

A Search Relevance Algorithm for Weather Effects Products

Justin Nevitt

US Naval Research Lab
4555 Overlook Ave., SW
Washington, DC 20375

justin.nevitt@nrl.navy.mil

Don Brown

Atlassian (Formerly of BAE)

ABSTRACT

This paper is concerned with providing the user with an efficient way to find information, specifically weather effects products within a Service Oriented Architecture. The work outlined in this paper pertains to searching and ranking weather effects products from the EVIS (Environmental Visualization) data provider. EVIS was a data provider to a Federated Search engine in the NCES (Network Centric Enterprise Services) ECB (Early Capabilities Baseline). Several off the shelf search solutions are examined and a custom search/relevance algorithm is discussed. This algorithm is based on the idea that searching weather products is more akin to a database search. The paper concludes with a look at cross provider relevance and the complications that arise with a larger scale, growing SOA.

1.0 Introduction

We are situated in an information rich time [1] in which there is an overabundance of information, both useful and useless. A basic search on the internet reveals hundreds or thousands of documents on almost any subject, no matter how obscure. Previously, access to such specialized information was limited to individuals working in specific fields. Not too long ago, most information storage was in the form of physical texts (books, magazines, etc.). Had one wished to find a scientific article on a very specific subject he would have to physically go to an academic library. Today, the same information is available 24 hours seven days a week on the internet. The technology to store and present all this information was available years ago, but was not widely used. The problem was an overabundance of information in an unusable, unsearchable format. For instance, in the early days of the internet most navigation was done by hypertext referrals from one site to another or by word of mouth. Presently, searching has become second nature to even the most casual internet user. According to Alexa and other page ranking services the highest ranked web pages (estimation of most accessed) are often search engines [4] [5]. This suggests that people are navigating the internet by searching and not through the traditional linking and referral system of the original internet.

As time moves forward the amount of information available increases quickly. Wading through it to get to the information one wants can be difficult and possibly detrimental to task performance [2]. Search engine companies have made fortunes for their owners with the deceptively simple act of searching and presenting websites. This speaks volumes to the importance of a good search tool. Search, and more generally the organization of information, is what makes a vast body of information useful.

The Department of Defense is trying to keep several steps ahead of the enemy within the information technology arena. They often stress the importance of having information at the fingertips of the warfighter with buzzwords like “information superiority” and “information warfare.” [3] The idea is that in order to win the conflicts of tomorrow the United States must have the highest quality information,

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE OCT 2009		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE A Search Relevance Algorithm for Weather Effects Products				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Naval Research Lab 4555 Overlook Ave., SW Washington, DC 20375				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADB381582. RTO-MP-IST-087 Information Management - Exploitation (Gestion et exploitation des informations). Proceedings of RTO Information Systems Technology Panel (IST) Symposium held in Stockholm, Sweden on 19-20 October 2009., The original document contains color images.					
14. ABSTRACT This paper is concerned with providing the user with an efficient way to find information, specifically weather effects products within a Service Oriented Architecture. The work outlined in this paper pertains to searching and ranking weather effects products from the EVIS (Environmental Visualization) data provider. EVIS was a data provider to a Federated Search engine in the NCES (Network Centric Enterprise Services) ECB (Early Capabilities Baseline). Several off the shelf search solutions are examined and a custom search/relevance algorithm is discussed. This algorithm is based on the idea that searching weather products is more akin to a database search. The paper concludes with at a look at cross provider relevance and the complications that arise with a larger scale, growing SOA.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 16	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

A Search Relevance Algorithm for Weather Effects Products

presented in a timely and easily understood manner. Again, it is simply not enough to have all the information available. It must be available and easily consumed by the user. To quote Alberts, et. al.:

Improvements in the ability to share information will contribute to improvements in the ability to generate and maintain shared awareness which in turn, together with the greatly enhanced facilities to collaborate (quality of interaction), will contribute to improved synchronization. Thus, advances in the information domain that result from an improved ability to push the envelope in the richness, reach, and interaction space will affect processes in the cognitive domain which in turn will be reflected in the physical domain in the form of responsiveness, adaptability, agility, and flexibility. These competencies will provide a source of competitive advantage in the Information Age [3]).

This paper is concerned with providing the user with a way to easily find information, specifically weather products within the NCES ECB (Net-Centric Enterprise Services, Early Capabilities Baseline) SOA (Service Oriented Architecture). The work outlined in this paper pertains to searching and ranking weather products from the EVIS (Environmental Visualization) system. EVIS is a weather effects product generation and retrieval tool. It allows users to make custom maps with overlays of weather effects on various military operations. The generation of these maps includes user selection of locations, rules to analyze weather effects, security properties such as dissemination controls, and the ability to add routes. For example, a person planning a coordinated attack with aircraft and personnel on the ground can make a product that shows potential weather impacts to both these types of units. This product could then be saved to the server and shown to anyone who has the appropriate credentials via a secure network. EVIS was part of a larger entity called the NCES ECB SOA (NetCentric Enterprise Services Early Capabilities Baseline). In the NCES enterprise there were a number of data providers other than EVIS that created and published many different types of information. This included video surveillance, intelligence reports, biographies, etc. With all these providers it is important to have well designed search functionalities so a user can find the products that pertain to his mission quickly and without much effort.

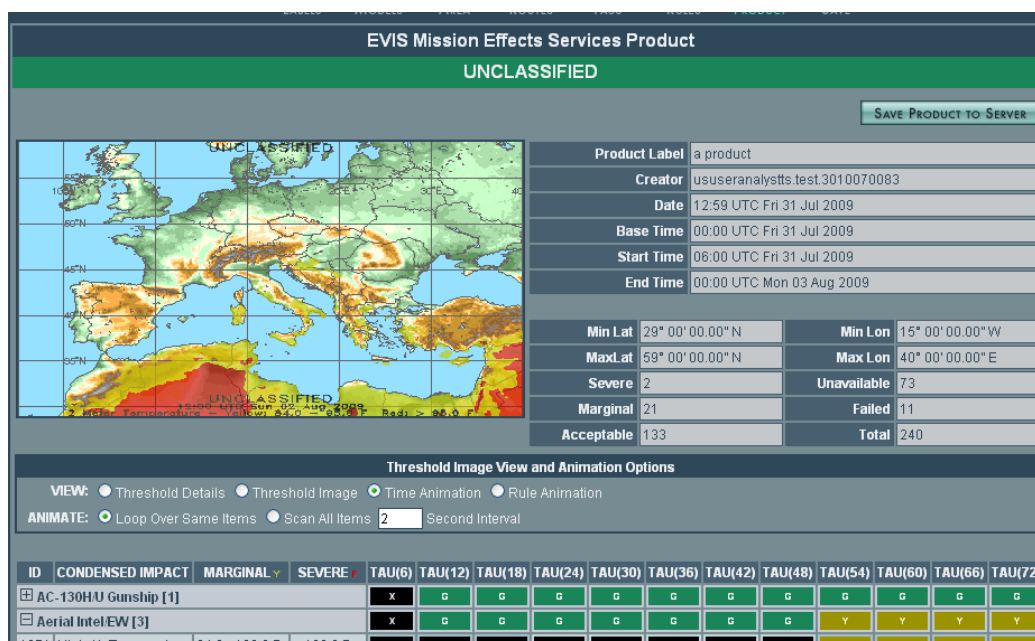


Figure 1 Sample EVIS Product

2.0 Motivations

EVIS was a content provider within the NCES ECB enterprise. In the NCES ECB architecture EVIS served two essential functions. One is to provide a workflow (called EMES, Environmental Mission Effects Services) that allows users to create weather effects products, as seen in Figure 1. These products are useful in the planning of military operations. The second duty is to present these weather effects products to end users. These end users are expected to be personnel possessing various levels of knowledge and at various levels on the chain of command. Some, if not all, of these people will be very busy. Some will even have people researching this information for them. It is very important that they are able to retrieve the appropriate products in a timely manner. In addition EVIS allows for the discovery of other theater weather products through a federated search capability.

Like any large enterprise, NCES ECB portal had its own specialized search functionality. This search capability has been dubbed Intelligent Federated Index Search (IFIS and sometimes referred to as FedSearch) and allowed the user to search all the content providers in the enterprise from one interface. It had some standard and some non-standard search features such as the ability to search by date, provider or geographic location. Unlike traditional search engines a Federated Search Engine does not scour all the data available in an enterprise and return matches. Instead, it uses a specialized content discovery process. First it refines the query based on stored knowledge of what the data providers can make available. This “knowledge” is ingested, in the form of xml metadata.

After this, the FedSearch engine routes the queries to only those providers that reasonably have data relevant to the query. Then the engine sends SOAP queries in a normalized form, based on DDMS (DoD Discovery Metadata Specification), to the data providers. The providers send a message back with reference to the matching documents. There are also provisions for sending messages to the providers to stop an ongoing search (providers may need to have long running searches) and request more results.

See Figures 2 and 3 below for a sample of the query and response schema. Query contains information about the search user in [profile], [constraints] on the query such as maximum results desired, the original form of the query (if available) in [originalForm], and a [normalizedForm] [10]. The normalized form has several nodes that can be filled with structured information based on the semantics of the query and the data provider.

A Search Relevance Algorithm for Weather Effects Products

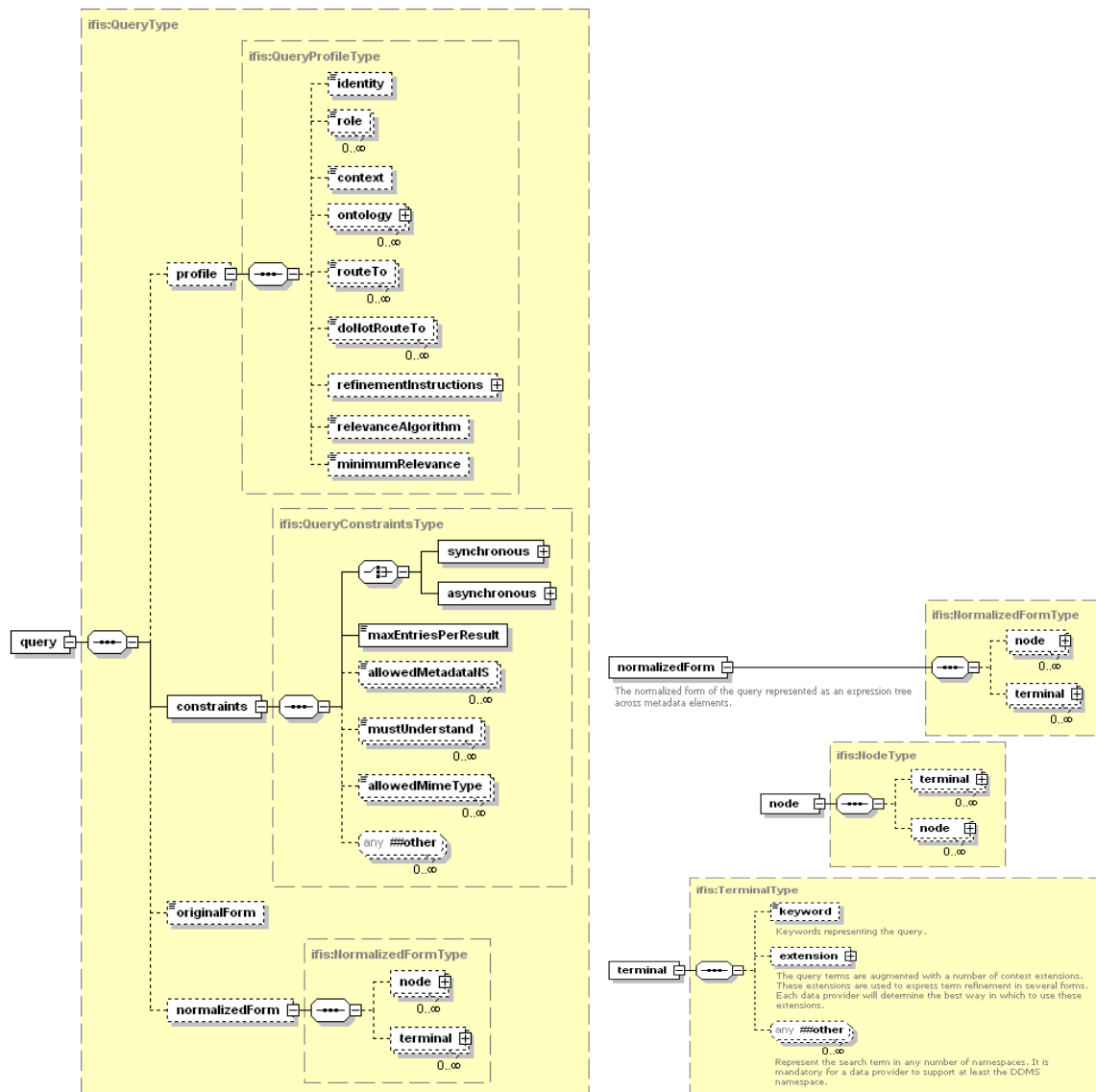


Figure 2 Query Specification

A Search Relevance Algorithm for Weather Effects Products

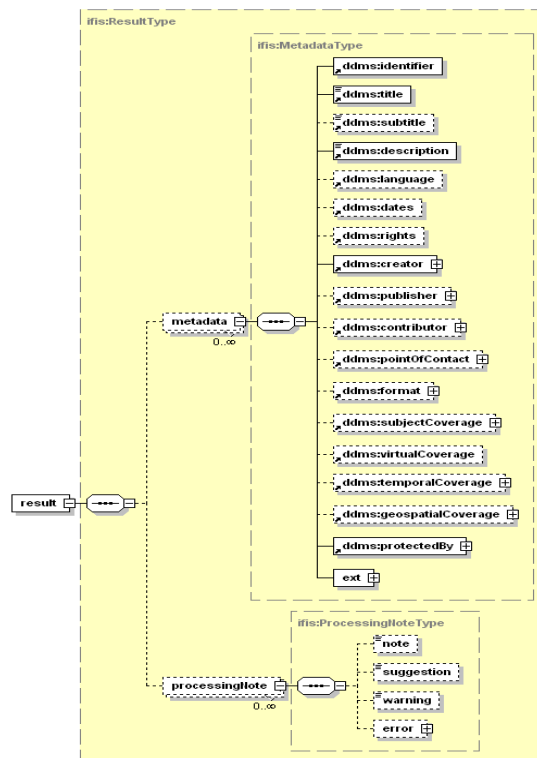
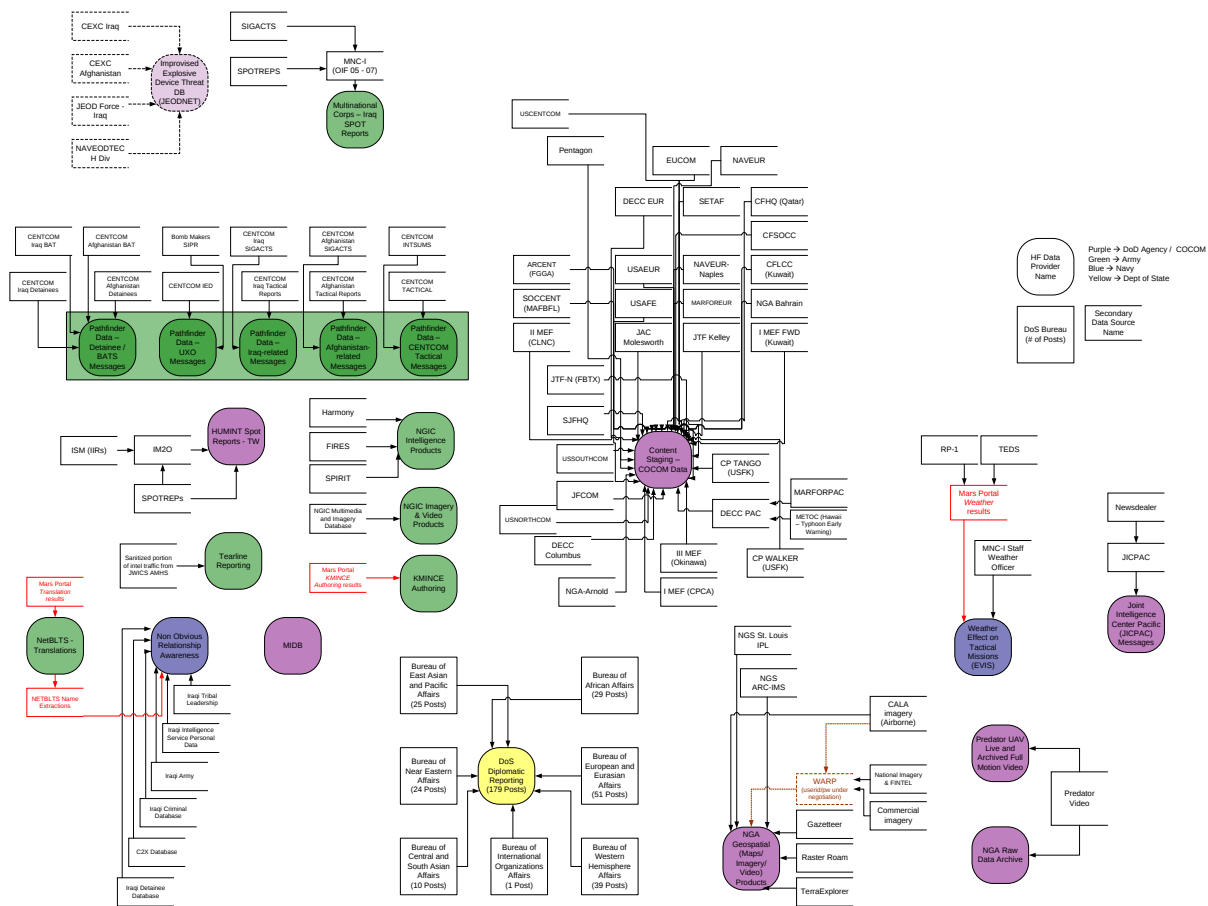


Figure 3: Return Schema

At first glance it may seem like a backwards and inefficient way of doing things. However, this is far from the truth. Searching the public internet with a search engine makes sense because most of the documents one is searching for are text-based, html web pages. Each web site does not have to provide search results back to the search engine for each query a user makes. This approach would be highly inefficient. In the case of NCES the content providers are serving very specialized types of information. Some examples include biographies, time sensitive reports, trusted intelligence, and video. These providers are expertly familiar with the content they serve. If FedSearch scoured all the content providers itself the operation would be highly inefficient. In addition, each time a content provider is added FedSearch would need to be modified. Instead, FedSearch remains unchanged and a new provider just has to pass a specification check and provide an endpoint for its search service. Content providers know when new content is added to their databases. They can add this content to their indexes or searches with minimal effort. FedSearch would have to periodically catalog all available data to achieve the same state.



Simple text search (TF-IDF)- A simple text search takes TF (term frequency) and compares that in some way with IDF (inverse document frequency). In other words it is a measure of how frequently a word appears in a document vs. how popular it is in multiple documents. This type of algorithm works, but not well with a system such as EVIS. EVIS products can be for one day or many days which would lead to problems with frequency of words vs. rarity of the word. Also, EVIS products can be searched by criteria that are not plain text such as time and lat/long location. [8]

A Search Relevance Algorithm for Weather Effects Products

Learning algorithms-An algorithm that learns or adjusts its weights has the possibility of being very good at predicting what the user needs. Development of this type of algorithm was not possible given the constraints on the use of cookies and other, similar technologies.

Google Pagerank-At the current time this is considered the gold standard in search engines. However, its ranking system is based, largely, on a measure of interconnectedness. A page that is referenced more often is given a higher ranking [9]. This type of measure is useless for EVIS as it does not link between products. Searching EVIS is more akin to querying a database.

3.0 Target Users

In designing the algorithm it was important to consider the users requesting EVIS products. In the commercial world software packages often fail because the user base is not properly assessed. The target user for EVIS is at its widest all military personnel who plan missions, forecast weather, or are involved in executing missions. At its widest scope “mission” can mean anything from a convoy of supplies to a covert SEAL team operation. In addition, users may have varying levels of technical knowledge and familiarity. This can cause problems at both extremes. Advanced users may expect querying to work in a fashion similar to commercial products. Also, the novice user may become frustrated if the item they are looking for does not appear within the first few hits. This information is somewhat useful, in that it provides motivation for creating a good search and relevance algorithm. More useful is information pertaining to how each of these people will use the software in their daily routines.

The categories of searchable information were selected based on possible queries users would submit to find weather products. The first step in designing the algorithm was to identify important categories of information in the weather products that people would use to query them. This was determined through interviews and interactions with potential users as well as what was available in the IFIS engine. The following categories were determined through interviews and interactions with potential users: time, title, person, location, keyword, effect, and mission.

It is unclear at this stage what type of query will be the most common, however several use scenarios are envisioned. Here are some examples:

Repeat user: A user responsible for getting weather effects reports and adding them to briefings on a particular region needs to access the same or similar products every day or so. This user would benefit from being able to search by location and time.

Mission planner: A mission is being planned in the very near future and weather effects would be useful in making choices. The mission planner may want to search by location, effect, time, and keyword (for type of mission). A mission planner may have people working under him who find these products and use them in briefings.

Mission executor: A mission has been planned and a user knows there is a specific weather effects product available for this mission. This user could find it by searching for a keyword which may appear in the title or summary.

On the surface searching and ranking EVIS products seems like a simple text query. All EVIS products contain text and for the most part users are inputting text (They can also specify date, location, etc.). Such a search service would be somewhat functional and would provide users with good results some of the time. However, the goal is to provide the user with useful and relevant results consistently. To do this a different search paradigm was necessary.

A Search Relevance Algorithm for Weather Effects Products

4.0 Search Approach

EVIS weather effects products are much like records in a database. Each one containing multiple entries of different categories (date, location, etc.) and differing amounts of entries in some categories (effects). To further complicate matters EVIS products are of varying sizes, which leads to complications if a simple text search is done.

The search and relevance algorithm created for EVIS is more akin to a database search tool than anything else. Like a database search tool it searches in bins (or categories) for matches to the user's query. Certain bins carry more weight in a search than others. If for instance someone is searching by date and location, which is more important, and by how much?

5.0 The Algorithm

The relevance algorithm is technically intertwined with the search algorithm. This choice was made mainly due to efficiency. Searching for products and assigning relevance scores on the fly is faster than sorting through a list of matching results. The search algorithm defaults to a simple AND (meaning that if there are multiple terms and no operators in a query they must all be present for a document to be counted as a hit) text search of all the fields in all the weather products that are currently available. An AND search was selected because it eliminates the possibility of getting back clutter from a specific query. EVIS and IFIS understand much more complicated Boolean keywords and symbols in many combinations. For instance "(not severe or moderate) and helicopter" is a legal query. As the scenarios mentioned above, most queries to this system are expected to be rather specific. However, due to the fact that the way in which the system is intended to be used and the actual real world use of the system could be different, this is highly configurable.

Relevance is determined based on several categories of information. These categories correspond, almost one to one, with the categories query terms are expected to come from. These categories also correspond nicely with the data as it is organized in each weather product. This makes for a relatively pain free implementation. The categories included in the algorithm are term, term_count, keyword, state_count, time, location, title, summary, creator, and not. Each of these makes up a portion of the total relevance score that a particular product receives. The portioning of this score is configurable so each category can be given as much weight, as needed, in determining the final relevance score. This allows for tweaking as EVIS is put to use. Currently the configuration is set so that the category term has the greatest sway on the relevance score.

5.1 Search categories

List of categories and how they are calculated:

- term – matches of query word in fields of a product. Fields include mission, evolution, operation, and parameter.
- term_count – number of times the term matches
- keyword – gets points if the query term matched a keyword. Keywords are specific to a data provider, for instance "weather."
- state_count – the number of matches for "severe", "marginal", or "acceptable" if specified.
- time – if the query time constraint matches the product. If the constraint is "before" or "after", the value is a function of how close the product is to the constraint. If "contains", then the max amount of points are given.
- location-if the product contains the query point or intersects with the query box. This value is Boolean. At this time Federated Search does not provide a location lookup for plain text location

A Search Relevance Algorithm for Weather Effects Products

names (gazetteer). In the future it may, but it is expected that it will return the same data type for location to content providers.

- title – do terms from the query match terms from the product title.
- summary - do terms from the query match terms from the product summary.
- creator – if a query by creator is done, is there a match.
- not – if the user specifies NOT in his query add score for term not being in the product.

For some of the categories a simple text search is done based on the number of chances to match vs. the number of matches. There is also the possibility of partial matches with this algorithm. These categories are summary, title, and term matches. Other categories are scored as stated above. Generally these are either Boolean or number values. Scores are temporarily stored for each of the results. After this is done the scores for each category are normalized against the highest score for that category.

The following formula expresses this (where S is a product's relevance score, c is the non-normalized score for the n th category of this product, m is the maximum score from any product in the n th category, and w is the configured weight for the n th category). Put more plainly, the score is scaled, per category, to the maximum score attained in all relevant products, in that category.

$$S = \sum_{i=1}^n \frac{C_n}{m_n} w_n$$

5.2 Current algorithm weights:

The current implementation of EVIS has used the following weights for each category:

Term: 15
Term Count: 10
Keyword: 5
State Count: 10
Time: 15
Location: 15
Title: 10
Summary: 10
Creator: 5
Normalize: false

These weights are based on the anticipated needs of the end users. As those needs change, so can the weights. Currently the weights emphasize time, term and location.

In order to explain how the algorithm will work in practice, and example EVIS product is needed. As described earlier the EVIS products are akin to database entries. In practice, the metadata for products is actually searched. An example is shown in Figure 5.

A Search Relevance Algorithm for Weather Effects Products

<SpreadsheetInfo xmlns="urn:us:gov:dod:don:usn:st:nrl:Evis:1:1">	
<Classification>U</Classification>	
Wednesday August 17, 2005 15:58 Z 15:58 D	
August 17, 2005	
August 18, 2005	
August 19, 2005	
August 20, 2005	
<ProductLabel>Location X August 17, 2005 </ProductLabel>	
<Summary>Weather Effects Forecast for Location X</Summary>	
<ProductURL>http://000.000.000.000/XXX1/index.php</ProductURL>	
<ServerName>https://servername.navy.mil</ServerName>	
<UserName> XXX1</UserName>	
<RequestID>Location X</RequestID>	
<MaxLat>11.0</MaxLat>	
<MinLat>22.0</MinLat>	
<MaxLon>33.0</MaxLon>	
<MinLon>44.0</MinLon>	
<CreationDateMillis>1124251199000</CreationDateMillis>	
<BaseTimeMillis>1124251199000</BaseTimeMillis>	
<StartTimeMillis>1124251199000</StartTimeMillis>	
<EndTimeMillis>1124596799000</EndTimeMillis>	
<RuleInfo>	
<Mission>AVIATION: Aerial Recon</Mission>	
<ParameterName>DUST</ParameterName>	
<TauInfo>	
<Offset>0</Offset>	
<State>SEVERE</State>	
</TauInfo>	
</RuleInfo>	
<RuleInfo>	
<Mission>AVIATION: Aerial Recon</Mission>	
<ParameterName>TURBULENCE</ParameterName>	
<TauInfo>	
<Offset>0</Offset>	
<State>SEVERE</State>	
</TauInfo>	
</RuleInfo>	

Figure 5: Sample Product Showing Searchable Data

A sample search of “weather” on the product in Figure 5: Sample Product Showing Searchable Data would produce hits in the algorithm for keyword, because EVIS is uniquely identified by the keyword “weather.” It would also get hits on the title, because “weather” is in the title. A search of “severe weather” would result in the hits for “weather” and a term count vs. possible term count for “SEVERE” in <State> would be calculated.

A further example of how the algorithm works is shown in Table 1. A sample search for the query terms listed in the first column would produce the relevance scores listed in the rows under each product, assuming that each product is of the following type:

- Products 1 & 2 are U.S. Marine Corps weather web pages
- Product 3 is for personnel and helicopter operations covering all of Iraq.
- Product 4 is for personnel and helicopter operations covering the east coast of the United States.

A Search Relevance Algorithm for Weather Effects Products

Table 1: Example relevance scores

Query Terms	Product 1	Product 2	Product 3	Product 4
Weather	10	10	5	5
Weather Iraq	0	0	40	0
human	0	0	25	25
Human	0	0	25	25
Iraq	0	0	10	0

Taken alone these results are exactly what a search should return. When querying for “weather” everything from EVIS is returned. The products returned have low scores for “weather” because there is a low likelihood that any single product is the one the user is looking for. However, they are all given a similar chance at being viewed. In the second query from the table above there were three terms “weather + Iraq + human”. One product was returned and this one product had a relatively high score. This is an acceptable result. The user’s query indicated that he wanted Iraq weather products that showed effects of weather on humans (vice airplanes or sensors) . This more specific query returned a higher result because it was easier to compute confidently that the user would like to view the product returned.

6.0 Federated Search Applet

The following are illustrative screen shots of the search interface applets a user encountered in the NCES ECB environment. The actual interface was similar to those shown below, but not exactly the same. Many pieces of the interface customize themselves based on a user’s role attributes and stored preferences.

6.1 Simple Search Interface:

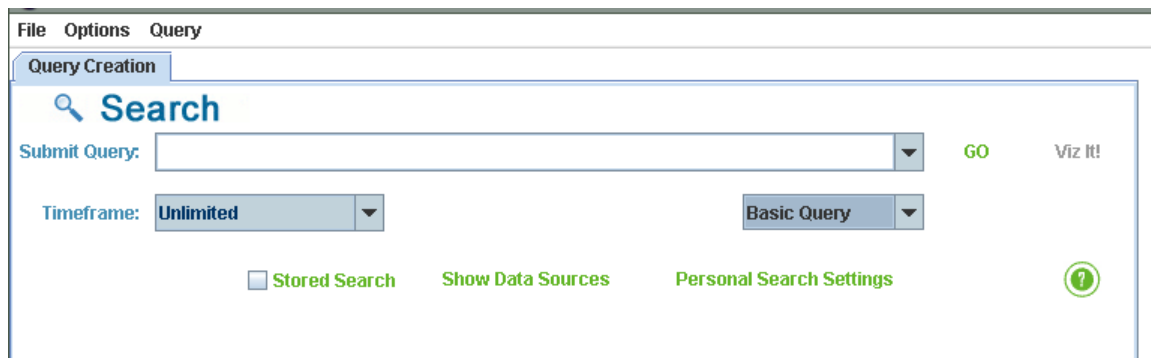
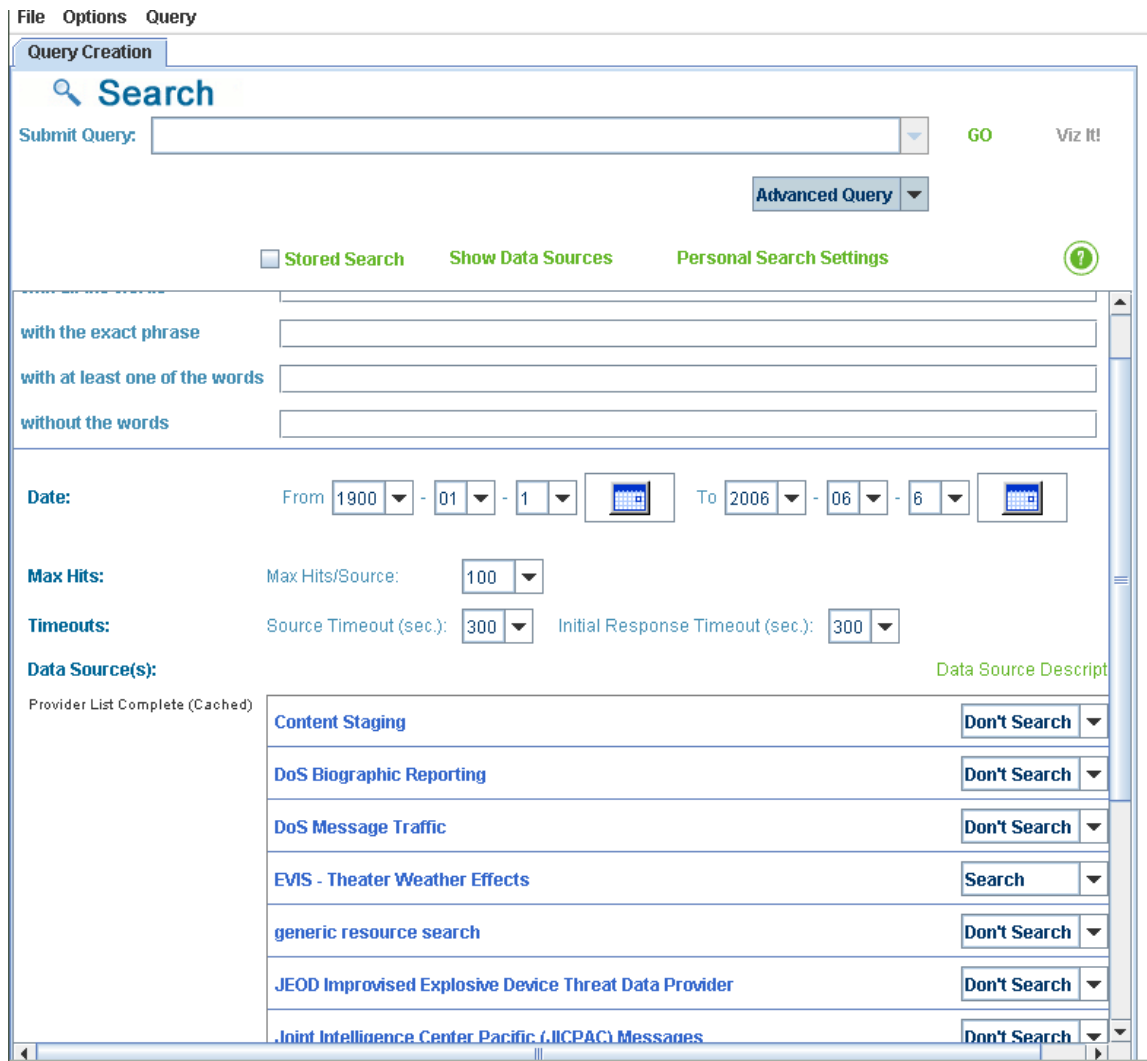


Figure 6: Basic federated search interface

This is the most basic interface through which users interacted with Federated Search. This interface was the default interface exposed when a user selected the search functionality in the portal. It has basic functionality in it such as the ability to use personal settings, the ability to use stored search criteria and a timeframe limitation setting.

A Search Relevance Algorithm for Weather Effects Products

6.2 Advanced search interface



File Options Query

Query Creation

Search

Submit Query: GO Viz It!

Advanced Query

☐ Stored Search Show Data Sources Personal Search Settings ?

with the exact phrase

with at least one of the words

without the words

Date: From - - To - -

Max Hits: Max Hits/Source:

Timeouts: Source Timeout (sec.): Initial Response Timeout (sec.):

Data Source(s): Data Source Descript

Provider List Complete (Cached)

Content Staging	Don't Search
DoS Biographic Reporting	Don't Search
DoS Message Traffic	Don't Search
EVIS - Theater Weather Effects	Search
generic resource search	Don't Search
JEOD Improvised Explosive Device Threat Data Provider	Don't Search
Joint Intelligence Center Pacific (JICPAC) Messages	Don't Search

Figure 7: Advanced search interface

Should a user wish to perform a more specific query they could do so by selecting “Advanced Query” from the drop down box on the right side of the basic search screen. This displayed the screen shown in Figure 3. In addition to the settings available in the basic search interface this interface offers the following: specific date constraint selection, configuring the output parameters of a search (maximum hits gathered, timeouts, etc.), and it allows for the selection/exclusion of specific data providers.

6.3 Search Results:

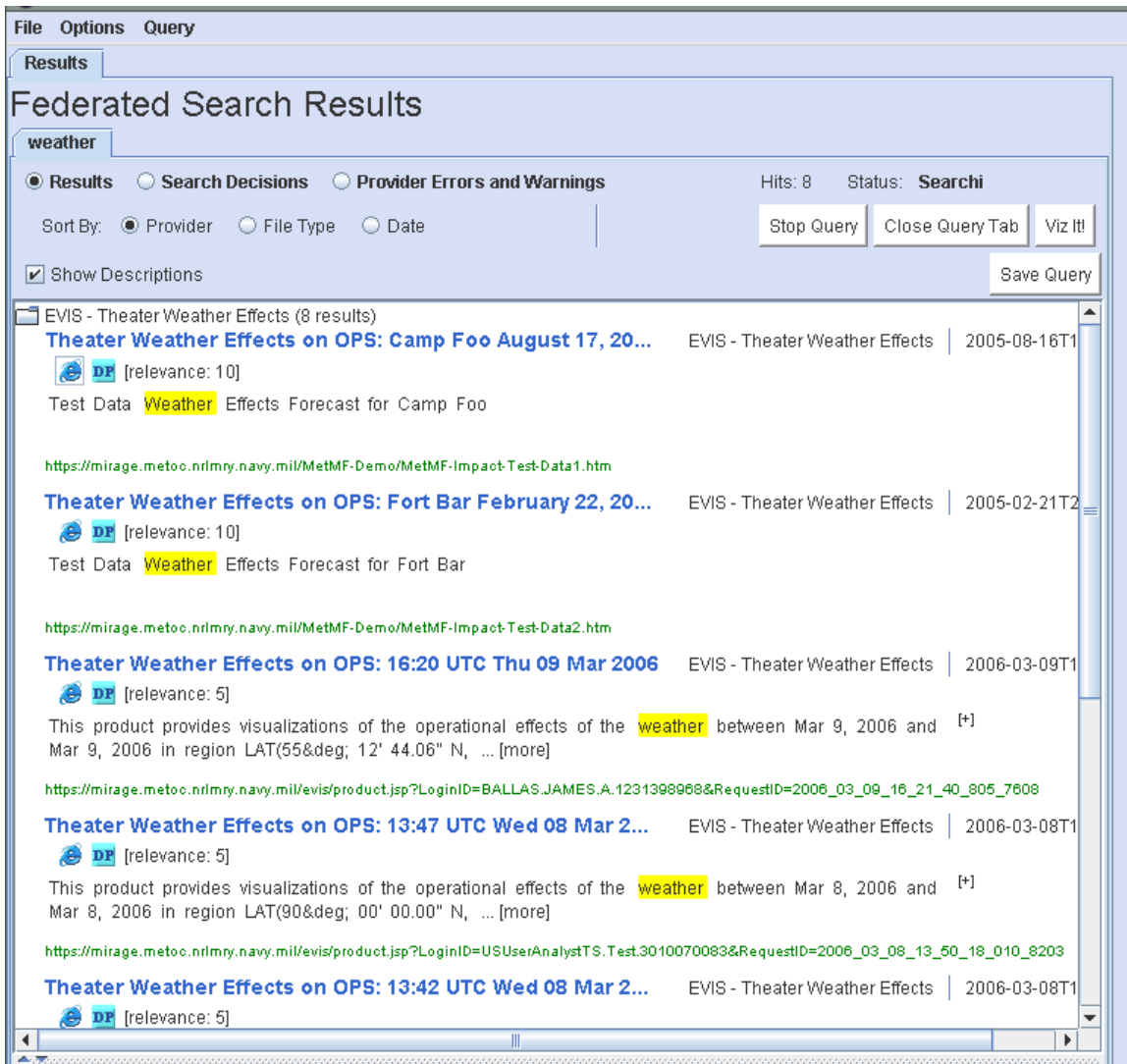


Figure 8: Search Results display

Upon the successful completion of a query the user was presented with a screen similar to that shown in Figure 8. In this interface users were able to sort results by provider, file type, and date as well choose to show descriptions or not. Each result item had the following information associated with it: title, provider name, date. Clicking on a product title will bring the user to the selected product.

A Search Relevance Algorithm for Weather Effects Products

6.4 Products Searchable via the GISA Tool

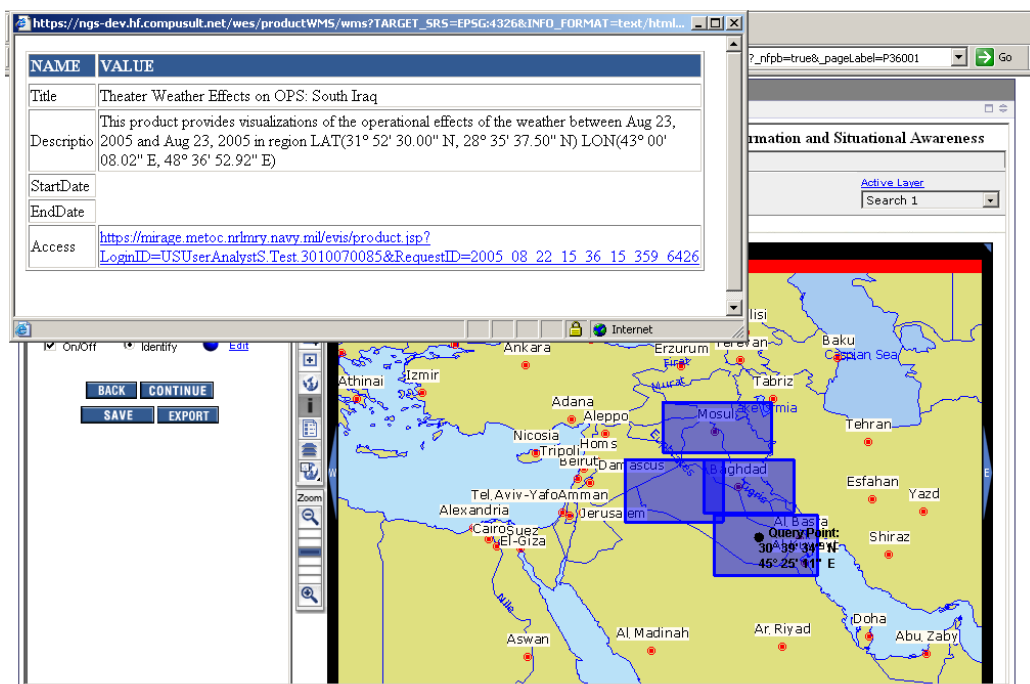


Figure 9: Geographical search interface

The GISA (Geospatial Information Situational Awareness) portlet shown in Figure 5 was a tool that allowed users to search for products via geographic location. It uses the IFIS engine to return a list of product which can be narrowed down using the interactive map or a list similar to the one in the Federated Search client.

7.0 Cross-Provider Relevance

Federated Search simultaneously returns results from multiple data providers, each providing different types of data using their own relevance algorithms. This becomes a problem if one provider writes its algorithm so that it always scores low (or always scores high) relative to the other providers. When its products are meshed with the products of other providers it may always appear low on the list and never see any traffic from users. This is the biggest problem in returning cross-provider relevance. A quick solution to this problem is normalizing the scores returned by each provider. This, however, is also problematic. For instance, if a user queries FedSearch for products pertaining to Iraq weather they should receive all EVIS products with Iraq in the title or summary. They may also receive products that mention weather and Iraq in passing, perhaps only a few from a single provider. The first product from this provider would be rated at or close to 100% relevant even though it might not be that relevant. And the last product returned by this provider might be rated at a floor of 1% relevant, where it should probably be somewhere at 30% or 40%.

Figure 6 is a typical output from the search query “weather”:

A Search Relevance Algorithm for Weather Effects Products

Format	Relevance	CLS	Title/Abstract	SourceDate
	 [100]	[S]	DoS Telegraph (06ISLAMABAD429): PAKISTAN - EARTHQUAKE: USAID/DART ASSESSMENT TRIP TO MUZAFFARABAD [Highlighted] ...effects of the severe winter weather during the previous days. Affected...remarkably well through the severe weather . That affected populations managed as...effect of the severe winter weather in previous days. Per Ref...the first wave of winter weather , including heavy rains and snowfall...due to the severe winter weather that occurred on January 1 and... http://ncddosi.citi-us.com/ncddos/cable/mrn_06ISLAMABAD429.html	[DoS Message Traffic] 2006-01-13T07:27:00+0000
	 [100]	[S]	DoS Telegraph (06ISLAMABAD437): PAKISTAN - EARTHQUAKE: USAID/DART ASSESSMENT TRIP TO MUZAFFARABAD [Highlighted] ...effects of the severe winter weather during the previous days. Affected...remarkably well through the severe weather . That affected populations managed as...effect of the severe winter weather in previous days. Per Ref...the first wave of winter weather , including heavy rains and snowfall...due to the severe winter weather that occurred on January 1 and... http://ncddosi.citi-us.com/ncddos/cable/mrn_06ISLAMABAD437.html	[DoS Message Traffic] 2006-01-13T09:08:00+0000
	 [099]	[U]	(U) nkor.pdf 77 Air Defense Support to Offensive Operations https://hfdportal2.spawar.navy.mil/rfPortal/portlets/kminceauthori ng/fedsearch.jsp?redirectURL=https%3A%2F%2Fhyperh1.us.hyperwave.com/m%3A443%2FNGIC%2520Imagery%2520and%2520Video%2520Products%2FStuff%2FNGIC_Products%2FNuclear%2520Weapons%2Fnuke%2Fguide%2Fdpk%2Fnpk.pdf	[NGIC Imagery and Video Products] 2004-01-23T01:09:46Z
	 [097]	[S]	South Korea: Weather Supercomputer Urgently Needed for the supercomputer being used for **weather** forecasting that is located in the Computerization Technology Center [Research Institute]. Local **weather** forecasting is expected to suffer. As a result, a strong demand is being made to purchase a dedicated supercomputer for quickly analyzing **weather** data https://pathfinder.us.hyperwave.com:8443/docServer/Retriever?dbname= /export/horizontal_fusion/databases/FBIS_1998&recnum=241285&qfile=H F_1143236201444_23254342.tmp	[Pathfinder Data - Iraq Related Messages] 1998-09-16
	 [089]	[S]	Nigeria: Nigeria--Meteorological Equipment Installed at Six Airports Article by Leo Collins: "Six Airports Get Automatic **Weather** Observing Equipment" < country. the across stations its of six in (AWOS) System Observing **Weather** Automatic > Also experimental farm at the Central **Weather** , Forecasting Station, Oshodi, also in Lagos. According https://pathfinder.us.hyperwave.com:8443/docServer/Retriever?dbname= /export/horizontal_fusion/databases/FBIS_1998&recnum=1504248&qfile=H F_1143236201444_23254342.tmp	[Pathfinder Data - Iraq Related Messages] 1998-05-28
	 [089]	[S]	South Korea: ROK, DPRK Exchange Notes on Possible Information Trade on a possible swap of **weather** -related information, the meteorological administration said thursday. The administration sent the telegram, in the name of kimp'o airport's **weather** office chief , stressed the urgency of exchanging **weather** information for various purposes, including safety of flights https://pathfinder.us.hyperwave.com:8443/docServer/Retriever?dbname= /export/horizontal_fusion/databases/FBIS_1998&recnum=62958&qfile=H F_1143236201444_23254342.tmp	[Pathfinder Data - Iraq Related Messages] 1998-04-25

Figure 10: An example search output

As you can see a lot of the products are not weather specific. EVIS products do not even show up in the first page of hits.

Currently work is being done to remedy this situation. Cross provider relevance is not an unattainable goal. A solution is in the work that looks at secondary characteristics and context to determine the proper cross provider relevance scores. These characteristics include the tendencies of each provider's algorithm, the content/syntax of the query, and other related attributes [6]. The hope is that, in the future it will be easy to pull data from various sources and present them to the user in an easy to digest form.

8.0 Beyond

There are several important points to take away from this development process. The first is that the user is paramount in making design decisions. In many situations it can be said that the user is what makes or breaks an endeavor. The difference between a good product and a not so good product is the way the user perceives it. In this case the users are military personnel who need accurate and timely weather effects reports. This meant tailoring the search algorithm to their needs. Specifically this meant weighing the categories in favor of ones they are expected to search more often.

The second lesson learned is that off the shelf solutions are not always the simplest or best for a particular problem. This problem required something specific and simple to solve. A custom algorithm was ultimately not overly complicated to devise and implement. In addition, this algorithm specifically answers the needs of its target users in a way that others might not be able to.

The NCES ECB that this technology was a part of is no longer operational. It was, as the title suggests, a baseline of early capabilities, not a deployed and sustained system. The lessons learned and concerns raised while it was operational are valid in current and future command and control systems. For instance,



A Search Relevance Algorithm for Weather Effects Products

DISA NCES program continues to be the DoD's core enterprise services program, providing federated search among other services.

REFERENCE:

- [1] Alberts, D. (2002). Information age transformation: getting to a 21st century military. Washington: DoD Command and Control Research Program.
- [2] Alberts, D. (1996). The Unintended consequences of information age technologies. National Defense University.
- [3] Alberts, D., Garstka, J., Hayes, R & Signori, D. (2001). Understanding information age warfare. Washington: DoD Command and Control Research Program
- [4] Alexa Web Search – Top 500. (Feb. 2006) [On-line] Available: <http://www.alexa.com/topsites>
- [5] Bosch, S & Fan, J. (Jan 18, 2006) Top search terms reveal web users rely on search engines to navigate their way to common web sites, according to Nielsen//Netratings. [On-line]. Available: http://www.nielsen-netratings.com/pr/pr_060118.pdf
- [6] Diamond, T. (July 1999) TIPSTER III final report. TextWise, LLC Data Fusion Segment.
- [7] Eastman, C. & Jansen, B. (2003). Coverage, relevance, and ranking: the impact of query operators on web search engine results. ACM Transactions on Information Systems, 21(4), 383-411.
- [8] Joachims, T. (1997). A Probabilistic analysis of the Rocchio Algorithm with TFIDF for text categorization. Proceedings of ICML-97, 14th International Conference on Machine Learning. (NOTE: for TFIDF explanation)
- [9] Google Technology. (2006). [On-line] Available: <http://www.google.com/technology/>
- [10] Horizontal Fusion Developers Reference v 1. (2004). Horizontal Fusion. [On-line] Available: http://horizontalfusion.dtic.mil/docs/specs/20041118_Final_Developers_Ref.pdf.