2001-02-28

newsAgent
A tool for automatic news surveillance and corpora building

Martin Hassel

NADA-KTH
100 44 Stockholm, Sweden
ph: +46 8 790 66 34
fax: +46 8 10 24 77
email: xmartin@nada.kth.se

Introduction

The SeaSum project consists of two main tracks. Track 1 concerns the integration of *natural language technology* with the current *search engine technology* existing at EuroSeek AB. Track 2 concerns the development of a *message routing system*, or more precisely, a user centred news surveillance and delivery system. Besides the obvious commercial aspects of the system, the purpose was to build a test bed for new natural language technology tools, i.e. *automatic text summarization*, *named entity tagging*, *stemming*, etc. In the process of building this system we also made it capable of gathering the news texts into a corpus, a corpus we have used to train and evaluate above mentioned tools.

This report will mainly consist of two parts. The first part concentrates on the news service, *Nyhetsguiden*, and the application running the service, *newsAgent*. Besides a technical description of the service, this part contains results from a brief user survey. The second part of this document concerns the development of the corpus of Swedish news texts built with newsAgent.

1. Nyhetsguiden – A user centred news delivery system

The system has a modular design and consists of three parts, the user interface, the user database and the main application, newsAgent. Being modular, the system can be run as a distributed system or on a single web server. When run as a distributed system, at least newsAgent must be run on a computer with Internet access. The user interface (Nyhetsguiden) and the user database can reside on either an Internet or Intranet capable server depending on the desired public access to the system.

1.1 newsAgent – at the heart of the system

newsAgent is the core of the system and is basically a web spider that is run in a console window. The spider is implemented in Perl, which makes it platform independent, that is, it can run on any platform running Perl (Unix/Linux, Windows, Macintosh, BeOS, Amiga, etc). On intervals of 3-5 minutes newsAgent searches the designated news sources (Appendix D) for new news texts, that is news texts not seen by the system before. When a new news text is encountered it is fetched, the actual news text and accompanying illustrations are extracted (by removing navigation panels, banners, tables of links, etc). The resulting document is then passed through the system (Appendix A, picture 1).

1.2 Nyhetsguiden – what the user sees

Each document passed through the spider is matched against a user database containing traditional search expressions (queries) as given to, for example, a traditional search engine. These search expressions have been entered by users, using the third part of the system, the user interface (Appendix A, pictures 2-4). Upon a match, the news text can be summarized and pushed, according to the user's desire, either by e-mail, WAP, SMS, fax or in a personalized ticker. The filters for the designated news sources and the plug-ins for e-mail, WAP, FTP, etc are developed separately and new and extended functionality is easily added. At the moment only push through e-mail is implemented but WAP and ticker plug-ins are under development and SMS and fax is easily implemented through the use of dedicated gateways.

The user interface gives easy access to all matching operators, several other options as well as guides and, when the ticker is fully implemented, the users news reports. The matching operators implemented so far are:

- Fuzzy matching
- Part-of matching
- Case sensitive matching

Fuzzy matching accepts small differences between a query term and the matched word. This is a way of coping with, for example, names that have several spelling alternatives (i.e. Kadaffi/Ghadaffi or Jansson/Janson/Janzon). The default matching behaviour is, however, to use exact matching, since fuzzy matching on badly chosen search terms can lead to the strangest matches. Part-of matching copes with the large amount, and many times creative use of compound words. The default behaviour here is to match whole words only, but if the user so wishes the search term *skjut* (shoot) can match *skjuta* (to shoot) as well as *skjutbana* (shooting range). The default behaviour is that the search term *Martin* matches both *Martin*, *martin* and for that matter *MaRtIn*. Case sensitive matching can however be used to match case in order to avoid unwanted matching between, say, an ordinary word and a name denoting a person (as an English example we can use *Bush* and *bush*). There are also plans on including our stemmer (Carlberger et al 2001) as an alternative to fuzzy matching. If there arises a user demand for a specific type of matching operator, this can of course be added at a later stage.

1.3 User survey

In order to evaluate the need for an automated news broker we did a brief news survey among 25 of our 36 current users. The 25 where chosen because they where, at the time, frequent users of Nyhetsguiden and had had time to experience the service during a period of at least a month. Of these 25, 18 answered our survey, which was done by non-anonymous e-mail. The survey was done in Swedish (Appendix B). The questions and a compilation of the users answers translated into English are given below (Table 1).

1) Has this service been useful for you?

Yes/No/Don't know

Yes 18

2) Do you receive too many news reports?

Yes/No/Don't know

No 15
Don't know 3

3) Do you receive too many news reports with the same or similar content?

Yes/No/Don't know

Yes 10
No 6
Don't know 2

4) Are the news reports too long?

Yes/No/Don't know

Yes 2
No 14
Too much diverse text around the actual news 2

5) Which channels or fields do you follow?

Domestic/World/Sports/Business/Technology/Society/Other

All channels/fields 2
Some channels/fields 3
One channel/field 9
Don't know 1
No answer 1

Domestic 3
World 4
Sports 6
Business 4
Technology 8
Society 7
Other 4

6) Are there any other channels/fields that you would like to follow?

Science and research 2
Foreign sources 4
Religion 1
Art 1
Culture 1
No (existing channels are enough) 1
Don't know 1
No answer 4

7) What are your age and your profession?

Age:

18-24 = 1
25-34 = 5
35-44 = 6
45-54 = 0
55-64 = 2
65+ = 1
No answer 3

Profession:
Administrative 2
Research 4
Software engineering 2
Leadership 3
Information distribution 2
Health care 1
Technical engineering 1
No answer 2

8) Would you like to have a feature that only allows "new" news reports to reach you and not news report containing the same news you've just read?
Yes/No/Don't know
Yes 12
No 1
Don't know 5

9) Would you like to have a feature that shortens the news reports to 2-3 lines of text?
Yes/No/Don't know
Yes 8
No 7
Don't know 3

10) Would you like to have a feature that shortens the news reports to 2-3 rows of text that focus on the key word(s) you have given? For example: if you follow the word Ericsson, then only the text around this word would be sent to you, so you won't have to read about, for example, Volvo or some other company if they are mentioned in the same news.
Yes/No/Don't know
Yes 9
No 4
Don't know 5

11) Would you like to have a feature that summarizes several similar news reports to one news report (8-10 lines)?
Yes/No/Don't know
Yes 10
No 3
Don't know 5

12) What type of matchning of the search term(s) do you use?
Whole words only/Partial Matchning/Fuzzy Matchning/Don't know

Several types of matching 2
One type of matching 11
Don't know 5

Whole words only 11
Partial Matching 4
Fuzzy Matchning 1

13) Did you find it hard to formulate your query?
Yes/No/Don't know
Yes 4
No 12
Don't know 2

Table 1. User survey results

As we can see, all 18 who answered have had use of the service, so clearly there seems to be demand for this type of service. Another conclusion to draw form this brief user survey could be that this type of service is most needed, or maybe most wanted, by researchers, developers and marketing people. The age of the users is though quite evenly spread, even if there is a peak at mid-career age. We can also see that the current system is not apprehended as doing as good a job as it could. For example 10 of the 18 users believe that they get too many similar news reports. This indicates that there is a need to not only be alerted when something interesting has been released through the news wire, but there also is a clear need to be alerted only when something new has happened. The users seem to be somewhat divided on the topic of summarization, both concerning single and multiple news reports, even if it would decrease the redundancy. This is however a feature we would like to implement. We believe that the users, when presented with the final feature, will appreciate it, and we also believe that a later user survey may prove us right.

2. Construction of a Corpus of Swedish News Texts

Traditionally it has been hard work constructing a corpus of news text. In Sweden there are no newspapers that on a yearly basis offer their paper in digital form, as some foreign newspapers do, meaning that obtaining this material has to be done on demand. Many Swedish newspapers are, when inquired, unwilling to release texts from their archives for research purposes, and even when they do, it is often the question of a small amount of news texts with an age of several years. This may potentially lead to the exclusion of contemporary words and giving unusually high, or low, occurrence frequencies to words related to phenomena limited to a certain period of time.

In the past, the solution would be to collect newspapers in their paper form and type or scan (using a Optical Character Recognition program) them in order to convert them to a format manageable by computers.

The World Wide Web is, on the other hand, today a large collection of texts written in different languages and thus giving an abundant resource for language studies already in a format, by necessity, manageable by computers. Many of the web pages are also frequently updated and so give a steady access to concurrent use of language in different fields. In this situation, neglecting the usability of Internet as a corpus would be foolish. In our case we used a tool called newsAgent that is a set of Perl scripts designed for gathering news texts, news articles and press releases from the web and routing them by mail according to subscribers defined information needs.

2.1. KTH News Corpus

The project with the KTH News Corpus was initiated in May 2000. We started out collecting news telegrams, articles and press releases from three sources but with the ease of adding new sources we settled for twelve steady news sources (Appendix D). As of February 2001 we have gathered more than 100.000 texts amounting to over 200Mb with an increase of over 10.000 new texts each month. The increase in word forms during the last month was almost 230.000. The lengths of the texts vary between 5 and 500 lines with a tendency towards the shorter and an average length of 193 words per text. No data on the total number of words or word forms are available at the moment.

The texts are stored in HTML tagged format but only the news heading and the body of the news text is preserved. All other page layout and all navigation tables and banners are removed. Each text is tagged with Meta tags storing the information on time and date of publication, source and source URL. Using the news sources own categorization of their news texts, instead of a reader based categorization (Karlgren 2000), we have stored the news in different categories (Appendix D) and thus giving the possibility to study the difference in use of language in, for example, news on cultural respectively sports event. The corpus is structured into these categories by the use of catalogue structure, a HyperText linked index and a search engine driven index thus giving several modes of orientation in the corpus.

For the purpose of evaluating a Swedish stemmer in conjunction with a search engine (Carlberger et al 2001) manually tagged 100 texts TREC style and constructed questions and answers central to each text. We also tagged each text with Named Entities and keywords for future evaluation purposes.

The purpose of the project is to create a corpus aimed at research on Information Retrieval, Information Extraction, Named Entity Recognition and Multi Text Summarization and for this purpose the system does not, contrary to (Hofland 2000), remove duplicated concordance lines. Unfortunately copyright issues remain unsolved, we have no permission from the copyright holders except fair use, and so the corpus can only be used for research within our research group. The tool for gathering the corpus, newsAgent, is on the other hand available for use outside our research group (with the exclusion of mail routing and FTP plug-ins).

2.2. Areas of use

So far the corpus has been used for three evaluation purposes. (Knutsson 2001) has used it for evaluating error detection rules for Granska, a program for checking for grammatical errors in Swedish unrestricted text. The tagged texts have besides, as mentioned above, being used for evaluation of a Swedish stemmer also been used for evaluating SweSum (Dalianis & Hassel 2001), an automatic text summarizer that among other languages handles Swedish unrestricted HTML tagged or untagged ASCII text.

In the near future parts of the corpus will be used for the training and evaluation of a Named Entity Tagger and for expanding SweSum with Multi Text Summarization. Other possible areas of use are for producing statistics and lexicons, and for developing Topic Detection Tracking systems for Swedish news.

2.3. The Future of the Corpus

We are now on the verge of rewriting the tools for building the corpus since we now are more fully aware of its potential uses. Among planned improvements are:

- Internal representation in XML
- Automatic tagging of:
  - Parts-of-speech
  - Clause and sentence boundaries
  - Named Entities (persons, locations, etc.)
- Automatic summarization of each text

- Automatic running statistics
  - Average increase per month/week in number of:
    - Texts
    - Sentences
    - Words
    - Word forms
  - Average:
    - Text length (in sentences, words & characters)
    - Sentence length (in words & characters)
    - Word length
  - Total number of:
    - Texts
    - Sentences
    - Words
    - Word forms
- Hopefully a solution to the current copyright issues
- A more balanced choice of channels/sources

References

H. Dalianis & M. Hassel 2001. *Development of a Swedish Tagged Corpora for Evaluating Summarizers*, draft.

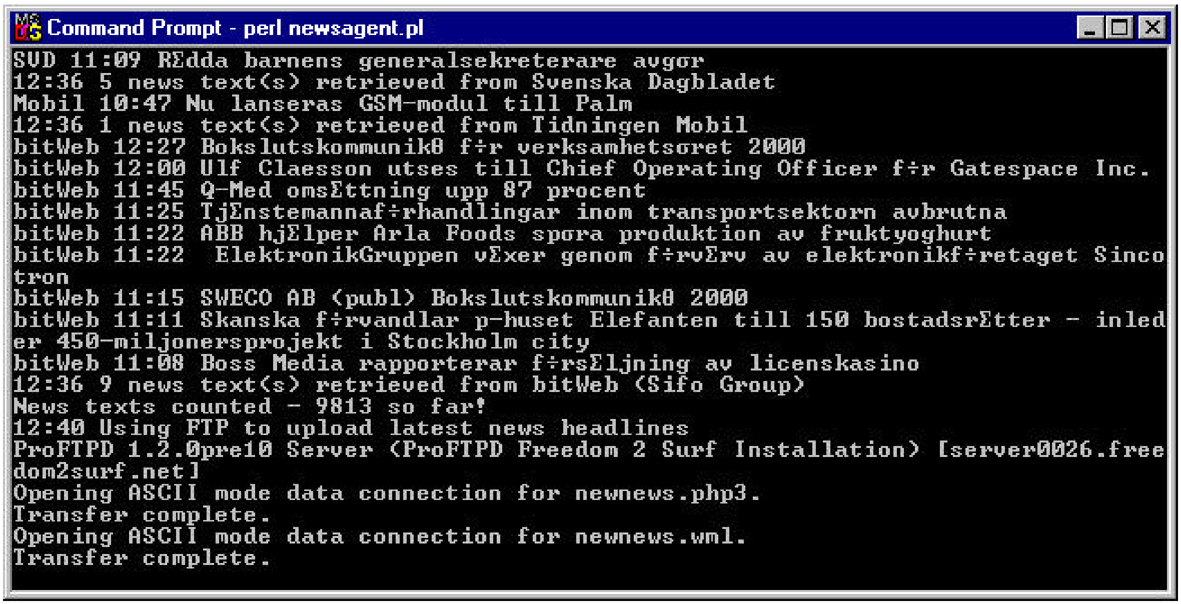
J. Carlberger, H. Dalianis, M. Hassel, O. Knutsson 2001. *Improving Precision in Information Retrieval for Swedish using Stemming*, draft.

K. Hofland 2000. *A self-expanding corpus based on newspapers on the Web*. In the proceedings of Second Internation Conference on Language Resources and Evaluation. LREC-2000 Athens, Greece, 31 May-2 June 2000. pp 1271-1272.

J. Karlgren 2000. *Assembling a Balanced Corpus from the Internet*. Stylistic Experiments for Information Retrieval, Dissertation for the Degree of Doctor of Philosophy, Stockholm University, Department of Linguistics. pp 99-104.

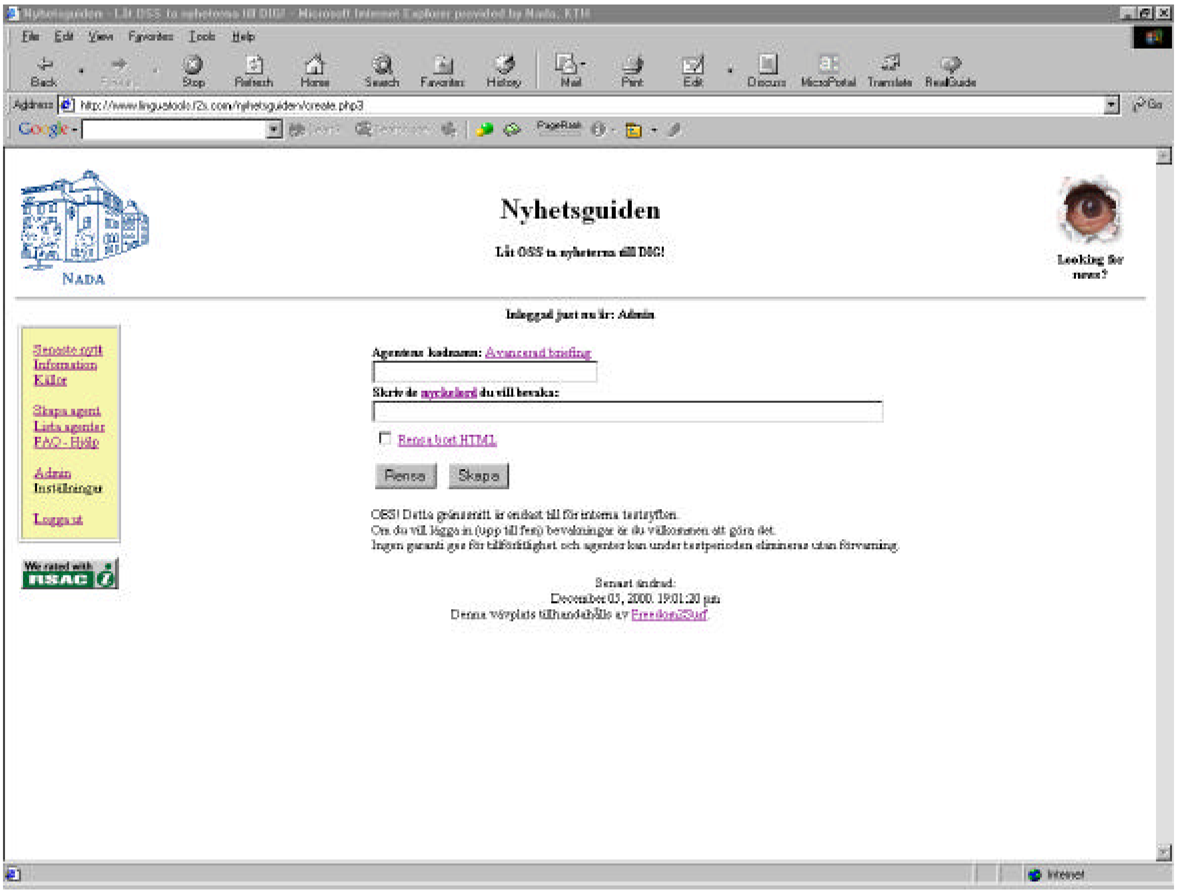
O. Knutsson 2001. *Automatisk språkgranskning av svensk text* (in Swedish), Dissertation for the Degree of Licentiate of Philosophy, Kungliga Tekniska Högskolan, NADA.

Appendix A
Screenshots from newsAgent and Nyhetsguiden.



```
SVD 11:09 REdda barnens generalsekreterare avgσr
12:36 5 news text(s) retrieved from Svenska Dagbladet
Mobil 10:47 Nu lanseras GSM-modul till Palm
12:36 1 news text(s) retrieved from Tidningen Mobil
bitWeb 12:27 Bokslutskommunik8 f÷r verksamhetsσret 2000
bitWeb 12:00 Ulf Claesson utses till Chief Operating Officer f÷r Gatespace Inc.
bitWeb 11:45 Q-Med omsΣttning upp 87 procent
bitWeb 11:25 TjΣnstemannaf÷rhandlingar inom transportsektorn avbrutna
bitWeb 11:22 ABB hjΣlper Arla Foods spσra produktion av fruktyoghurt
bitWeb 11:22  ElektronikGruppen vΣxer genom f÷rvΣrv av elektronikf÷retaget Sinco
tron
bitWeb 11:15 SWECO AB (publ) Bokslutskommunik8 2000
bitWeb 11:11 Skanska f÷rvandlar p-huset Elefanten till 150 bostadsrΣtter - inled
er 450-miljonersprojekt i Stockholm city
bitWeb 11:08 Boss Media rapporterar f÷rsΣljning av licenskasino
12:36 9 news text(s) retrieved from bitWeb (Sifo Group)
News texts counted - 9813 so far!
12:40 Using FTP to upload latest news headlines
ProFTPD 1.2.0pre10 Server (ProFTPD Freedom 2 Surf Installation) [server0026.free
dom2surf.net]
Opening ASCII mode data connection for newnews.php3.
Transfer complete.
Opening ASCII mode data connection for newnews.wml.
Transfer complete.
```

Picture 1: The console for the NewsAgent engine.



http://www.linguatools.f2s.com/nyhetsguiden/create.php3

Nyhetsguiden

Låt OSS ta nyheterna till DIG!

Inloggad just nu är: Admin

Agentens kodnamn: Avancerad briefing

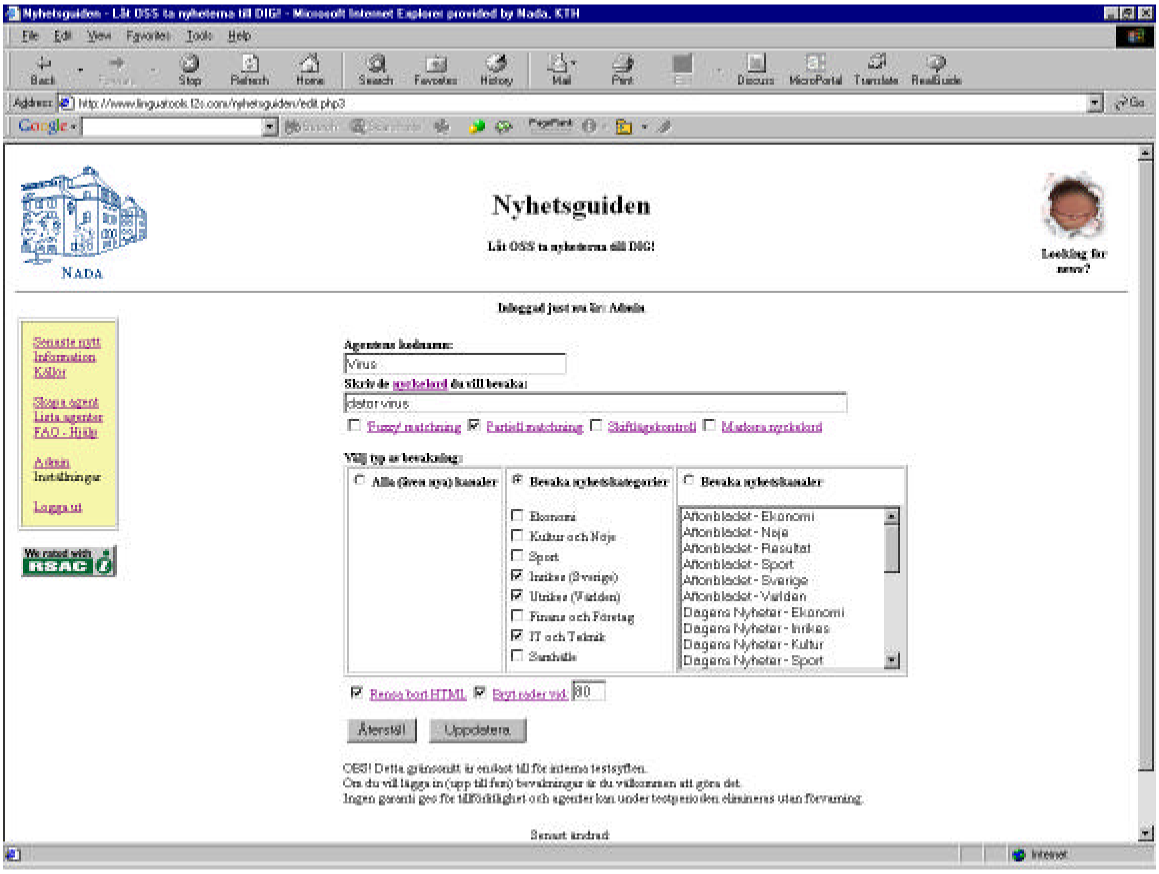
Skriv de nyckelord du vill bevaka:

□ Rensa bort HTML

OBS! Detta gränssnitt är endast till för interna testsyften.
Om du vill lägga in (upp till fem) bevakningar är du välkommen att göra det.
Ingen garanti ges för tillförlitlighet och agenter kan under testperioden elimineras utan förvarning.

Senast ändrad:
December 05, 2000. 19:01:20 pm
Denna webplats tillhandahålls av Freedom2Surf.

Picture 2: Configuring an agent the easy way.



Nyhetsguiden - Låt OSS ta nyheterna till DIG! - Microsoft Internet Explorer provided by Nada, KTH
Address http://www.linguatools.f2s.com/nyhetsguiden/edit.php3
NADA
Looking for news?
Senaste nytt
Information
Källor
Skapa agent
Lista agenter
FAQ - Hjälp
Admin
Inställningar
Logga ut
Återställ
Uppdatera

Nyhetsguiden

Låt OSS ta nyheterna till DIG!

Inloggad just nu är: Admin

Agentens kodnamn: Virus

Skriv de nyckelord du vill bevaka: dator virus

☐ Fuzzy matchning ☑ Partiell matchning ☐ Skiftlägeskontroll ☐ Markera nyckelord

Välj typ av bevakning:

○ Alla (även nya) kanaler

● Bevaka nyhetskategorier

- ☐ Ekonomi
- ☐ Kultur och Nöje
- ☐ Sport
- ☑ Inrikes (Sverige)
- ☑ Utrikes (Världen)
- ☐ Finans och Företag
- ☑ IT och Teknik
- ☐ Samhälle

○ Bevaka nyhetskanaler

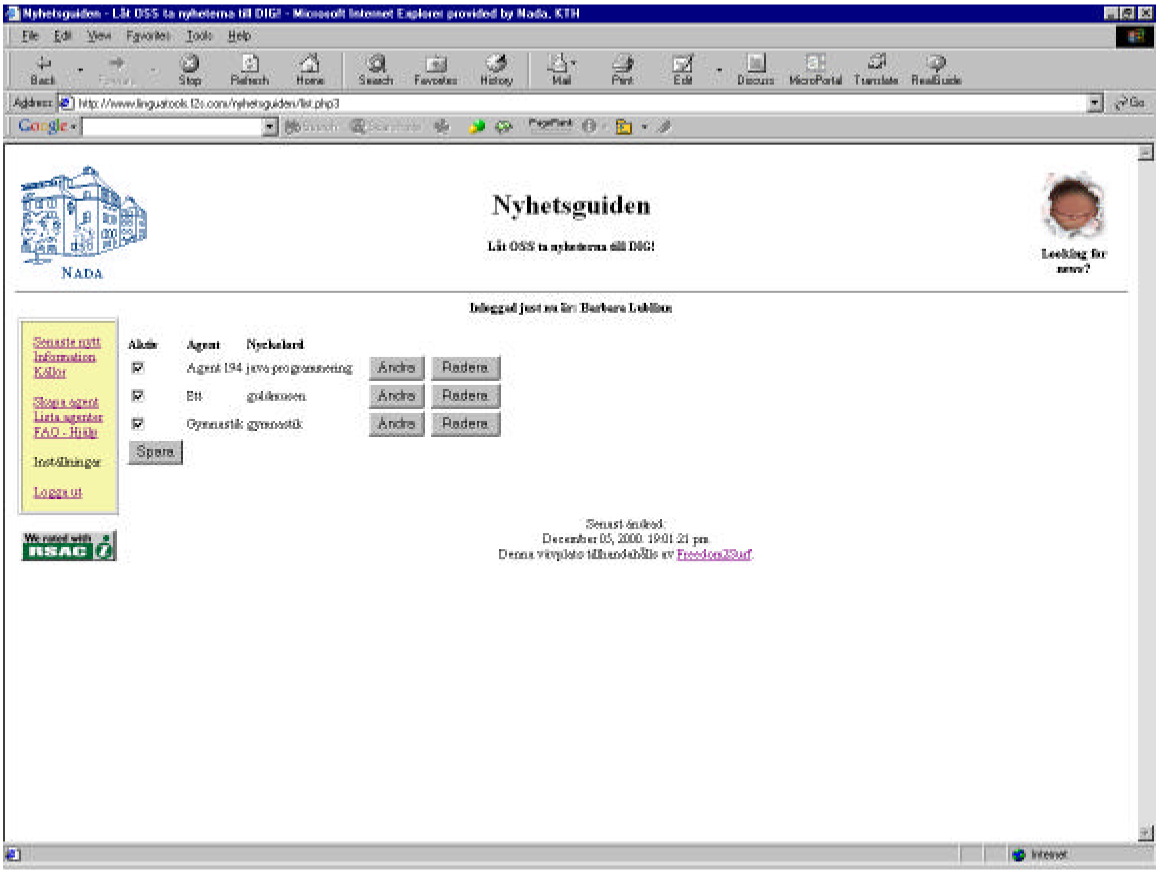
- Aftonbladet - Ekonomi
- Aftonbladet - Nöje
- Aftonbladet - Resultat
- Aftonbladet - Sport
- Aftonbladet - Sverige
- Aftonbladet - Världen
- Dagens Nyheter - Ekonomi
- Dagens Nyheter - Inrikes
- Dagens Nyheter - Kultur
- Dagens Nyheter - Sport

☑ Rensa bort HTML ☑ Bryt rader vid 80

OBS! Detta gränssnitt är endast till för interna testsyften.
Om du vill lägga in (upp till fem) bevakningar är du välkommen att göra det.
Ingen garanti ges för tillförlitlighet och agenter kan under testperioden elimineras utan förvarning.

Senast ändrad

Picture 3: The advanced mode gives more power to users with specific needs.



Nyhetsguiden - Låt OSS ta nyheterna till DIG! - Microsoft Internet Explorer provided by Nada, KTH
Address http://www.linguatools.f2s.com/nyhetsguiden/list.php3
NADA
Looking for news?
Senaste nytt
Information
Källor
Skapa agent
Lista agenter
FAQ - Hjälp
Inställningar
Logga ut
Spara

Nyhetsguiden

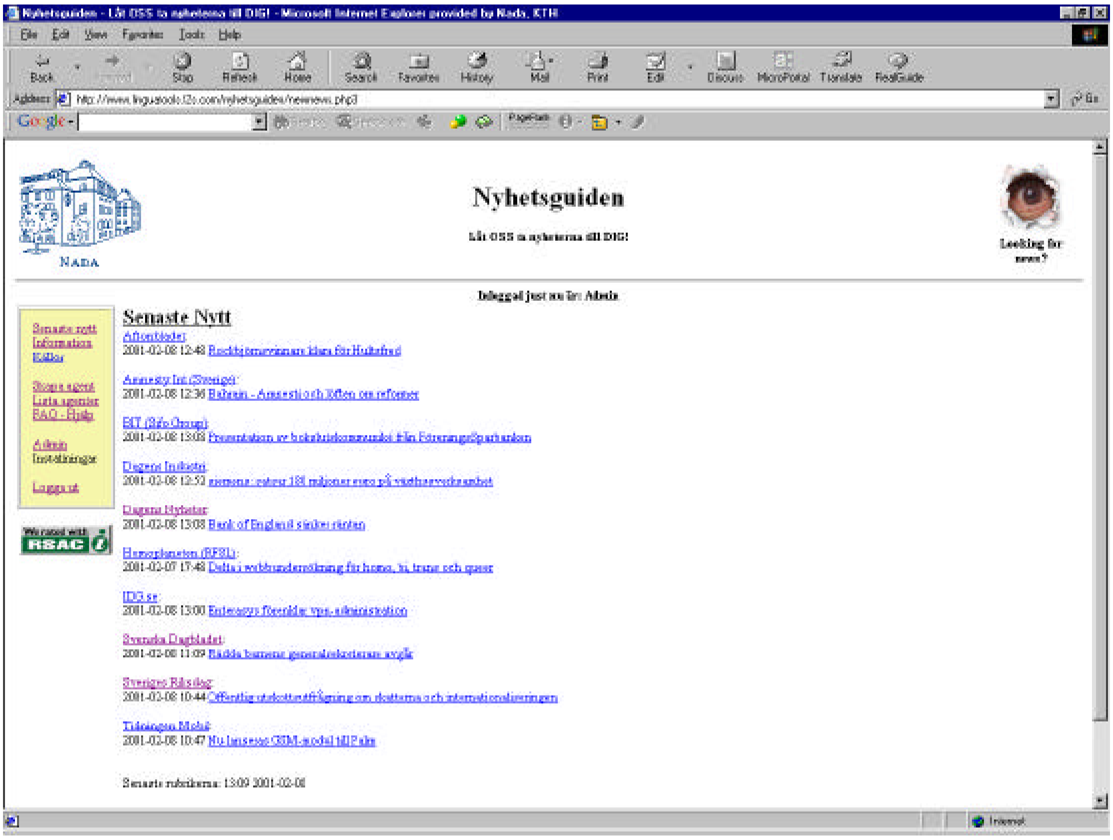
Låt OSS ta nyheterna till DIG!

Inloggad just nu är: Barbara Lublin

Aktiv	Agent	Nyckelord		
☑	Agent 194	java programmering	Ändra	Radera
☑	Ett	guldkorn	Ändra	Radera
☑	Gymnastik	gymnastik	Ändra	Radera

Senast ändrad:
December 05, 2000. 19:01:21 pm.
Denna webbplats tillhandahålls av Freedom2Surf.

Picture 4: Agents can be activated/deactivated, re-briefed or deleted.



Nyhetsguiden - Låt OSS ta nyheterna till DIG! - Microsoft Internet Explorer provided by Nada, KTH
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss MicroPortal Translate RealGuide
Address
Go
NADA
Looking for news?
Internet

http://www.linguatools.f2s.com/nyhetsguiden/newnews.php3

Nyhetsguiden

Låt OSS ta nyheterna till DIG!

Inloggad just nu är: Admin

Senaste nytt
Information
Källor

Skapa agent
Lista agenter
FAQ - Hjälp

Admin
Inställningar

Logga ut

Senaste Nytt

Aftonbladet:
2001-02-08 12:48 Rockbjörnvinnare klara för Hultsfred

Amnesty Int (Sverige):
2001-02-08 12:36 Bahrain - Amnesti och löften om reformer

BIT (Sifo Group):
2001-02-08 13:08 Presentation av [illegible] från Föreningssparbanken

Dagens Industri:
2001-02-08 12:53 siemens: satsar 181 miljoner euro på [illegible]

Dagens Nyheter:
2001-02-08 13:08 Bank of England sänker räntan

Homoplaneten (RFSL):
2001-02-07 17:48 Delta i webbundersökning för homo, bi, trans och queer

IDG.se:
2001-02-08 13:00 Enterasys förenklar vpn-administration

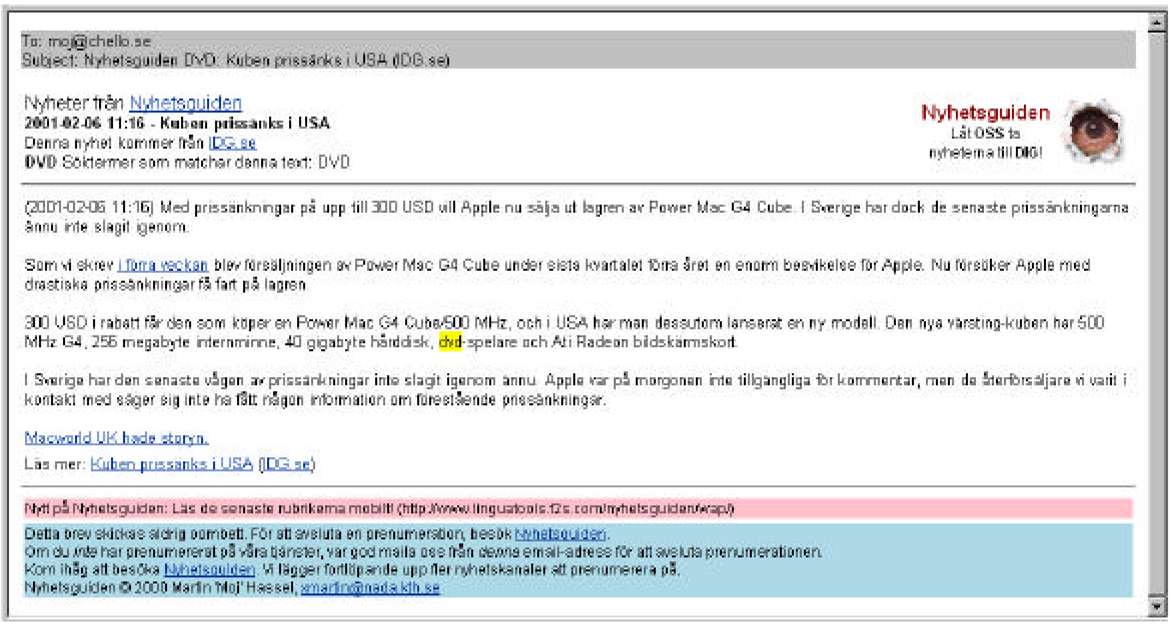
Svenska Dagbladet:
2001-02-08 11:09 Rädda barnens generalsekreterare avgår

Sveriges Riksdag:
2001-02-08 10:44 Offentlig utskottsutfrågning om skatterna och internationaliseringen

Tidningen Mobil:
2001-02-08 10:47 Nu lanseras GSM-modul till Palm

Senaste rubrikerna: 13:09 2001-02-08

Picture 5: Browsing the latest headlines in the web interface.



To: moj@chello.se
Subject: Nyhetsguiden DVD: Kuben prissänks i USA (IDG.se)

Nyheter från Nyhetsguiden
2001-02-06 11:16 - Kuben prissänks i USA
Denna nyhet kommer från IDG.se
DVD Söktermer som matchar denna text: DVD

(2001-02-06 11:16) Med prissänkningar på upp till 300 USD vill Apple nu sälja ut lagren av Power Mac G4 Cube. I Sverige har dock de senaste prissänkningarna ännu inte slagit igenom.

Som vi skrev i förra veckan blev försäljningen av Power Mac G4 Cube under sista kvartalet förra året en enorm besvikelse för Apple. Nu försöker Apple med drastiska prissänkningar få fart på lagren.

300 USD i rabatt får den som köper en Power Mac G4 Cube/500 MHz, och i USA har man dessutom lanserat en ny modell. Den nya värsting-kuben har 500 MHz G4, 256 megabyte internminne, 40 gigabyte hårddisk, dvd-spelare och Ati Radeon bildskärmskort.

I Sverige har den senaste vågen av prissänkningar inte slagit igenom ännu. Apple var på morgonen inte tillgängliga för kommentar, men de återförsäljare vi varit i kontakt med säger sig inte ha fått någon information om förestående prissänkningar.

Macworld UK hade storyn.

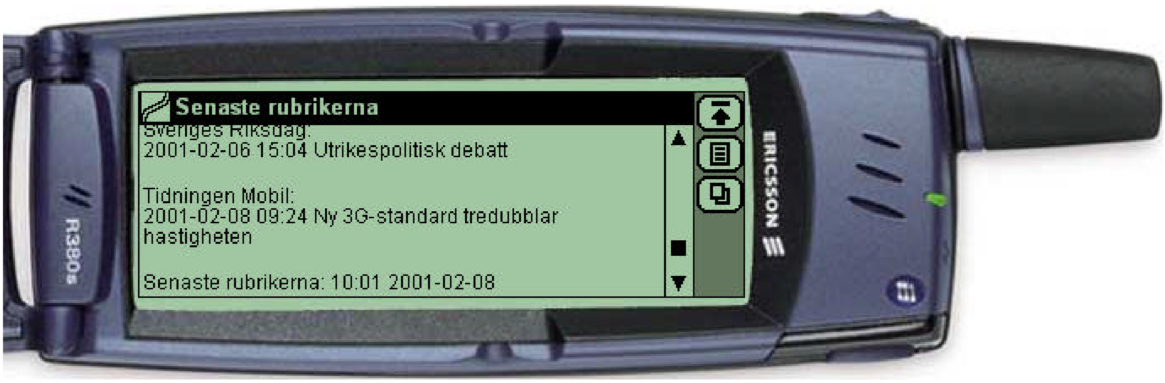
Läs mer: Kuben prissänks i USA (IDG.se)

Nytt på Nyhetsguiden: Läs de senaste rubrikerna mobilt! (http://www.linguatools.f2s.com/nyhetsguiden/wap/)

Detta brev skickas aldrig oombett. För att avsluta en prenumeration, besök Nyhetsguiden.
Om du *inte* har prenumererat på våra tjänster, var god maila oss från denna email-adress för att avsluta prenumerationen.
Kom ihåg att besöka Nyhetsguiden. Vi lägger fortlöpande upp fler nyhetskanaler att prenumerera på.
Nyhetsguiden © 2000 Martin 'Moj' Hassel, xmartin@nada.kth.se

Picture 6: Matched news, here delivered as HTML-formatted mail.

Two examples of accessing Nyhetsguiden via WAP.



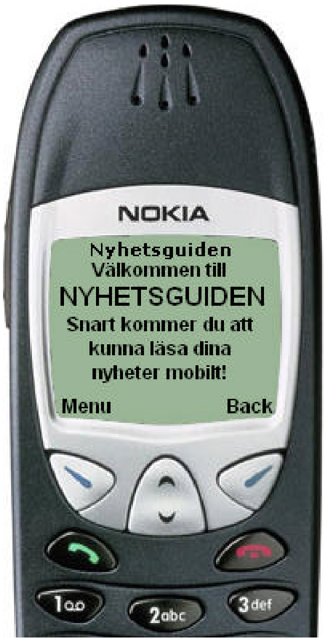
Senaste rubrikerna

Sveriges Riksdag:
2001-02-06 15:04 Utrikespolitisk debatt

Tidningen Mobil:
2001-02-08 09:24 Ny 3G-standard tredubblar hastigheten

Senaste rubrikerna: 10:01 2001-02-08

Picture 7: Browsing the latest headlines on an Ericsson R380 WAP phone.



Nyhetsguiden
Välkommen till
NYHETSGUIDEN
Snart kommer du att kunna läsa dina nyheter mobilt!

Picture 8: Nyhetsguiden through WAP.

Appendix B

Original Swedish form used in the user study.

Hej Nyhetsguideanvändare !

Ni är 25 flitiga användare i Sverige som använder sig av Nyhetsguiden. Vi vill ställa några frågor till dig för att ytterligare förbättra tjänsten, som tack för besväret skickar vi en biocheck till dig. Svara genom att fylla i direkt i mailet och maila tillbaka till mig. Gärna innan den 26 januari 2001

1) Har denna tjänst varit till någon nytta för dig?

Ja/Nej/Vet inte

2) Kommer det för många nyheter till dig?

Ja/Nej/Vet inte

3) Kommer det för många likadana nyheter eller nyheter med samma innehåll?

Ja/Nej/Vet inte

4) Är nyheterna för långa?

Ja/Nej/Vet inte

5) Vilka kanaler eller områden bevakar du?

Inrikes/Utrikes/Sport/Ekonomi/Teknik/Samhälle/Annat

6) Är det några andra områden som du skulle vilja ha möjlighet att bevaka?

7) Vilken ålder och vilket yrke/funktion har du i yrkeslivet.

8) Skulle du vilja ha en funktion som bara släpper fram "nya" nyheter till dig och inte nyheter om sådant som du precis läst?

Ja/Nej/Vet inte

9) Skulle du vilja ha en funktion som kortar ner nyheterna till 2-3 rader?

Ja/Nej/Vet inte

10) Skulle du vilja ha en funktion som kortar ner nyheterna till 2-3 rader centrerat kring det som de nyckelord du angivit bevakar? T.ex. om du bevakar ordet Ericsson så blir bara texten kring ordet Ericsson skickat till dig, så att du inte behöver läsa om t.ex. Volvo eller andra företag också om de nämns i nyheten.

Ja/Nej/Vet inte

11) Skulle du vilja ha en funktion som sammanställer flera liknande nyheter till en nyhet på 8-10 rader?

Ja/Nej/Vet inte

12) Vilken typ av matchning av nyckelorden/sökorden använder du?

Hel Matchning/Partial Matchning/Fuzzy Matchning/Vet ej

13) Tycker du det är svårt att formulera din sökfråga?

Ja/Nej/Vet inte

14) Övriga kommentar

Tack för hjälpen och glöm inte att skicka med din vanliga adress så att vi kan posta biochecken. Gärna innan den 26 januari 2001

Hälsningar
Hercules Dalianis
Projektledare – SeaSum

Dr. Hercules Dalianis		Numerical Analysis and Computing Science,Nada
ph:	+46 8 790 91 05	Royal Institute of Technology, KTH
mobile ph:	+46 70 568 13 59	SE-100 44 Stockholm
fax:	+46 8 10 24 77	Sweden
email:	hercules@nada.kth.se	
www:	http://www.nada.kth.se/~hercules/	

Appendix C
User comments in Swedish.

- Bra tjänst det här :-)

- Jag tycker att det är en jättebra tjänst.

- Bra system, jag håller mig uppdaterad om t.ex. Linux, någonstans kan man ju poängtera att Nyhetsguiden inte kan svara för att det inte kommer några mail om t.ex. Ola Knutsson.

- Bra grej och koncept. Intressant teknik. Trist webbsida - skulle behöva en uppfräschning. Enklare text/kommunikation på webbsidan. Kan säkert bli något stort om ni utvecklar och marknadsför det på rätt sätt. Vi skulle behöva någon typ av "specialbevakning" på våra kunder och de branscher de verkar i. Ni behöver en presskampanj och publicitet!

- Bra tjänst.

- Även om jag har svarat nej på 13 så kan det vara bra att ibland erhålla en blänkare om hur man kan ändra eller lägga till saker i sin sökning. Man tenderar ju att lägga in ett sökbegrepp vid ett tillfälle och sen får det "tuffa och gå".

- Bra tjänst som ännu är i sin linda.

- Inga, förutom att det ibland har kommit in lite för mycket post pga. dålig matchning från min sida. Men sådant går att rätta till.

- Nya layouten på era nyhetsmail bra!!

Appendix D
News sources and categories used by newsAgent.

Source:	**Categories:**
Aftonbladet	- Economics, cultural, sports, domestic and foreign news
Amnesty International	- Press releases and news on human rights
BIT.se (Sifo Group)	- Press releases from companies
Dagens Industri	- News on the industrial market
Dagens Nyheter	- Economics, cultural, sports, domestic and foreign news
Homoplaneten (RFSL)	- News concerning rights of the homosexual community
Tidningen Mobil	- News articles on mobile communication
International Data Group	- News articles on computers
Medströms Förlag	- News articles on computers
Senaste Nytt.com	- News flashes (discontinued)
Svenska Dagbladet	- News flashes
Svenska Eko-nyheter	- News flashes
Sveriges Riksdag	- Press releases from the Swedish Parliament