nl

⁴ Telematica Instituut, Enschedé, The Netherlands
Luit.Gazendam@telin.nl

⁵ Dutch Institute for Sound & Vision, Hilversum, The Netherlands
adjong@beeldengeluid.nl

Abstract. In this article we report on a user study aimed at evaluating and improving a thesaurus browser. The browser is intended to be used by documentalists of a large public audio-visual archive for finding appropriate indexing terms for TV programs. The subjects involved in the study were documentalists of the Dutch National Audiovisual Archives and of broadcasting corporations. The study provides insight into the value of various thesaurus browsing and searching techniques.

1 Introduction, Objectives and Approach

In this paper we report on a user study with a thesaurus developed for cataloguers of a audio-visual broadcast archive. This work is part of the CHOICE project¹ which aims to support annotation and search of the broadcast archive of the Dutch Institute for Sound & Vision. As part of this project we built a thesaurus browser for the GTAA² thesaurus. The thesaurus browser is a general SKOS/RDF browser [7]. We converted the original database representation of the thesaurus to SKOS (for conversion principles and representation details see [9]).

The purpose of the browser is to support cataloguers both of Sound & Vision and of the broadcast corporations in finding the appropriate indexing terms. Indexing is still mainly a manual process. Sound & Vision is in the process of moving to a completely digital archiving process and as a consequence heavier demands are put on the cataloguers. In fact, the browser is considered a simple baseline tool. The project also works on semi-automatic techniques for extracting

¹ <http://www.nwo.nl/catch/choice/>

² Dutch abbreviation for “Common Thesaurus Audio-visual Archives”.

indexing terms from context documents (TV guides, articles). Initial results of the semi-automatic support can be found in another paper [3]. This paper only discusses the browser.

The objective of the evaluation study was to improve the efficiency of the thesaurus browser in finding terms. We were particularly interested in how the browser aligns with cataloging practice. The subjects were people who are cataloguing audio-visual programs as part of their daily job. The study consisted of two parts. First we had a number of evaluation sessions with an initial version of the browser (Secs. 2-3). Based on the results of this first evaluation, the browser was adapted and evaluated in a second study (Secs. 4-5). In Sec. 6 we reflect on the outcomes and discuss related work.

From a general knowledge-engineering perspective, this paper focuses on questions related to user access to large knowledge structures, such as thesauri. Large knowledge structures typically incorporate many different viewpoints that one can take on the concepts involved [2]. As knowledge engineers we are used to organize concepts into large subtype hierarchies, but this may not always be the most appropriate way for accessing these concepts, given the user and his task context. Finding concepts in large concept structures for semantic-annotation purposes is becoming an increasingly important knowledge-access task [4]. This paper gives a detailed insight into knowledge-access problems in the domain of annotation of audio-visual archives.

The CHOICE project is part of the Dutch CATCH (Continuous Access To Cultural Heritage) Programme, funded by NWO (Dutch Science Foundation). A special characteristic of CATCH is that the teams of researchers are working part-time in the heritage institution. At the moment 10 of such projects are underway.

2 Thesaurus Browser

Cataloguers at Sound & Vision index TV programs by assigning to these a set of controlled terms, selected from the GTAA thesaurus. Currently, they only have access to these terms in the form of alphabetically sorted flat lists. Although the GTAA has internal structure this is not exploited by the current generation of software tools. Therefore, as a first step to improve the cataloging process, a thesaurus browser was designed and implemented.

2.1 Requirements

We identified the following requirements:

- Because the GTAA Browser will be used by both incidental and regular users and because these users are located both inside and outside of Sound & Vision, a *web application* was preferred.
- The thesaurus content is regularly updated, for example person's names and locations are regularly added. There is one authoritative resource for the GTAA, which is a relational database system maintained at Sound & Vision. The browser should therefore directly interact with this database.

- The browser should be able to display and exploit all structures that are present in the thesaurus in appropriate and intuitive ways. The same is true for structures and information that we add to the thesaurus. It should provide at least the existing searching and browsing functionalities, *i.e.* direct access to terms according to the facet to which they belong (this notion of facet is detailed in the following section) and an alphabetical search facility.
- For interoperability with other CATCH projects the thesaurus should be accessible through open web standards.

2.2 Browser – Version 1 Implementation

Fig. 1 shows the architecture of the GTAA Browser. The browser is implemented as a web application which can retrieve thesaurus data from an extensible set of data sources. One of those is Sound & Vision’s primary source of the GTAA, a relational database. Using this source, radio and television professionals will always have the latest modifications of the GTAA available. To accommodate the needs of researchers in CHOICE and CATCH the browser can also use an RDF/OWL representation of the thesaurus as its data source. This RDF/OWL store can be updated on request using a separate web application.

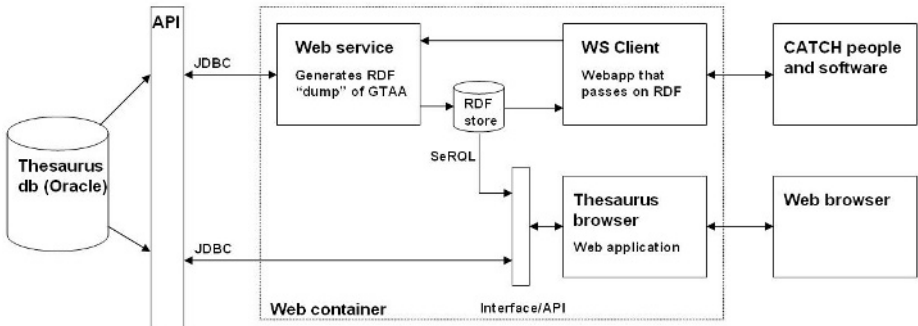


Fig. 1. Architecture of the browser

A screen shot of version 1 of the GTAA browser is shown in Fig. 2. The interface is divided into three main parts:

- the upper part, with 6 tabs (number 1 on the figure) representing the different dimensions (“facets”) of the GTAA;
- the middle part, where different information about the Terms are displayed (number 2 on the figure);
- the bottom part, consisting of an alphabetical search engine (number 3 on the figure).

Each of these three parts is discussed in more detail in the following paragraphs.

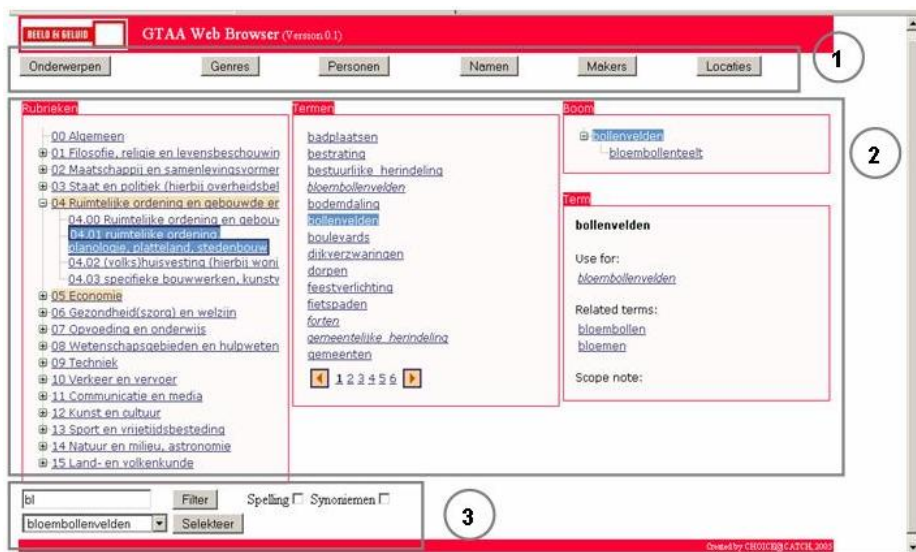


Fig. 2. First interface of the browser

GTAA Facets. The six tabs represent the different facets of the thesaurus: six disjoint groups of Terms, divided into top level categories. These facets are (between parentheses the Dutch term): Subjects (*Onderwerpen*), Genres (*Genres*), People (*Personen*), Names (*Namen*), Makers (*Makers*) and Locations (*Locaties*). The facets correspond to different fields in the indexing scheme of Sound & Vision for TV programs. They are given by the thesaurus structure and cannot be personalised by the user. The browser gives direct access to terms belonging to any of these facets by clicking on the corresponding tab. Fig. 2 shows the Subjects facet. The Subjects and the Genres facets are organized according to the ISO 2788 relationships: **BroaderTerm/NarrowerTerm**, **RelatedTerm**, and **Use/UseFor**. The **BroaderTerm/NarrowerTerm** is a hierarchical relationship that represents a description of subsets of documents (the **NarrowerTerm** should be used to describe a subset of the documents that can be described by the corresponding **BroaderTerm**). **BroaderTerm/NarrowerTerm** can represent a subclass relationship, as well as a part-of relationship, or some application specific relationship. **RelatedTerm** links two Terms that are closely related in a specific domain, like a Ship and a Sailor for example. Some Terms are also associated with **ScopeNotes**, textual comments about the use of the given Term. Terms from the Subjects facet are also grouped into **Categories**. These are an alternative way of grouping Terms, beside the **BroaderTerm/NarrowerTerm** hierarchy, by generic domain: Philosophy, Economy, etc. Terms from the four other facets are alphabetical flat lists, sometimes associated with **ScopeNotes**. As the Subjects facet is the most structured one, we detail its display in the middle part of the browser window (number 2 of Fig. 2) in the following subsection.

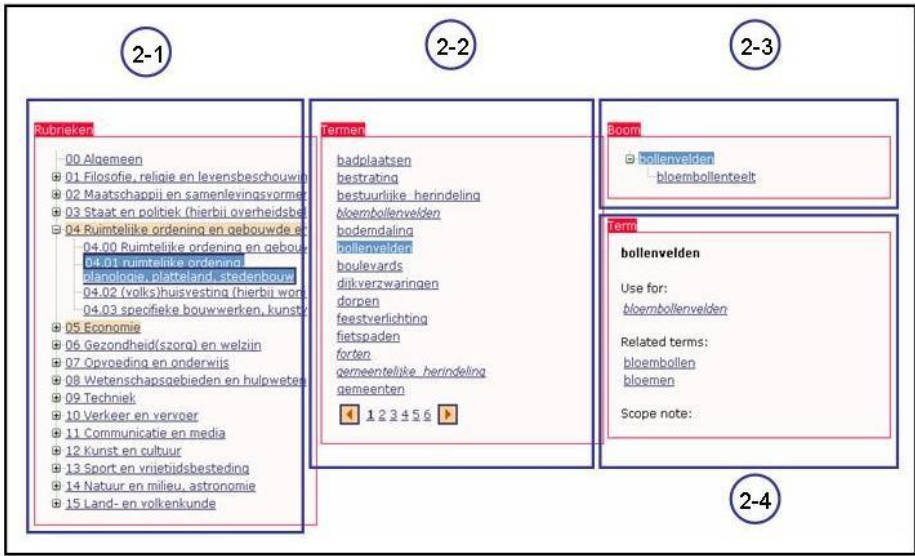


Fig. 3. Middle panel of the GTAA browser

Browsing Relationships Between Terms. A close-up of the middle panel of the Web Browser is shown in Fig. 3, where we can see that it is divided into four parts.

The left part (panel 2-1) displays the different Categories and Sub-Categories (*Rubrieken* and *Sub-Rubrieken* in Dutch) into which the Subjects Terms are categorized³. Clicking on a Category or Sub-Category displays in panel 2-2 the list of the Terms which belong to it⁴. The sample screen shot displays the terms from the sub-category *Urbanism and Organization of Public Space* highlighted in blue. Preferred terms are displayed in normal font and non-preferred in italics. Clicking on a term in this panel selects it, while the Category(ies) to which it belongs to are highlighted in orange in panel 2-1. Panels 2-3 and 2-4 are also instantiated or updated when a term is selected. They display the relevant **BroaderTerm/NarrowerTerm** tree (2-3) and other available information about the term such as **Related Terms** (2-4). Terms displayed in panels are clickable, enabling the user to navigate through neighbors of the selected term.

Alphabetical Search. In version 1, the search functionality was only valid in the facet that was active: if the user submits a query in the Subjects facet, the alphabetical search is limited to this facet. When the user types the first letters of a term, a refinement button (labeled *Filter*) gives the list of the preferred and non-preferred terms of the facet that begin with the same characters.

³ A term can be categorized in up to three different (Sub-)Categories.

⁴ If the Category contains more than 14 Terms, the first 14 Terms are displayed in alphabetical order, and other ones can be reached by clicking the different page number at the bottom of this panel.

3 User Study: Part I

3.1 Setup

Formative evaluation of the GTAA thesaurus browser in two parts was performed to determine whether it supports the cataloguers internal and external to Sound & Vision in their tasks of annotating audiovisual material, in particular in terms of navigation, browsing and searching. With the analysis of the study results we aim at answering questions about:

- the usefulness of browsing a hierarchical structure of terms versus alphabetical lists for finding out relevant terms;
- the intuitiveness of the search and navigation facilities;
- the effectiveness of the presentation of the controlled vocabulary, of the cross-links between the terms, of the categories and of the different dimensions of the GTAA thesaurus (namely the facets).

Subjects. The first user study concerned in total nine cataloguers: five thesaurus experts from Sound & Vision, two domain experts from NOS and two domain experts from EO⁵. Most of them (7 out of 9) are using annotation software daily.

Procedure. The experimental session lasted around 60 minutes per subject supervised by an examiner and video-recorded. To make sure the testing conditions are similar to all users we started with a brief (about 5 minutes) introduction of the experiment and the browser. Next, each of the subjects spent time for a “directed play-around” to get acquainted with browser’s functionality, reading a list of guidelines⁶ and reporting on problems. Subsequently, they watched an audiovisual document with a duration of 2 minutes and we asked to provide indexing terms for that document using the browser. They could use three strategies to find these terms:

- Use the Categories hierarchy to display lists of terms (henceforth *Browsing search* or *Browsing functionality*), in the Subject facet;
- Type in some letters in the alphabetical search box and check for a matching term by clicking on the *Filter* button (henceforth *Filter search*);
- Type in a whole term in the alphabetical search box and check for a matching term in the thesaurus (henceforth *Alphabetical search*).

From this first step on, the different relationships of the Subject facet could also be used to navigate in the thesaurus’ content, as well as the alphabetical lists in the other facets. No complex query composition functionality was provided to search for a term.

At places, where problems occurred the examiner initiated a dialog with the user in order to clarify the problem and to gather additional information on

⁵ NOS and EO are Dutch broadcasting organizations.

⁶ The guidelines can be found at <http://www.cs.vu.nl/~guus/public/choice-guidelines.pdf>. These are the adapted guidelines user for the second study.

it. Finally, each of them filled one usability questionnaire with five clusters of questions (overall interface, search facilities, term browsing, subject facets and additional functionality) and one personalia questionnaire focusing on sex, age and proficiency. All subjects were allowed to also use pre-selected on-line reference material on the topic of the audiovisual document during the annotation session.

Metrics. We evaluated the efficiency, satisfaction, learnability and effectiveness of the GTAA thesaurus browser by using the following metrics:

Number of problems during play-around: is used to calculate the overall learnability of the browser by counting the problems occurred over the number of steps and the overall time spent during the play-around;

Total time spent during play-around (in minutes): idem for number of problems;

Number of problems during annotation: is used to calculate the overall effectiveness by counting the number of problems over the steps and the overall time spent during annotation;

Number of times alphabetical search was used during annotation: gives an estimate of the efficiency of the alphabetical search in the two user studies.

Number of times hierarchy search was used during annotation: idem for alphabetical search;

Number of times filter search was used during annotation: idem for alphabetical search;

Number of steps during annotation: is used in the calculation of the efficiency of the browser for the annotation tasks;

Number of resulting indexing terms during annotation: is used to calculate the success factor in terms of overlapping with the terms indicated in the gold standard;

Number of steps per index term during annotation: is used as a measure of efficiency;

Total time spent on search tasks during annotation (in minutes): is used in the calculation of search efficiency of the browser in both user studies;

Total time spent during annotation (in minutes): is used in the calculation of the efficiency and effectiveness of the browser in both user studies.

We mark something as a *problem* when the user indicates that there is an obstacle to perform a task. For example, the user searches for “Afghanistan”, types the term in the search field (in the “Genres” facet), and it brings no results back, because the user didn’t select the facet “Locations”. Software bugs were also identified during the user studies, but were not counted as problems for the calculation of the effectiveness of the browser, nor in the total time spent with the browser.

A *step during annotation* is defined as a set of meaningfully connected atomic actions to perform an annotation task. For example, when the user is searching for “Afghanistan” he may first try to use the hierarchy in the Subject facet by error, then go to the Location tab, then type in some letters and get results from the filter list. This would result in three steps.

The *gold standard* was defined by thesaurus experts from Sound & Vision. It contained twelve indexing terms from the GTAA thesaurus, which they considered as appropriate to annotate the audiovisual document used in both user studies. We counted the total number of GTAA indexing terms that each subject used for the annotation and the total number of the ones which match the gold standard. We only considered exact matches in this evaluation, but an option for future studies could be to use a similarity function, for example based on the hierarchical structure of the thesaurus, to compare selected terms to the gold standard.

Questionnaire. In order to assess the usability of the GTAA browser the participants were asked to fill out a questionnaire⁷ (50 questions on a 7-point scale, 8 open questions). User satisfaction is expressed as a normalized value in the range [1,7], where 1 is highly satisfied and 7 is highly not satisfied. In order to identify trends in the user groups and discriminate different levels of expertise the participants we asked to also fill in a questionnaire about their personal characteristics with respect to gender, age, computer and annotation proficiency (10 questions).

3.2 Results

Table 1 shows the results of the study for the defined metrics. We can observe the following:

- Both during play-around and annotation a significant number of problems were encountered (on average $3.69 + 2.89 = 6.56$). On analysis these problems were mainly concerned with relatively trivial issues. For example, case-sensitive search and lack of auto-completion proved problematic. Another cause of problems was the lack of synchronization when updating parts of the screen (categories, term lists, etc.). Before the second user test the browser was accordingly adapted.
- A less trivial cause of problems was the confusion about combined use of the “filter” and “select” buttons for alphabetical search. For example, consider the following scenario:

```
SEARCH GOAL: "peace troops"
1. choose facet "Subjects"
2. enter query "peace" and click Filter button
3. select from drop-down list "peace troops" click
   Select button to activate search on "peace troops"
```

From the videos we observed that such scenarios were quite common. Participants would typically get stuck in step 3. It turned out that it was unclear to them where the drop-down box was and also that an additional “select” click was required. To remedy this, we improved the guidelines. Also, the organization on the screen was not logical (jump from bottom to top). This is improved in the adapted browser.

⁷ The questionnaire can be found at <http://www.cs.vu.nl/~guus/public/choice-quest.pdf> (in Dutch).

Table 1. Results for the first part of the user study. Time measurements are in min.

Stage/Metric	Subjects									Average
	S1	S2	S3	S4	S5	S6	S7	S8	S9	
During play-around										
#problems	3	4	5	4	3	2	3	4	5	3.67
total time spent	11	19	22	19	20	14	15	20	18	18
During annotation										
#problems	3	4	4	2	3	2	3	2	3	2.89
#alphabetical search used	8	9	7	6	7	4	7	9	6	7.00
#hierarchy search used	1	2	0	1	0	5	1	1	1	1.33
#filter search used	2	2	2	2	1	1	1	1	1	1.44
#steps	47	30	22	29	31	28	35	37	40	33.22
#resulting index terms	9	10	8	4	9	6	8	5	7	7.33
#steps per index term	5.22	3.00	2.75	7.25	3.44	4.67	4.38	7.40	5.71	4.87
total time spent on search	29	24	18	21	20	12	18	10	13	18
total time spent	38	33	29	30	31	20	25	19	22	27

- From the results it is clear that most of the subjects were mainly using alphabetical search. The explanation they gave was that they already knew the term they were looking and therefore hierarchical search is not appropriate. From the hierarchy search they mainly used the use-for and related-term relationships. The filter-search was also used infrequently; this appeared mainly to be caused by the problem reported before. The added value was in fact unclear to them.
- Subjects were performing a large number of total steps (on average 33.22) to find indexing terms. Analysis of the videos showed that the main reasons for this were (i) inefficiency of the screen layout, (ii) insufficient feedback on the action performed, and (iii) the filter-search problem mentioned earlier.
- On average the resulting number of indexing terms was 7.33, i.e. roughly 60% of the gold standard. We think this is adequate.
- On average 4.87 steps were needed to find an index term. The minimum number of steps needed to find a term would be 3 (see scenario above). This means there is definitely room for efficiency improvement.

Table 3 further on in this paper shows aggregated results of the questionnaire on user satisfaction. We discuss the results in Sec. 5 in relation to those of the second study.

4 Thesaurus Browser – Adapted Version

In accordance with the results of this first user study, we adapted the browser (Fig. 4). The next paragraphs describe the most important modifications that were made.

Alphabetical Search (Number 1 on Figure 4). Alphabetical search turned out to be important for users, so we made some small technical improvements, such as

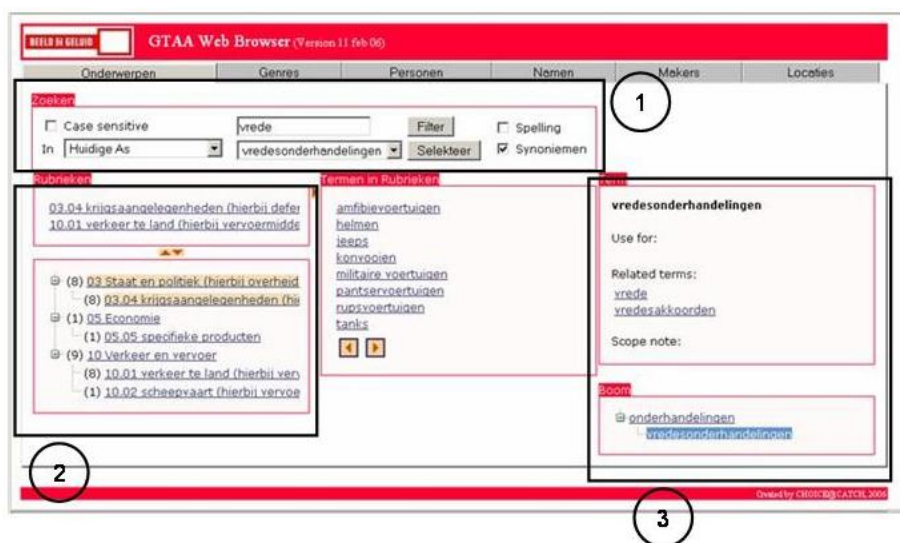


Fig. 4. Adapted interface of the Web Browser

a default behavior of case-insensitive search and the possibility to search within (i) the active facet, (ii) a given facet or (iii) any facet. As an additional facility in cases where the characters typed in do not match a thesaurus Term, the browser displays also:

- A list of spelling suggestions
- Terms that match the input through an intermediate list of synonym terms.

The spelling suggestion tool was adapted from a generic module and the synonym list has been computed using the online versions of the Van Dale and the Muiswerk dictionaries⁸.

Selection of Multiple Categories (Number 2 on Figure 4). One of the reasons why the browsing search was not preferred as a first step to search for a term is that the Categories are too broad: they contain too many terms to make the display of the whole list interesting for finding out a term. But we took advantage of the fact that most Subject terms are part of more than one Category to offer the user an additional filtering functionality. Categories and sub-categories are now displayed in association with the number of Terms belonging to them. When the user selects a category, its Terms are still displayed in the middle part, but panel 1 on Fig. 4 is also updated with the list of *other* categories these Terms can belong to, and the number of overlapping terms. For example, if a user selects the Category **Military Issues**, the terms related to **Military Issues** are displayed,

⁸ Respectively at the URL <http://www.vandale.nl/opzoeken/woordenboek/> and <http://www.muiswerk.nl/WRDNBOEK/INHOUD.HTM>.

and all other categories in which the displayed terms also appear are proposed for narrowing down the number of terms. If the user selects also **Traffic and Transportation**, he will get the list of military vehicles in the thesaurus. He can narrow down his query even further by selecting **Vessels**, in which case the list is narrowed down to military vessels. The number of terms to be displayed can thus be narrowed down to a dozen by two or three clicks. It is a kind of faceted search, but on the term level⁹.

BroaderTerm/NarrowerTerm Display (Number 3 on Figure 4). We solved a problem of ergonomy in changing the display of the **BroaderTerm/NarrowerTerm** tree. In the previous version, displaying the tree could lead a bad display of the other information about a given Term.

Cross-Facet Links. We extracted some information provided in the scope notes of the **People** facet to generate cross-facet links: if a scope note states that a person has a specific occupation, say **King**, and if this occupation is in the **Subjects** facet, then we generate a browsable link between the person and the subject **Kings**. This helps the user to browse directly (potentially) other relevant parts of the GTAA than the current facet.

5 User Study Part II

5.1 Setup

The second user study targeted evaluation of the adapted version of the GTAA browser according to the same measures and experimental goals as in the first study. In total seven subjects, six from Sound & Vision and one from NOS, followed the same experimental design and procedure (with an improved explanation form). None of the subjects participated in the first study.

Results. Table 2 shows the results for the second user study. We observed the following:

- There is a decrease in the number of problems (2.14 vs. 3.67). Thus, there is a clear indication of a higher level of effectiveness.
- The time used for play-around is longer (25 min. vs. 18 min.). This is logical because the guidelines were more elaborate and the complexity of the search increased. So, a longer learning curve is needed.
- The total time spent during annotation decreased (16 min. vs. 27 min.)m as well as the total number of steps per session (25.57 vs. 33.22) and the number of steps per index term (3.74 vs. 4.87). This indicates an increase in the efficiency of the search. As the minimum number of steps per index term is 3 (see Sec. 3), the result in the second study is actually approaching maximum efficiency.

⁹ As opposed to the document level, for which facets would be the broadcasting date, the genre, etc. on top of the different controlled terms used as metadata.

Table 2. Results for the second part of the user study. Time measurements are in min.

Stage/Metric	Subjects							
	S10	S11	S12	S13	S14	S15	S16	Average
During play-around								
#problems	1	3	3	4	3	0	1	2.14
total time spent	29	24	26	16	26	33	22	25
During annotation								
#problems	1	2	2	1	3	2	3	2.00
#alphabetical search used	7	9	9	12	6	9	5	8.14
#hierarchy search used	4	1	0	2	1	1	2	1.57
#filter search used	2	2	0	0	5	0	0	1.29
#steps	26	25	28	28	37	18	10	24.57
#resulting index terms	6	7	6	7	5	9	6	6.57
#steps per index term	4.33	3.57	4.67	4.00	7.40	2.00	1.67	3.74
total time spent on search	12	9	19	14	10	10	15	13
total time spent	15	12	22	19	13	14	18	16

- The average time used for search was 13 min. This is close to 80% of the total annotation time. In comparison: in the first study it was 67%.
- The use of alphabetical search increased slightly (8.14 vs. 7.00), while at the same annotation time went down. This means we achieved at least partially the goal of making alphabetical search more effective (see previous section).
- The hierarchy search increased marginally (1.57 vs. 1.33); the filter search marginally decreased (1.29 vs. 1.44). The number of times these functionalities were used prevents any generalization. Our hypothesis is that due to the improved alphabetical search there was no real need for the other search types. In this context it is worthwhile to point out that the subjects were used to alphabetical search already, and had little to no experience with other search types. The Categories were not displayed in their previous annotation tool, and thus these groups of terms were not yet used in a real-life annotation task. They were not yet adapted to fit this task, contrary to the thesaurus content, which is updated on a daily basis. The Categories proved to be too broad to enhance a browsing type of search.

Table 3 shows the aggregated results of the user-satisfaction questionnaire. We observe here a marginal increase in the satisfaction of the users with respect to the general browser functionality, the subject-facet functionality and the search functionality, as well as a marginal decrease in the satisfaction for the browser functionality. The differences in the aggregated values of the first and second studies are too small to be able to make any generalization. However, some of the values on individual questions support the hypothesis that the satisfaction of the users increased with the adapted browser, while the number of steps required to perform the task decreased.

While the users in the first study were doubting the usefulness of the browsers hierarchical structure (3.11), the participants in the second study show strong

Table 3. Results of the questionnaire about user satisfaction in both studies. Results were aggregated per question group (left column).

Question group	Average score 1 = lowest, 7 = highest	
	User study 1	User study 2
General browser functionality	4.79	5.26
Subject-facet functionality	4.64	5.10
Search functionality	4.82	4.87
Browse functionality	5.02	4.57

consensus that the thesaurus structure in the adapted browser helped them discover related terms (5.29) and the relationships between them (5.17). Most of the users in the first study preferred to use the alphabetical search above the hierarchical one in the “Subject”-facet (6.44), where in the second study we can see a clear change in a positive direction (5.00) although still preferring the alphabetical search to the hierarchical. The level of complexity in the hierarchical search was appreciated more by the users of the adapted browser in the second study (5.43 vs 4.48), as well as using hierarchical structure in combination with the search (4.71 vs 4.13). Many of the users both in the first and in the second test were not happy that it took too long to find the appropriate main category in the hierarchy (2.82 vs 2.86). This comment is mainly concerning the broadness of these groupings. Further, there were no significant improvements in the hierarchy presentation in the browser, thus no major changes were expected, as shown from the previous values.

6 Discussion and Related Work

This study shows some insights of the use of knowledge structures like thesauri in application settings. The cataloguers were used to quite basic tools for finding index terms and were dazzled by the complexity of the browser interface. They are used to alphabetical search and therefore we gained most performance value by optimizing this part of the search, as can be seen in the second user study. For searching a hierarchical representation is apparently not of much value. However, for disambiguating terms, showing the respective places in the hierarchy could be a quick means for selecting the right concept. We could also notice a difference of strategies between the subjects from Sound and Vision (experts of the thesaurus content) and from broadcasting corporations. The later were more eager to start searching a term by browsing. Thus, such a browsing facility could be helpful to the general public for searching the public Website of Sound and Vision.

Despite the learning curves (see the times needed for the play-around sessions), the cataloguers were in general positive about the use of such a tool in their daily work. This is apparent from the questionnaire, but also from the fact that,

based on the results of this study, Sound & Vision is seriously considering of incorporating the thesaurus browser in their archiving process.

Other studies evaluate thesaurus browsers by user studies, but they usually focus on the task the thesaurus helps achieving, and not on the thesaurus usability and functionalities themselves. Several authors [8,6,1] have considered the selection of a term as a particular part of their evaluation, but they evaluate it against the recall or precision of documents retrieved. Blocks [1] explicitly stresses the fact that, as their interface enables query expansion on the basis of the **NarrowerTerm** relationship in the thesaurus, the tendency of users to search for the most specific query term is a waste of time: the set of terms that they choose for formulating the query would be taken into account with a query involving their common hypernym, by which the users started browsing the thesaurus in the first place. Our purpose is the opposite: making sure that the browser proposes relevant functionalities for different search strategies in order to retrieve the *most specific and relevant term* for indexing a document. The perspective of evaluating a tool dedicated to helping the selection of a keyword for *indexing* has not been taken into account very often, and this indexing task has specific requirements.

As mentioned in the introduction, the thesaurus browser is just a small piece in the larger puzzle of supporting semantic annotation. We see it as a baseline tool for cataloguers, who may always have the need to do some manual work on annotation. The majority of the research is aimed at providing automatic tools for generating candidate indexing terms [3]. In the digital archiving process of the future we expect the emphasis to lie on semi-automatic annotation, with the role of the cataloguer shifting to the person who performs quality control on suggested indexing terms and/or selecting the most appropriate ones from the terms suggested. Also this process would benefit from a usable thesaurus browser. The fact that the tool is based on the RDF/OWL specification makes it a good candidate for reuse with other RDF/OWL-based thesauri. This is in fact a realistic extension. Many institutions still rely on their in-house thesaurus, but would benefit from larger a wider scope of thesauri [5]. For example, geographical data in GTAA are likely to be incomplete and it might be a better approach to use geo-spatial data from other sources, such as the Getty Thesaurus of Geographical Names¹⁰.

Acknowledgments

The authors are grateful to the 16 participants in the experiments from Sound & Vision, NOS and EO. Johan Oomen provided useful comments. This research is supported by the CHOICE, which is part of the DUTCH CATCH Programme funded by the Dutch National Research Foundation NWO. The evaluation of the GTAA browser was performed as a usability case study at LaQuSo Lab, Eindhoven University of Technology.

¹⁰ http://www.getty.edu/research/conducting_research/vocabularies/tgn/

References

1. Dorothee Blocks. *A qualitative analysis of thesaurus integration for end-user searching*. PhD thesis, University of Glamorgan, Hypermedia Research Unit, School of Computing, UK, 2004.
2. P. Borst, J. M. Akkermans, and J. Top. Engineering ontologies. *Int. J. Human-Computer Studies*, 46:365–406, 1997.
3. L. Gazendam, V. Malaisé, G. Schreiber, and H. Brugman. Deriving semantic annotations of an audiovisual program from contextual texts. In *Semantic Web Annotation of Multimedia (SWAMM'06) workshop, held in conjunction with WWW'06, Edinburgh, UK*, 2006.
4. S. Handschuh and S. Staab. Annotation of the shallow and the deep web. In S. Handschuh and S. Staab, editors, *Annotation for the Semantic Web*, volume 96 of *Frontiers in Artificial Intelligence and Applications*, pages 25–45. IOS Press, Amsterdam, 2003.
5. L. Hollink, A. Th. Schreiber, J. Wielemaker, and B. J. Wielinga. Semantic annotation of image collections. In S. Handschuh, M. Koivunen, R. Dieng, and S. Staab, editors, *Knowledge Capture 2003 – Proceedings Knowledge Markup and Semantic Annotation Workshop*, pages 41–48, 2003.
6. Eric H. Johnson and Pauline A. Cochrane. A hypertextual interface for a searcher's thesaurus. In *Proc. Second Annual Conference on the Theory and Practice of Digital Libraries (Digital Libraries 95), Austin, Texas, USA, June*, pages 77–86, 1995.
7. A. Miles and D. Brickley. SKOS core guide. Technical report, W3C Working Draft, 2 November 2005. <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102>.
8. Heiner Stuckenschmidt, Anita de Waard, Ravinder Bhogal, Christiaan Fluit, Arjohn Kampman, Jan van Buel, Erik van Mulligen, Jeen Broekstra, Ian Crowlesmith, Frank van Harmelen, and Tony Scerri. A topic-based browser for large online resources. In E. Motta and N. Shadbolt, editors, *Proc. 14th Int. Conference on Knowledge Engineering and Knowledge Management (EKAW'04), Northamptonshire, UK, 5-8 October*. Springer-Verlag, 2004.
9. M. van Assem, V. Malaisé, A. Miles, and G. Schreiber. A method to convert thesauri to SKOS. In *Proc. Third European Semantic Web Conference (ESWC'06), Budvar, Montenegro*, June 2006. Accepted for publication. <http://www.cs.vu.nl/~mark/papers/Assem06b.pdf>.