TIDES Portal Evaluation

The MITRE Corporation Evaluation Working Group

September / October 2000

Table of Contents

TIDES Portal Evaluation
by the MITRE Evaluation Working Group

Laurie Damianos and Jill Drury

September/October 2000

Executive Summary

A formative evaluation was performed to elicit users' reactions to the TIDES Portal and to gather information on potential utility. The TIDES Portal is an early prototype of the DARPA Translingual Information Detection, Extraction, and Summarization (TIDES) program. The TIDES Portal was designed to deliver usable information on demand from multiple sources and media and in multiple languages. It allows users to search current data from television, radio, electronic news, and digital libraries. Multilingual sources are also searchable, with both input search terms and output results in English.

The TIDES Portal consisted of an integrated suite of five search engines, a collaborative presentation tool, and a clipboard for storing selected documents. BBN's Oasis automatically segmented radio news into spoken documents that could be navigated and searched. MITRE's Broadcast News Navigator automatically captured, annotated, segmented, and summarized television news; users could browse the digitized news, view summaries, read the closed caption text, and view the actual video. Lincoln Labs' Common Coalition Language System was a translingual information system which retrieved and summarized both Korean and English documents based on English language queries. The Foreign Broadcast Information Service tool provided worldwide open-source news translated from multiple languages. MITRE's Web Harvester collected on-line newspapers, in particular, the International Herald Tribune. Global Infotek's TIDES Situation Book allowed the user to create and share a report by dragging material into a collaborative repository and annotating it. A clipboard was built into the TIDES Portal user interface to allow users to store pointers to useful documents by dragging the document icons into the reserved space.

The evaluation of the TIDES Portal focused on finding answers to the following high-level questions:

- What could users do with the TIDES Portal and would they find it useful?

- Would the TIDES Portal have any advantages over other means of information retrieval, such as an Internet search engine?

The Evaluation Working Group designed a within-subjects, counterbalanced experiment in which each participant was given two sets of two questions to answer using the TIDES Portal and their favorite Internet search engine. Each question set consisted of one fact-based and one open-ended question pertaining to the North Korean Ballistic Missile Program. In the first part of the experiment, each participant was asked to answer one set completely using the TIDES Portal and then to choose one of those questions and verify the answer via another search engine. In the second part of the experiment, the user was asked to answer a similar set of questions with a search engine and then to verify one answer using the TIDES Portal. We alternated the order in which the question sets were given but always started the experiment using the TIDES Portal.

We collected both quantitative and qualitative data throughout the experiment. Server logs and browser logs automatically captured time stamped events, URLs, and search terms. Observers were recorded comments and critical incidents. Participants provided background information via questionnaires, and more feedback was gathered at the end of each session during an oral interview.

Users liked knowing the source of the data in the TIDES Portal because they could trust those data. The variety of information available and the physical separation of data sources in the display were also mentioned as positive features

of the TIDES Portal. Users also liked the highlighted named entities and list of keywords in Oasis, and they liked the summary and named entities in BNN.

Users remarked that the TIDES Portal would be useful for specific tasks like the one they were asked to perform in this evaluation. Not one said that he/she would use it in his/her work, but they all envisioned it to be useful to someone performing intelligence analysis, tracking events, information gathering and organizing, data mining, and looking for trends.

Although the focus of this evaluation was on the utility of the TIDES Portal, the participants had difficulty ignoring the usability issues. In particular, the slowness of the systems and the under-developed user interface were obstacles to the participants while performing the given task.

Overall, participants scored higher on task completion using their favorite Internet search engine than when using the TIDES Portal. Only half as many questions were answered via the TIDES Portal. In general, users found answers to questions more quickly via their favorite Internet search engine because of the textual clues provided in the list of documents. For example, some search engines show the first two lines of text associated with each document, and some show a relevant piece of text with the search terms embedded. The TIDES Portal did not provide any clues, and users were forced to examine each retrieved document – a slow process due to performance issues. In many cases, users gave up on trying to find answers to questions in the TIDES Portal.

The display of document proxies presented other problems. Titles were not entirely visible; some wrapped within the pane, others were viewable only through horizontal scrolling which users found annoying. Users also had trouble resizing the panes to view the titles.

Several users had difficulty keeping track of which documents they had already examined and opened a few documents more than once. They wanted the ability to "cross off" the documents they had reviewed or have some other sort of indication that the link had been visited.

This feedback will be useful in designing the next iteration of the TIDES Portal interface. Clearly, this experiment shows that we must spend sufficient time on usability so that the underlying utility of the new services can be properly assessed.

Overview

This formative evaluation effort was designed to elicit users' reactions to the TIDES Portal and to gather information on the integrated tools' potential utility. For various reasons discussed below, the experiment produced usability feedback as well as data for a utility analysis.

The TIDES Portal was an early prototype of the DARPA Translingual Information Detection, Extraction, and Summarization (TIDES) program. The TIDES Portal was designed to deliver usable information on demand from multiple sources and media in multiple languages. The TIDES Portal allows a person to search current data from television, radio, electronic news, and digital libraries, including multilingual sources, with both input search terms and output results in English.

The TIDES Portal Components

The TIDES Portal consists of an integrated suite of five search engines, a collaborative presentation tool, and a clipboard for storing selected documents. Each of the components is described in detail below.

Oasis

BBN's Oasis uses speech recognition combined with text mining tools to provide video or other collaborative audio that can be navigated and searched. The user interface presents transcribed material segmented by speaker and topic, and named entities are highlighted within the transcribed text. Retrieval is based on any of the content-based features. Playback of the audio material is also available at various levels.

Broadcast News Navigator

MITRE's Broadcast News Navigator (BNN) is a web-based news-on-demand system that automatically captures, annotates, segments, summarizes, and presents stories from broadcast news video. The tool was developed to allow a user to collect and analyze stories of interest from a large, diverse repository of news sources, assisting the user in information gathering and trend analysis. Users can browse the digitized news, view summaries, read the closed caption text, and view the actual video.

Common Coalition Language System

Lincoln Laboratory's Common Coalition Language System (CCLINC) is a translingual information system which facilitates human-human interactions via two-way English-Korean translation of text and speech, and human-machine interactions via (1) translingual information retrieval on Korean and English documents based on English queries, (2) Korean to English translation of the retrieved documents, and (3) translingual question answering.

CCLINC is based on cutting edge natural language understanding, generation, and speech recognition technology, and has been configured under the DARPA Communicator architecture.

Foreign Broadcast Information Service

The Foreign Broadcast Information Service (FBIS) tool provides worldwide open-source news translated from multiple languages.

Web Harvester

MITRE's Web Harvester collects on-line newspapers and news sources such as the International Herald Tribune and ProMed. For this experiment, data was available from the International Herald Tribune only.

TIDES Situation Book

Global Infotek's TIDES Situation Book (TSB) is a collaborative application that provides notebook representations of organized data pertaining to a particular topic. The TSB allows the user to drag or insert multimedia material into a notebook-like representation to create a highly organized presentation of the topic. The TSB interface allows the user to create chapters and pages in the notebook and annotate the data placed within.

Clipboard

A clipboard was built into the TIDES Portal user interface to allow users to store pointers to useful documents by dragging the document icons into the reserved space. Clipboard contents can be saved and loaded for later use.

The Experiment

Since the TIDES Portal prototype was built to showcase and integrate the different tools, user interface design was not a high priority for the development team. Therefore, we did not design this experiment as a classic usability evaluation and instead were more interested in getting reactions of representative users to the utility of the TIDES Portal.

In particular, we wanted to find answers to the following high-level questions:

- What could users do with the TIDES Portal?

- Would the TIDES Portal have any advantages over other means of information retrieval, such as an Internet search engine?

- Would users find the TIDES Portal useful?

Design

We designed a within-subjects, counterbalanced experiment in which each participant was given two sets of two questions to answer using the TIDES Portal and their favorite Internet search engine. Each question set consisted of one fact-based and one open-ended question pertaining to the North Korean Ballistic Missile Program. In the first part of the experiment, each participant was asked to answer one set completely using the TIDES Portal and then to choose one of those questions and verify the answer via another search engine. In the second part of the experiment, the user was asked to answer a similar set of questions with a search engine and then to verify one answer using the TIDES Portal. We alternated the order in which the question sets were given but always started the experiment using the TIDES Portal.

Development and Rationale

To answer the high-level questions about the TIDES Portal utility listed above, we modeled a scenario demonstrated during the Strong Angel exercise in RIMPAC 2000 in which a user created a report based on intelligence analysis of the North Korean Ballistic Missile Program. We trained a set of MITRE employees to use the TIDES Portal and its tools, asked them to perform a task similar to the Strong Angel demonstration scenario, and gathered feedback.

Ideally, we would have preferred to have people retrieve information using the TIDES Portal and to compare their performance with a different baseline group using an Internet search engine (a between-subjects design). However, it was difficult to structure the experiment as a between-subjects evaluation because of multiple subject differences such as familiarity with an Internet search engine versus the TIDES Portal, knowledge of the task domain, experience in gathering/analyzing open source intelligence data, etc. Even isolating just one of those variables, e.g., familiarity with search tools, it would still be difficult to run a controlled experiment because training on the use of the TIDES Portal could never approach parity with the many hours people have likely spent using a search engine. This inability to eliminate confounding effects drove the need for a counterbalanced, within-subjects study.

Task Development

To simulate the demonstration scenario, we considered asking the participants to create a report on a specific topic. An initial idea was to have the users create the end report in the TIDES Situation Book and compare that to the end report in Microsoft PowerPoint. We quickly realized that the additional training needed for the TIDES Situation Book would make each experiment session too long. Even just creating a report in Microsoft PowerPoint would tempt users to spend time making the output attractive rather than concentrating on the process of extracting information from the TIDES Portal. To avoid distracting users from the "IDES" components of the TIDES Portal, we asked users to produce a simple Microsoft Word document as their end product.

Asking users to answer structured questions using the TIDES Portal and an Internet search engine was a good way of exercising both types of information retrieval systems without having to control for user creativity. It was difficult to formulate specific questions on a topic that an intelligence analyst might research and that would require information that was accessible to both (or neither!) of the systems. We eventually chose a topic that was used as part of the demonstration scenario: the North Korean Ballistic Missile Program and developed a list of ten fact-based and open-ended questions.

A pilot test (see below) showed that it took the participant a long time to answer just a few questions. Therefore, we reduced the number of questions so that each participant would answer two sets (Part A and Part B) of two questions using both their favorite search engine and the TIDES Portal. The questions were balanced so that one question in each set was fact-based, and the other was open-ended. We counterbalanced the order in which the question sets were answered. Always giving the TIDES Portal question set first immediately after the training helped reduce the number of trials needed to get a fair distribution. Alternating question sets helped eliminate the extremely difficult task of devising pairs of questions where information was equally accessible via Internet search engines and the TIDES Portal.

We decided not to require the participants to answer all the questions so that we could stop the participants after some arbitrary time limit or at some point of diminishing returns. We found that ten minutes per question was an appropriate amount of time, and if users could not answer the question within that limit, it was clear that they were frustrated and were ready to move on.

Pilot Study

We ran a single pilot trial to ensure that questions were comprehensible and answerable. The pilot user found all answers in both the TIDES Portal and by using Google. The pilot study helped reduce the question sets so that the experiment sessions could be run in no more than 1.5 hours total. Comments and feedback made by the pilot participant also helped debug our questionnaires.

Number of Trials

After running the pilot study, we knew we would not need to run a large number of trial sessions. Because the TIDES Portal ran so slowly and the user interface was not completely developed, we were prepared to get more anecdotal data than quantitative data. We decided to run four trials (two with Part A followed by Part B second, and two with the order reversed) because it was clear that we would get little additional information from a larger study.

Experiment Execution

Of the four trials, two started with question set A and two started with question set B. Each trial session lasted not more than 1.5 hours including training, task execution, and questions.

Participant Population

The five volunteers in this experiment were all MITRE employees, chosen randomly. MITRE employees were judged acceptable as surrogate users because, similar to the expected user population of open-source intelligence analysts, they are extremely computer-literate and have experience doing data searches using Internet search engines.

The participants consisted of two senior artificial intelligence engineers, a senior human factors engineer, a software systems engineer, and a technical co-op student. None could speak or read Korean, but three had some experience investigating the North Korean Ballistic Missile Program and/or gathering/analyzing open-source (non-secret) intelligence data.

All users had at least one Internet search engine preference including Alta Vista, Google, Yahoo, Hotbot, dogpile, Ask Jeeves, Ask.com, and webcrawler.

Setup and Trials

A dedicated PC was used for the evaluation. Before each experiment session, we launched Internet Explorer 5.0 and logged into the TIDES Portal site. Both sets of questions were opened in Microsoft Word files which were then placed in the background.

Upon arrival, users read and signed a consent form. They were then asked to complete a one-page questionnaire designed to gather background information on job title, experience in intelligence analysis, knowledge of the task topic, and familiarity with computer platforms, browsers, and Internet search engines.

We read an overview of the TIDES Portal to the users and guided each user through a hands-on exercise consisting of a basic two-term search with an AND conjunction. The users were able to examine documents retrieved through each tool, and we discussed the type(s) of data available. We also showed the participants how to use the Clipboard and suggested that they store relevant documents there for subsequent use. We explained how to widen the panes, use the horizontal scroll bars, and view up to 30 documents (by displaying sets of 10 at a time).

Users were given a guide sheet with a simple description of each tool and encouraged to ask questions and make comments about the TIDES Portal and each of the component tools.

We read task instructions to users and gave them hard copies of the question sets along with the Microsoft Word versions of the same question sets. We asked the users to find evidence supporting their answers to each question even if they already knew the answer. To ensure an actual search, the users had to list the URL(s) associated with documents containing the answers.

All users first had to answer the questions using the TIDES Portal and then verify a single answer with an Internet search engine of their choice. They then answered the second set of questions via the search engine and then verified one of those answers with the TIDES Portal. If a user was not able to answer a question within 10 minutes, he/she was asked to continue on to the next question. The user was also permitted to give up at any point. In addition, we did not restrict the user to answering the questions within a set in any particular order.

Answers to questions were captured in the Microsoft Word documents. Users could type or paste information and were asked to provide the name of the tool used and the URL of the document with the correct answer. There was flexibility in the answers, e.g., a date could either be absolute (February 20, 1999) or relative (a year ago).

When the question sets were completed, we verbally gave another questionnaire designed to elicit feedback on the TIDES Portal and discussion of its utility. Due to the open-ended nature of the questions, we also gathered feedback on usability.

Methodology: Metrics and Data Collection

From the high-level view, we wanted to find out the following:

- What did the user do?
- How did the user accomplish the task?
- What TIDES Portal features were used and how?
- How well was the task accomplished? (Were all questions answered, and were they answered correctly?)
- Were participants able to answer questions faster using the TIDES Portal?
- What advantage did the TIDES Portal have over an Internet search engine, if any?
- Did the user find the TIDES Portal useful?
- For what other applications could the TIDES Portal be useful?

Metrics

The metrics collected related to the TIDES Portal usage as well as to the task performed by the participants.

Metrics Relating to features of the TIDES Portal

- tools usage
- frequency of tool usage (estimated)
- utility of each tool relevant to task (based on user feedback plus contribution to task)
- clipboard usage

Metrics Relating to Task

- time to completion (because of the TIDES Portal performance issues, this metric was not particularly relevant)
- quality of outcome (# of questions answered)
- path to outcome (# of searches performed and # of documents examined)

Data Collection

This evaluation focuses on both quantitative data as well as qualitative, anecdotal reactions.

Quantitative Data Collection

Due to time constraints, lack of access to source code, and lack of resources, we did not attempt to instrument any of the tools for automatic data collection. Instead, we took advantage of whatever automatic methods were already available to us.

TIDES Portal server logs tracked queries made via the TIDES Portal (including search terms) and also provided minimal data on use of the clipboard. The Internet Explorer History provided a rough estimate of queries made via the TIDES Portal as well as via other search engines. The History also recorded which documents were examined and when. Unfortunately, the History logs stored information (time stamp, URLs, and search terms) for the latest hit only. In other words, if a user made the same query twice with the same search engine or looked at the same document more than once, the latest event would overwrite its duplicate and all information about similar activities done previously would be lost.

To supplement automatic data collection, we observed the users while they made queries and examined documents for appropriate answers. We recorded the search terms for each query and kept a rough estimate of time on subtask (how much time was spent answering each question, etc.)

The Microsoft Word documents were used to capture answers to each question as well as which tools were used and the URLs of the documents that provided each answer.

In addition, the pre-experiment questionnaires provided some quantitative background information on the participants. We asked them about their experience using browsers and search engines as well as what they knew about gathering open-source intelligence data and any prior knowledge of the task topic. (See the appendix.)

Qualitative Data Collection

During the experiment, we observed the users and made notes of critical incidents related to both the usability and utility of the TIDES Portal. We also wrote down the participants' comments and questions.

Oral interviews, based on open-ended questions, gathered data targeted at the users' reaction to the TIDES Portal and potential applications for its use. We also asked the users to make suggestions on improving the TIDES Portal, both in terms of usability and functionality.

Data Preparation and Analysis

Quantitative Data Preparation and Analysis

Quantitative background data from the pre-experiment questionnaires were tabulated.

The TIDES Portal server logs and the Internet Explorer History were imported into Microsoft Excel spreadsheets and annotated with high-level descriptions of events (e.g., 'user did Google search,' 'user opened FBIS document,' 'user looked at BNN closed caption text,' etc). These data were then merged with observations made during the experiment to provide a more complete picture of user activities.

We note that these combined data are not totally accurate and are certainly not complete. We had no reliable way of recording which TIDES Portal documents were examined or even how many were examined since these data are not

captured in the server logs. The Internet Explorer History captured some of this information, but as stated above, only the latest visit to a particular URL was stored. In addition, times were not recorded at the same precision (e.g., observer's watch versus computer time stamp), and clocks were not synchronized (the TIDES Portal server clock was off by several hours).

We did not investigate additional server logs of the usage of individual TIDES Portal tools because not all were available to us and those that were available did not have standardized formatting or contents (e.g., different levels of events, different levels of details). The value added by analyzing the available subset of these logs was not worth the time or effort.

Correct answers were counted. (Basically, any question answered was considered correct because of the requirement to provide supporting evidence from one or more documents.) Combined with data from server logs and observations, we looked at how many queries were executed in order to answer each question and approximately how many documents were examined for each question. We compared these results using the TIDES Portal versus using another search engine. We also looked at which TIDES Portal tools were used and how often.

Qualitative Data Preparation and Analysis

Observations made during the experiment trials and comments made by the participants were typed up and combined with feedback from the post-experiment interview / questionnaire. These data were then compiled and organized by topic.

Results and Discussion

This section begins with a discussion of the biases in this experiment. It then addresses each of the original high-level questions regarding the TIDES Portal utility, followed by our findings, in greater detail, on overall utility and usability.

Biases

A number of biases most likely influenced the results and analysis of this evaluation. Below, we briefly describe several biases: performance, user interface, data, domain/task, and training.

Performance Bias

The TIDES Portal performance was exceedingly slow, primarily due to the research nature of the tools and congested network traffic. It took several minutes for all the TIDES Portal tools to retrieve and display documents and often minutes to display the contents of a single opened document. The speed issue essentially rendered useless some of the quantitative data we collected such as time-on-task measurements.

User Interface Bias

Because the TIDES Portal user interface was not the focus during Portal development, a number of features were not fully designed or not fully implemented. Deficiencies included poor display given the limited real estate, partially completed Clipboard functionality, no available help, inability to view more than 30 documents from any tool, etc. There were also actual bugs, e.g., CCLINC returned a random set of documents independent of the query, automatic launching of search when the user went back to the search page, automatic resetting of the date range drop down box after each search, etc.

Data Bias

The TIDES Portal data reflected a relatively narrow time window, with some documents dating from 1999, but most were from January 2000 through March 2000. The data retrievable via the Internet were virtually unlimited by such time constraints. In addition, the translated versions of Korean documents retrieved via CCLINC were so poor that little or no information was usable.

Domain / Task Bias

While attempting to create a task similar to the Strong Angel demonstration scenario, we also tried to make the task comprehensible to participants untrained in intelligence analysis. However, this type of research and the domain itself were both somewhat unfamiliar to the participants. Had a similar experiment been performed with intelligence analysts as users, we might have obtained different results.

Training Bias

We provided 10 to 15 minutes of training on the TIDES Portal. We felt this was enough time to learn the basics in order to accomplish the task. We also provided an overview of the types of data sources each component searched and were available to answer any questions the participants might have. However, use over time would have helped the user become more familiar with the tools and the types of data available through each.

High-Level Questions

What did the user do?

The user attempted to answer each question by performing an appropriate query and searching the resulting documents for relevant information. More questions were answered via an Internet search engine than through the TIDES Portal. In general, users found it easier to scan through a list of documents found via their favorite Internet search engines because of the textual clues provided. For example, some search engines show the first two lines of text associated with each document, and some show a relevant piece of text with the search terms embedded. Users had a harder time viewing the document titles in the TIDES Portal interface, and, since no other textual clues were provided, the users had to open and scan the documents in order to understand what they were about.

What features were used and how?

Because of the small sample size and limitations on logging capability, not much can be said about this. When searching for answers to the questions, users looked at documents from all the TIDES Portal tools. Documents retrieved from each tool provided at least one answer to each question, with the exception of Oasis. Participants preferred some tools over others, and these results are discussed in the Utility section below.

The TIDES Portal Clipboard was used only once by one of the participants. In the post-experiment interviews, most participants said they liked the Clipboard and would use it if they were doing ongoing research on a particular subject. However, they did not feel the need to use it while doing this particular task.

How well was the task accomplished? Were all the questions answered, and were they answered correctly?

Since the participants were not able to answer the questions unless they had some sort of supporting data from a document, we considered all answered questions to be answered correctly.

Only one user was able to answer all four questions correctly, but he was not able to verify any of his answers using the TIDES Portal. Only two unique questions were answered by all four participants, one fact-based and one open-ended. Overall, a total of 5 questions were answered using the TIDES Portal (4 questions plus 1 verification), and 10 questions were answered using an Internet search engine (6 questions plus 4 verifications). This could be an indication of question bias (some questions were not as easily answered in the TIDES Portal).

In fact, using the TIDES Portal, participants were able to answer 3 of the 4 different questions. Participants were able to answer all questions using an Internet search engine. [Note: the one question not answered via the TIDES Portal was answered by just one of the participants using an Internet search engine.]

Were participants able to answer the questions faster using the TIDES Portal?

No. Not only were participants not able to answer the questions faster using the TIDES Portal, but participants were not able to answer as many questions using the TIDES Portal. Tool performance was certainly a factor. Because participants were limited to 10 minutes per question and did not answer as many questions in the TIDES Portal, we cannot conclude that tool speed alone was the reason participants could not answer questions faster using the TIDES Portal.

What advantage did the TIDES Portal have over an Internet search engine, if any?

Users liked knowing the source of the data because of trust issues. The variety of information available and the segregation of search engine output were also mentioned as positive features of the TIDES Portal. Users also liked the highlighted named entities and list of keywords in Oasis, and they liked the summary and named entities in BNN. Participants liked the idea of having a Clipboard, but they made suggestions for improvement in design.

Did the user find the TIDES Portal useful?

The users would find the TIDES Portal useful for specific tasks such as the one they were asked to perform in this evaluation. Not one said that he or she would use it in his/her work, but they all envisioned it to be useful to someone performing intelligence analysis. One of the participants thought that he might use the TIDES Portal to search on political issues.

For what other applications could the TIDES Portal be useful?

According to our participants, an analyst could use it to track events. It would also be good for information gathering and organizing, data mining, and looking for trends. Another suggestion was to use it in the battle management cell in the US/Canadian ground based air defense system.

Utility

Utility to Particular Task

As mentioned in the previous section, users were able to answer more questions using Internet search engines than using the TIDES Portal. Overall, 5 questions were answered using the TIDES Portal (4 original questions, 1 verification), and 10 questions were answered using an Internet search engine (6 original questions, 4 verifications). See Table 1, below.

This could be an indication of question bias (some questions were not as easily answerable in the TIDES Portal). Note that results are not significant in this study, due to small sample size.

Only one user (*user4*) was able to answer all four questions. Only two questions were answered by all four participants, one fact-based and one open-ended. All questions were answered using some Internet search engine, but 3 of the 4 different questions were answered using the TIDES Portal. The one question that could not be answered via the TIDES Portal was answered via Google by just one participant.

Search tool(s)	successes	failures	success rate
TIDES Portal	5	7	42%
Overall Internet search engine	10	2	83%
Google	5	1	83%
Yahoo	2	1	67%
Hotbot	3	0	100%

Table 1 Success rate at answering questions in the TIDES Portal versus an Internet search engine.

user1 could not answer any questions using the TIDES Portal but was able to verify one using the TIDES Portal. Two users could not make any verifications using the TIDES Portal, but were able to answer one of the questions using the TIDES Portal.

Users found it harder to answer the second set of questions (Part B) than the first set of questions (Part A). See Table 2, below. They had to do multiple searches and look in multiple documents to find the "embedded" answers. One user requested a better, faster search engine with more advanced features.

Task question	# times answered		# times verified	
	TIDES Portal	Internet search engine	TIDES Portal	Internet search engine
Part A, Q1	2	2	1	2
Part A, Q2	1	1	0	0
Part B, Q1	0	1	0	0
Part B, Q2	1	2	0	2

Table 2 Frequency of answering and verifying questions in the TIDES Portal versus an Internet search engine.

Users had the most difficulty with Part B, Question 1, a question about the cost of a specific missile. The one user to answer this question performed multiple queries on synonyms for the word "cost" before he was able to find the answer. The other participants stated that this type of information would not generally be found in a news story.

From Table 3, it appears that responses to questions came from a random spread of the TIDES Portal tools. Oasis was never used to provide the final answer, but we cannot draw any conclusions from such a small sample size. We did not verify that each question could be verified by each TIDES Portal tool, only that each question could be answered by at least one TIDES Portal tool.

TIDES Portal Tool	# times tool provided final answer
Oasis	0
BNN	1
CCLINC	2
FBIS	1
Web Harvester	1

Table 3 TIDES Portal tools used to provide final answers to questions.

Overall, there was no difference in the average number of queries performed to answer questions in either the TIDES Portal or via an Internet search engine. Table 4 shows that users performed approximately 1.4 queries per question. These averages are approximate since the data are incomplete.

Search tools(s)	(approximate) average # queries per question	(approximate) average # documents examined per question
TIDES Portal	1.4 / (1.85)	2.25 / (5.5)
Overall Internet search engine	1.4	2.2
Google	1.75	3
Yahoo	1	unknown / (3)
Hotbot	1	1

Table 4 Average queries performed and documents examined per question. The numbers in parentheses pertain to questions unsuccessfully answered within 10 minutes; all other numbers are based on successfully answered questions. The numbers in this table are approximate because complete data were not available for every question and every user.

Although the table shows that, on average, users opened almost the same number of documents in the TIDES Portal as they did in Internet search engines to find an answer, these data obscure what was really happening. Participants examined more documents in the TIDES Portal before finding an appropriate answer but often looked at multiple documents in the search engine to verify what they had already found because opening documents was so much faster. Another reason was that the search engines displayed the retrieved documents either with the first few lines of text or with a chunk of text with the search terms highlighted. The tools in the TIDES Portal displayed just the document title which was hard to read in the narrow pane. Participants had to use the horizontal scroll bar to read the titles or struggle to resize the pane. (As expected from Fitt's Law, which predicts the time for users to select a target given the target's size and distance from the cursor, it was difficult and time-consuming for some users to resize panes.) Since there was little clue as to documents' contents, users had to open each document to discover its relevance.

Participants were generally able to find an answer using the Internet search engines after looking at just over two documents. We know little about the number of documents examined via an Internet search engine before a user failed to answer a question. However, participants looked at approximately 5.5 documents before either giving up or running out of time. Since the success rate of answering questions in the TIDES Portal was so much lower than that of the Internet search engines (42% versus 83%, see Table 1), it appears that simply providing information in documents is not enough; making the information easily accessible is also of great importance and utility.

TIDES Portal Tool Usage

Specific TIDES Portal tools were used more often by some users for different reasons. One participant stated that he used whichever TIDES Portal search engine returned the fewest documents. Another participant preferred to use whichever tool returned the documents first (FBIS, in his case).

Table 5 summarizes user comments on each of the TIDES Portal search engines.

TIDES Portal tool	Positives	Negatives
Oasis	- easy to scan because of highlighted named entities - keywords useful	- slowest to retrieve documents - slowest to open documents - may contain more opinions / editorials than facts (note: this could be a positive attribute for some tasks)
BNN	- synopses - listed named entities - key frames	- slow to open documents - strange story segmentation found (key frame and title looked like weather report, but content was about North Korea- the weather report was after the North Korea story) - not used much
CCLINC	document listings were easy to scan (in English-only documents, titles were clear and little space was used for graphics)	- returned random documents unrelated to query - poor translation proved to be of no use to participants - unclear what to click on to view document - unclear what to select in order to drag document to Clipboard
FBIS	- retrieved many documents - quickest to return retrieved documents - quick to open documents - good translations	
Web Harvester	- straightforward, easily understood by users - quick to open documents	suggestion to incorporate other newspapers in addition to IHT

Table 5 Summarized user comments on the TIDES Portal search engines.

We also asked users for feedback on the TIDES Portal Clipboard which was not often used by participants during the experiment. When asked why, participants responded that there was no particular need to store links to documents. Below are some of the users comments.

> *"It would have been more useful if Word was not available to record things and I needed to compare two or more articles. For this task, I only needed to look at a single article at one time."*

> *"I didn't need to save data from one search to another."*

> *"It would be useful for deeper research, multiple queries."*

Most participants liked the clipboard, however, because of its capability to store documents links for repeated perusal.

> *"Excellent. Good. Really nice. Something I could use. I get 500 documents and I'm always going back and forth."*

Trust

Several users stated that they trusted the data coming from the TIDES Portal tools more than what they found on the Internet.

> *"I need to know the source of the information my tool is giving me. This information is important in understanding how much I should trust the data. I know that BNN information and MITRE information is trustworthy. I might not have that feeling about the information I find via Google."*

Usability

The two most common complaints made about usability of the TIDES Portal were related to the speed and the display of document proxies in the user interface.

Although participants had been warned, in advance, about the slowness of the systems, they still found the delays formidable. One user stated, "*I did not have control and had to wait for the rendering machine before I could do anything. There was so much time to wait for a search to return and then to wait for a document to open.*"

The display of document proxies presented other problems. Titles were not entirely visible; some wrapped within the pane, others were viewable only through horizontal scrolling which users found annoying. Users also had trouble resizing the panes to view the titles.

For the evaluators, the biggest problem was the instability of the servers. During the pilot and several testing phases, not all servers were available. Twice, experiment sessions were cancelled due to server downtime.

Figure 1, below, shows a screen capture of the TIDES Portal user interface after a search has been performed. Retrieved documents are shown for just three of the systems. The Clipboard is also open. The figure is labeled with participants' comments recorded during and after each experiment session. Even with just three panes visible (plus the Clipboard), information in the document titles is not entirely visible without horizontal scrolling or resizing.

Specific usability issues are described in more detail in the following sections.
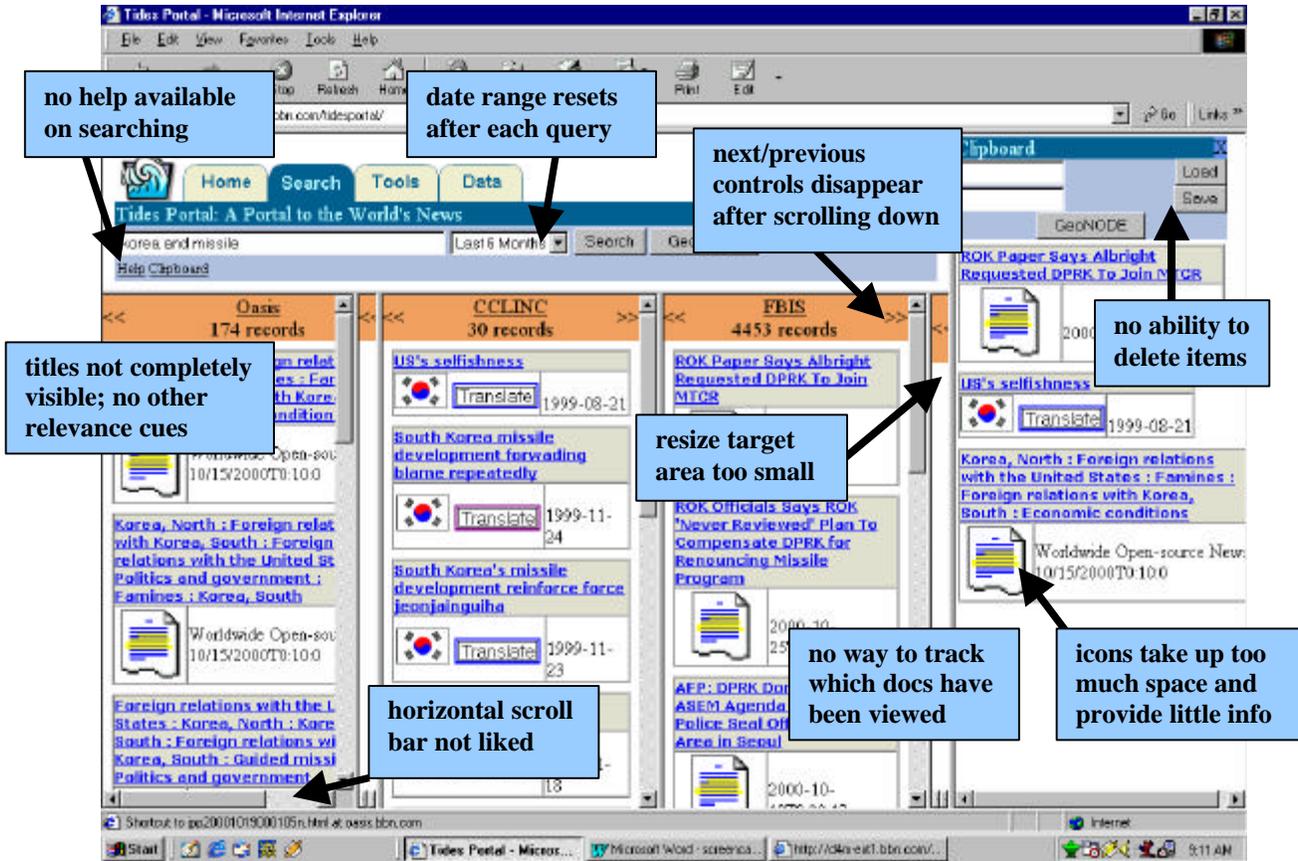
Figure 1 The TIDES Portal, with search results displayed from three systems. The Clipboard is open on the right.

Search

Because there was no help on how to perform a proper query, users were not sure which advanced features they could use. It was also not clear that all TIDES Portal tools handled advanced queries in the same way (for example, the necessity of the 'AND' conjunction).

If a user selected 'All Dates' in the date range dropdown box and clicked on the Search button, the data value would automatically reset to the default 'Last 6 months.' In a number of cases, users did not notice the change at first and had to perform an additional query. Partly because of this, some users requested a feature to cancel or stop a search.

Users complained that they lost control during the search and subsequent document retrieval. It would be better if they could start viewing returned documents while the other search engines were still busy.

When the participant used the <Back> or History buttons to return to the TIDES Portal after using a search engine, the TIDES Portal would automatically re-launch the last search. This caused both confusion and delays and is another argument for a cancel search button.

CCLINC always showed 30 as the number of results it returned, independent of the actual number of documents.

Document Proxy Display

As stated earlier, users had trouble scanning the lists of document proxies. While some titles wrapped and were easy to see, others required horizontal scrolling or pane resizing. None of the users liked the horizontal scrolling, and all but one had difficulty selecting the small target area to resize one of the panes. One person suggested a little button in the corner of each pane to maximize/minimize the pane. Another suggested listing the document titles horizontally rather than vertically.

Other users suggested making the icons smaller or removing them altogether since the screen real estate they occupied added little information as to the contents of the documents and caused more scrolling (since fewer titles fit in a screen).

Users seemed to forget they could view more documents, suggesting that the controls for requesting more documents should be at the bottom of the document list.

Visual Indicators

Several users had difficulty keeping track of which documents they had already examined and therefore opened a few documents more than once. They wanted the ability to "cross off" the documents they had reviewed or to have some other indicator that the link had already been visited. Color-coding the links as non-visited and visited would be helpful.

Textual Clues / Document Relevance

Users complained about having to open each document just to get an idea of the contents. They liked the Google search because it provided a relevant phrase or summary with the search terms highlighted. Some search engines also display the first two lines of text with each document proxy.

Viewing Document Contents

Several users suggested highlighting search terms in the retrieved documents.

"When I open a doc, I'd like to see the search terms highlighted. I'd also like to be able to jump from search term to search term. I had to use the Find for that."

"It was like a double search. First I'd do the search and then I'd have to search for the search terms in the resulting documents. I had no idea what was in the docs pertaining to my original search."

Most users did use the browser "Find" feature to look for relevant search terms. When there were multiple search terms in the query, this became a complicated task.

The Google search engine does provide highlighted search terms (see Figure 2, at left) via cached documents. However, none of the participants took advantage of this feature during this experiment.
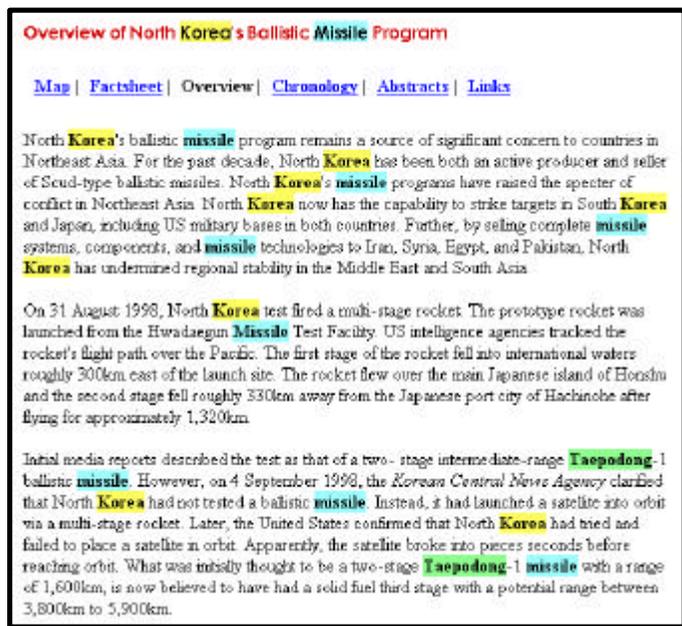

Figure 2 Google document with highlighted search terms.

One user requested the ability to read through all returned documents by scanning them in a long list.

Clipboard

Although the TIDES Portal Clipboard was not often used, its features were apparently well-liked. Users wanted the ability to clear the entire Clipboard (or create a new one) and to delete selected documents from it.

One user liked the idea of the Clipboard but did not like the way the document links were displayed. He felt the graphics were a waste of space, and the document titles were not sufficient information.

> *"It is sort of like bookmarks. But I like bookmarks better because I can bookmark sites related to docs. Then I can change the names to make more sense of things. It's more compact. I don't need all that info in the clipboard. – just one cue that makes the most sense to me."*

Summary and Future Evaluations

Although this experiment was intended to evaluate the <u>utility</u> of the TIDES Portal, the users were impeded by the system's <u>usability</u>. System performance, as well as the display of retrieved documents, were obstacles to completing the given task. Participants took more time to answer questions using the TIDES Portal and were less successful at answering them.

In general, users liked the variety of data available through the TIDES Portal system and trusted the data sources. Users also praised various features that aided in locating information quickly (e.g., highlighted named entities, listed topics and keywords, summarization features) once documents were open.

Participants most often criticized the lack of textual clues available in the display of retrieved documents. Titles were hard to read in the narrow panes, and there was no other information available on content relevance (e.g., first few lines of document, a brief summary, or relevant text with embedded search terms).

This experiment stresses the need for iterative evaluation of systems early on in the development stage. Usability issues, such as the user interface and system performance, are often not the focus during development of research tools, but they cannot be ignored. Users are less likely to value the utility of a system when usability is a negative factor.

A suggested next step is to incorporate useful experimental design features into the TIDES Portal. Once user interface design and network performance issues have been optimized (or stabilized), we would like to re-run this experiment on any future versions of the TIDES Portal to get a better view of the utility of the whole system as well as of the actual contributions of each component. It would be interesting to research and compile criteria and experimental designs that are currently used to evaluate Internet search engines. Others' approaches may be relevant to subsequent evaluation of the TIDES Portal or any other integrated TIDES prototype.

Acknowledgements

We would like to acknowledge Frank Linton whom we consulted on experimental design and formulation of questions with easy-to-score answers.

Appendices: Experiment Materials

Experimenter's Instructions
Consent Form
Pre-Experiment Questionnaire
Overview and Guidesheet
Experiment Questions, Parts A and B
Post-Experiment Questionnaire

Appendix A: Pre-Experiment Questionnaire

Thank you for participating in our experiment. Please answer the following questions to help us better understand some aspects of your background that may prove to be relevant to our evaluation.

1. What is your job title?

- Which computer platform(s) do you use most often?

- Have you ever gathered or analyzed open-source (non-secret) intelligence data?

    ___ Yes   ___ No

- Have you ever investigated the North Korean missile program?

    ___ Yes   ___ No

    If "Yes," what is your level of expertise?

    ___ I know the basics

    ___ I know a moderate amount

    ___ This is an area I know a lot about

- Please tell us how much you use Internet Explorer

    ___ Never; I normally use another browser (please state): _____

    ___ sometimes

    ___ always (it is my normal browser)

- Please state Internet Explorer version number, if known: _____

- Which search engines do you normally use? Please list.

Appendix B: Overview and Guidesheet

TIDES Portal: Window to the World's News

Usable information on demand from multiple sources and media, in multiple languages
- Television
- Radio
- Electronic news
- Digital libraries

Demo search korea AND disease

TIDES Portal Components

System	Description
Oasis	Transcribed radio broadcasts
Broadcast News Navigator	Indexed video and transcribed audio
CCLINC	Translated Korean documents
FBIS	Worldwide open-source news (multiple languages)
Web Harvester	On-line newspapers
Clipboard	Selected document storage (drag and drop)

Appendix C: Experiment Questions

Part A

1. On which date did North Korea pledge to uphold the moratorium on long range missile tests?

Tool or search engine used to answer question:	URL

2. Which countries are threatened by the North Korean Ballistic Missile Program?

Tool or search engine used to answer question:	URL

Part B

- How much does the Taepodong I missile cost?

Tool or search engine used to answer question:	URL

- To which country or countries has North Korea exported ballistic missiles?

Tool or search engine used to answer question:	URL

Appendix D: Post-Experiment Questionnaire

1. Were any questions particularly difficult to answer? Which? Why?

2. Would you use the TIDES Portal in your work? Suggest possible applications for the TIDES Portal.

3. Which tool did you use most in the TIDES Portal? Why?

4. Please comment on each of the tools:

TP Tool	Comments
Oasis	
BNN	
CCLINC	
FBIS	
Web Harvester	

5. Did you use the clipboard? Did you find it a useful feature? Suggest other uses for the clipboard.

6. What did you like most about the TIDES Portal?

7. What did you like least about the TIDES Portal?

8. What would you change about the TIDES Portal?

9. Was there something you wanted to do but could not do?

10. Which search engine(s) did you use?

11. Is this what you would normally use for a search? If not, why not?

12. Did you use any advanced search feature? Which?

13. Was there something you wanted to do with your search engine but could not do?

14. If you could design a web search engine of your choice, what would you do / change?