



governmentattic.org

"Rummaging in the government's attic"

Description of document: **Documents labeled 'non-responsive' and removed/redacted in previous FOIA Case 47415**

Requested date: 3-April-2007

Released date: 26-September-2007

Posted date: 18-October-2007

Date/date range of document: March – May 1994

Source of document: NSA FOIA Requester Service Center:
National Security Agency
Attn: FOIA/PA Office (DJ4)
9800 Savage Road, Suite 6248
Ft. George G. Meade, MD 20755-6248
Telephone: (301) 688-6527
Fax: (443) 479-3612

Notes: Among other topics:
AQUAINTANCE - sorting/retrieval of data
PARENTAGE - computer processing and
visualization of data and linked data

The governmentattic.org web site ("the site") is noncommercial and free to the public. The site and materials made available on the site, such as this file, are for reference only. The governmentattic.org web site and its principals have made every effort to make this information as complete and as accurate as possible, however, there may be mistakes and omissions, both typographical and in content. The governmentattic.org web site and its principals shall have neither liability nor responsibility to any person or entity with respect to any loss or damage caused, or alleged to have been caused, directly or indirectly, by the information provided on the governmentattic.org web site or in this file



NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND 20755-6000

FOIA Case: 52660A
26 September 2007

This responds to your Freedom of Information Act (FOIA) request of 3 April 2007 for "a copy of the documents labeled 'non-responsive' and removed/redacted within the previous FOIA Case 47415." A copy of your request is enclosed. The information previously redacted as 'non-responsive' in our response to case 47415 has been processed under the FOIA and the documents you requested are enclosed. Certain information, however, has been deleted from the enclosures.

Some of the information deleted from the documents was found to be currently and properly classified in accordance with Executive Order 12958, as amended. This information meets the criteria for classification as set forth in Subparagraph (g) of Section 1.4 and remains CONFIDENTIAL as provided in Section 1.2 of the Executive Order. The information is classified because its disclosure could reasonably be expected to cause damage to the national security. Because the information is currently and properly classified, it is exempt from disclosure pursuant to the first exemption of the FOIA (5 U.S.C. Section 552(b)(1)).

In addition, this Agency is authorized by various statutes to protect certain information concerning its activities, as well as the names of its employees. Such information is exempt from disclosure pursuant to the third exemption of the FOIA which provides for the withholding of information specifically protected from disclosure by statute. The specific statute applicable in this case is Section 6, Public Law 86-36 (50 U.S. Code 402 note). We have determined that such information exists in these records, and we have excised it accordingly.

Subsection (b)(2) of the FOIA exempts from disclosure matters related solely to the internal personnel rules and practices of an agency. This exemption has been held to apply to matters that are "predominantly internal," the release of which would "significantly risk circumvention of agency regulations or statutes." Crooker v. Bureau of Alcohol, Tobacco, and Firearms, 670 F.2d 1051, 1074 (D.C. Cir. 1981). Information contained within the enclosures meets the criteria for Exemption (b)(2) protection as that statutory

provision has been interpreted and applied by the Federal Judiciary. The information being protected under Subsection (b)(2) is limited to e-mail routing information that appears before the text of a message and would reveal how NSA's information network is constructed. The release of such information could expose the network to unauthorized access and is also classified CONFIDENTIAL.

Also, information regarding other individuals has been deleted from the enclosures in accordance with 5 U.S.C. 552 (b)(6). This exemption protects from disclosure information which would constitute a clearly unwarranted invasion of personal privacy. In balancing the public interest for the information you request against the privacy interests involved, we have determined that the privacy interests sufficiently satisfy the requirements for the application of the (b)(6) exemption.

Since these deletions may be construed as a partial denial of your request, you are hereby advised of this Agency's appeal procedures. Any person denied access to information may file an appeal to the NSA/CSS Freedom of Information Act Appeal Authority. The appeal must be postmarked no later than 60 calendar days from the date of the initial denial letter. The appeal shall be in writing addressed to the NSA/CSS FOIA Appeal Authority (DJ4), National Security Agency, 9800 Savage Road STE 6248, Fort George G. Meade, MD 20755-6248. The appeal shall reference the initial denial of access and shall contain, in sufficient detail and particularity, the grounds upon which the requester believes release of the information is required. The NSA/CSS Appeal Authority will endeavor to respond to the appeal within 20 working days after receipt, absent any unusual circumstances.

Sincerely,

A handwritten signature in black ink, appearing to read "Rhea D. Siers", with a stylized flourish at the end.

RHEA D. SIERS
Deputy Associate Director for Policy

Encls:
a/s

From: [REDACTED]
To: [REDACTED] jdcohen@po, mdamash@po,
[REDACTED]
Subject: NSASAB VISIT
Cc: [REDACTED]
Classification: (U) UNCLASSIFIED
X-[REDACTED]-In: [REDACTED] <940311181329.baaa001tO>
X-[REDACTED]-Out: mdamash [REDACTED] <940311181330.gaaa001tO>

This is a resend due to a classification problem.

I wish to express mine and [REDACTED] appreciation for the outstanding support you provided for the recent NSASAB Telecommunications Panel visit. Your briefings and demonstrations were recieved with many complementary comments from the visitors, who were impressed with the breadth and depth of your activities related to the GNI thrust. The panel is now well informed, thanks to your efforts, and in a much better position to serve the DIR's needs. Thanks also to those on the staff who orchestrated a flawless visit.'

(b) (3)-P.L. 86-36

Date: Fri, 11 Mar 94 14:07:43 -0500

From: Marc Damashek <mdamash>

To: [REDACTED]

Subject: ACQUAINTANCE description

Cc: mdamash

Reply-To: mdamash@ [REDACTED] nsa (Marc Damashek)

(b) (3)-P.L. 86-36

Classification: UNCLASSIFIED

Jim:

Here is a brief description of ACQUAINTANCE:

=====

R534 has developed an extremely simple, fast, completely general method of sorting and retrieving machine-readable text according to language and/or topic. The method is totally independent of the particular languages or topics of interest, and relies for guidance solely upon exemplars (e.g., existing documents, fragments, etc.) provided by the user. It employs no dictionaries, keywords, stoplists, stemming, syntax, semantics, or grammar; nevertheless, it is capable of distinguishing among closely-related topics - in any language - heretofore considered inseparable, and it can do so even in text containing a nontrivial error component (typically 10-15% of all characters). The technique can be quickly implemented in software on any computer system, from micro to super, and can easily be implemented in hardware as well. It is directly scalable to very large data sets (millions of documents). Practical applications include:

a) language- and topic-independent sorting and retrieval of documents satisfying dynamic criteria defined only by existing documents (I want to see more documents that look like THIS);

b) clustering of topically related documents, with no prior knowledge of the languages or topics that may be present. This activity can, if desired, automatically generate "document selectors" for use in application (a);

c) specialized sorting tasks, such as the identification of duplicate or near-duplicate documents in a large set.

Sender: [REDACTED]
Reply-To: [REDACTED]
Date: Fri, 11 Mar 94 09:15:40 EST
From: [REDACTED]
To: mdamashe [REDACTED] nsa
Subject: ASCII version

(b) (1)
(b) (2)
(b) (3)-P.L. 86-36

(b) (3)-P.L. 86-36

Date: Tue, 15 Mar 94 13:07:24 EST
From: [REDACTED]@t.nsa (T Wide Mailing)
Sender: [REDACTED]@t.nsa
Reply-To: [REDACTED]@t.nsa
To: [REDACTED]
Subject: R55 Vacancy Announcement
Classification: ~~FOUO~~
Mailclass: 4

(b)(1)
(b)(2)
(b)(3)-P.L. 86-36

submitted by: [REDACTED]
released by: [REDACTED]

(b)(3)-P.L. 86-36

Classification: ~~FOUO~~

T VACANCY ANNOUNCEMENT

POSITION: Electronic Engineer - 212FN/Sr. Electronic Engineer - 213FN
Computer Scientist - 232AN/Sr. Computer Scientist - 233AN
Crypt Mathematician - 252AN/Sr. Crypt Mathematician - 253AN

BILLET NUMBER: R55/011

GRADE: GG 12/13

ORGANIZATION: R55

LOCATION: R&E

SELECTION OFFICIAL: [REDACTED] Chief, R55, 961-8595

----- End Included Message -----

----- End Included Message -----

----- End Included Message -----

----- End Included Message -----

Sender: [REDACTED]
Reply-To: [REDACTED]
Date: Thu, 17 Mar 94 07:33:06 EST
From: [REDACTED]
To: mdamash@r53.r.nsa
Subject: Acquaintance
Classification: u
X-Sun-Charset: US-ASCII

(b)(3)-P.L.
86-36

Marc,
would you have some time to discuss the language ID portion of Acquaintance?
Is there someone else I should talk to?
Thanks,

[REDACTED]
P043

From: [REDACTED]
Date: Thu, 24 Mar 94 10:49:43 -0500
To: jdcohen, mdamash
Subject: Visit by MCTSSA for PROPELLER Demo

Here is the outcome of J22's presentation of PARENTAGE for a Marine application.

Mark G.

Begin forwarded message:

Sender: [REDACTED]
Reply-To: [REDACTED]
Classification: ~~(FOUO) FOR OFFICIAL USE ONLY~~
From: [REDACTED]

DOB & POB, please notify M51 to forward your clearance (TSSI minimum required)
for 1 year to [REDACTED] ATTN: [REDACTED]
(Security Manager/CSSO), 365-6507 (secure) or [REDACTED] (non-secure).

DATE: June 9, 1994
TIME: 10:30 a.m.
ROOM: A100
SPEAKER: [REDACTED] Z032

(b)(3)-P.L.
86-36

TITLE: Data Visualization, Clustering, and Multivariate
Nonparametric Density Estimation
ABSTRACT

In this talk, we illustrate the application of density estimation (histograms, for example) to problems in data exploration for multivariate structure. Clustering and discrimination will be considered in particular. Clusters may be associated with features such as modes in a density estimate. Adequacy of a Normality assumption for LDA or QDA can be assessed graphically. Estimation and visualization of densities in three and four dimensions via the "averaged shifted histogram" will be presented in a short videotape. We introduce a new diagram called the mode tree that captures how the location of a mode varies with the smoothing parameter. We end with an application in image processing of an MRI of a patient with a brain tumor.

HOST: Dan Heifetz

*****SNTP*****SNTP*****SNTP*****

(b)(1)
(b)(2)
(b)(3)-P.L. 86-36

Sender: [REDACTED]
Reply-To: [REDACTED]
Date: Thu, 2 Jun 1994 07:43:12 +0500
From: [REDACTED]
To: mdamash@po
Subject: Re: Acquaintance
Cc: [REDACTED]
Classification: (U) UNCLASSIFIED
X-Sun-Charset: US-ASCII
X-[REDACTED]-In: [REDACTED] <940602114516.baaa001oJ>
X-[REDACTED]-Out: mdamash [REDACTED] <940602114518.baaa001NY>

Marc - Let me try again to explain the using one problem that I would like to solve. I currently have about 11MB of saved mail messages. I would like to go thru all the saved messages and ?hash?

them and save the results.

As new messages come in each day, I would like to ?hash? them and save them in my master file(s). The thing I don't see how I can easily do is add the new messages without having to go back and reprocess data from the previously saved messages.

It could be that I don't understand the "normalization" process. I can see that all of the old messages do not need to be rehashed but can the "normailization" process be more like addition than I am realizing.

(b)(1)
(b)(2)
(b)(3)-P.L. 86-36

I have no objection in spending some time with you except that I am working on this problem along with a bunch of others and at the same time spending almost full time on the TDY travel team. If we can't work this out via e-mail I guess I will have to either come and see you or have somebody else work with you.

> From: mdamash@po Wed Jun 1 06:35 EDT 1994
> Date: Wed, 1 Jun 94 06:35:59 -0400
> From: <mdamash@po.nsa>
> To: [REDACTED]
> Subject: Re: Acquaintance
> Cc: mdamash, [REDACTED]
> Classification: (U) UNCLASSIFIED
> X-In: [REDACTED] <940601103600.baaa002cR>
> X-Out: [REDACTED] <940601103602.baaa004YB>
>

(b)(3)-P.L. 86-36

> Classification: (U) Unclassified

> Fred:

> Rather than your trying to understand what's going on from
> the source code, I suggest that you come down for a visit, bringing
> along some of the messages you'd like to work with (100-1000; we can
> probably get to them electronically as well, with a day or two to
> fiddle with the network filters that will undoubtedly be trying to
> keep them out). I'll give you as complete an explanation as you can
> stand, and we can process the data on the spot.

> I suspect the whole process is a lot less complicated than
> you may believe. Give me a call when you have a few minutes
> (961-1073).

> Regards,
> Marc Damashek

From: Jon D. Cohen <jdcohen>
Date: Thu, 2 Jun 94 12:42:46 -0400
To: [REDACTED]
Subject: acquaintance/par BAS
Cc: mdamash, jdcohen

=====

TITLE:

ACQUAINTANCE - Language- and Topic-Independent Sorting,
Retrieval, and Clustering

AUTHOR:

Marc Damashek

AFFILIATION:

National Security Agency
9800 Savage Road, Ft. Meade, MD 20755-6000
(301) 688-0863

DATE:

March 7, 1994

DESCRIPTION:

The purely statistical characterization of text in terms of n-grams (sequences of n consecutive characters) has been of interest since the earliest work on communication theory. Its inherent advantages include low sensitivity to character errors (garbles), language independence, ease of implementation, and suitability for reduction to hardware.

Since the work of Shannon, n-gram analysis has sporadically been applied to a number of straightforward (and generally uninspiring) problems in textual analysis and document processing, such as the flagging of possible spelling errors, document compression, and - perhaps more interestingly - language identification. Despite the apparent utility of n-gram analysis, it has also come to be widely believed that useful sorting according to topic lies well outside its scope: the prevailing sentiment has been that N-Grams Don't Work (for topic identification). The principal reasons for this conviction seem to include a vague feeling that sufficiently meaningful information could not possibly be captured by n-grams to render topic discrimination possible, and the observation that elements necessarily common to a large number of documents (for example, the inevitable conjunctions, prepositions, stems, case and conjugation endings, prefixes, suffixes, etc.) must be automatically suppressed if there is to be any hope of success. Heretofore, no universally applicable (i.e.,

language- and topic-independent) means of addressing the latter problem has been proposed, and the former has simply been accepted uncritically.

It turns out that n-grams can in fact be made to work, and to work so well (with no need whatever for word-based parsing, keywords, dictionaries, stop lists, etc.) that entirely new modes (and fields) of inquiry become immediately accessible. I have developed an effective, language-independent n-gram based method of sorting, clustering, and retrieving textual documents according to language and topic, and that method has been productively employed at several test sites in the federal government since 1991. In particular, the problem of suppressing commonality across a group of documents (where the group can be large or small) has been solved in a completely language-independent manner requiring no human intervention. The basic technique is fast and trivially simple (and can be implemented with the simplest of hardware and software), but it yields startling (and useful) results. Completely language-independent sorting, retrieval, and clustering are available "out of the box"; with a suitable visualization tool (we employ one such tool -- developed in-house -- for our demonstrations), one can automatically illustrate and appropriately label clusters identified in the blind from among large numbers of documents, with no prior information about the languages or topics that may be present. It is not uncommon to recognize meaningful, enlightening taxonomic relationships among the various items presented in this way.

Patent applications have been filed for ACQUAINTANCE in the United States and a number of foreign countries. NSA is actively soliciting expressions of interest from prospective licensees, and from potential collaborative researchers (via Cooperative Research and Development Agreements -- CRADAs).

=====

TITLE: PARENTAGE
 AUTHOR: Jonathan D. Cohen
 ADDRESS: National Security Agency
 9800 Savage Road
 Ft. Meade, MD 20755-6000
 (301) 688-0863

DESCRIPTION:

PARENTAGE is a set of algorithms and a computer program that serves as an aid to visualizing and sorting linked data. It was developed as an adjunct to a wide variety of applications, offering the ability to study associations and communications among many pieces of data or entities.

The program provides the means to display such associations in an intuitive, graphical manner. The ranking of linkages and messages or entities, based on their apparent importance in association networks, is accomplished easily and displayed graphically. Users can also retrieve text (even in foreign scripts), sound, and pictures associated with the data. Such background text can be searched, analyzed for content, and automatically labeled (in any language). The retrieved information, as well as the graphical displays, can be pasted into word-processing and graphics programs to make reports, viewgraphs, and publishable documents.

Linkage diagrams can be arranged with an array of automatic and manual tools. Displays can be overlaid with maps and charts and other annotation. Temporal linkages can be animated for studying communication patterns, and activity information can be plotted in a variety of formats.

Working in tandem with **ACQUAINTANCE**, **PARENTAGE** can topically cluster textual documents and map out topics, subtopics, and their relationships, providing a graphical representation of a text corpus. All algorithms are language-independent and require no dictionaries or other training.

The program provides a delightfully friendly and efficient user interface offering more than 150 menu commands,

(b)(3)-P.L. 86-36

From: Marc Damashek <mdamash>
Date: Fri, 6 May 94 07:49:11 -0400
To: [REDACTED]
Subject: Business card authorization
Cc: mdamash
Reply-To: mdamash [REDACTED] nsa (Marc Damashek)

Barb:

What is the status of the business card authorization request that I sent in mid-April? It's getting closer to the time I leave on TDY.

Thanks,

Sender: [REDACTED] (b)(3)-P.L. 86-36
Reply-To: [REDACTED]
Date: Mon, 9 May 94 16:42:18 EDT
From: [REDACTED]
To: [REDACTED]
Subject: TSIS-94 TECH TRANSFER EXHIBITS
Cc: [REDACTED]
Classification: U

TO TECH TRANSFER CONTRIBUTORS/EXHIBITORS:

This is the first scheduling update for your exhibits and demonstrations at TSIS. It is the schedule as we know it at this time and there will undoubtedly be some changes. Since its better that you have as much advanced notice to arrange your schedules and budget your time, even if there are minor schedule

adjustments, I am sending out this e-mail at this time.

The Technology Transfer Exhibits and Demonstrations are a part of the TSIS-94

program. Staffing of the exhibits will be required for the senior DDT management walk-thru (with video taping and photographs) on Tuesday (May 17th)

afternoon, early Thursday (May 19th) morning and the two hour lunch breaks on

Wednesday (May 18th) and Thursday (May 19th). May 20th morning is scheduled

for viewing by Agency employees. A detailed schedule appears later in this

e-mail. Although this is a substantial commitment, it affords you and your

technology tremendous opportunities through interaction senior management,

the TSIS participants and the Agency at-large.

The display panels will be delivered on or before the morning of 16 May and

will be set up at that time. Your participation is not required for this setup,

but I encourage you to come down and see as the areas begin to take shape. Power

should become available that morning. To ensure adequate lighting, track

lighting is requested to be installed on the ceiling, but in the event of

schedule slippage, we are attempting to acquire other forms of suitable

supplemental lighting.

FOR APPROVED DEMONSTRATIONS:

Please ensure the stand-alone hardware for your demonstration is ready for

set-up on Tuesday (May 17th) morning starting at 9:00 am. The technologies

listed below are scheduled to have demonstrations and you should arrange to

have your equipment brought over, set up and checked out by noon on Tuesday

(May 17th) in preparation for the DDT senior walk-thru that afternoon:

1. E21 HEBREW IN CONTEXT - LANGUAGE TRAINING

(TT#5)

2. R22
3. J12
4. R51
 & M5
5. R53
6. R53/MARC DAMASHEK

7. R52
8. P04
 A25

SPEECH BASED BIOMETRIC ID (TT#06)
AUTOBOND CHIP INTERCONNECT (TT#11)
SOFT DECISION MAKING/POLYGRAPH (TT#15)
" "
VISUALIZATION (PARENTAGE) (TT#18)
SORTING/RETRIEVAL (ACQUAINTANCE (TT#19)
 (Shared hardware with TT#18)
DUAL CHANNEL BEAMFORMER (TT#24)
SCREENWORK (TT#27)
" "

Think about how you intend to secure (read as protect) you equipment overnight.

Possibly a black cloth might be sufficient to discourage rummaging. We are exploring a nearby area for lockable storage if necessary.

FOR ALL TECHNOLOGY EXHIBITORS:

For all of you who worked with us to develop the text/graphics for your exhibit now's the time for you to showcase your technology. Please arrange to have your exhibit staffed by a person conversant in your technology and, for EXHIBITS 1-8 above, capable of presenting a demonstration at the R&E building interior lobby area to accommodate the following schedule:

MAY 17th (TUESDAY) 2:30 pm to 4:30 pm (DDT SENIORS REVIEW & VIDEOTAPING)
MAY 18th (WEDNESDAY) 11:55 am to 2:00 pm (TSIS PARTICIPANTS)
MAY 19th (THURSDAY) 7:45 am to 8:45 am (TSIS PARTICIPANTS)
MAY 19th (THURSDAY) 11:00 am to 1:00 pm (TSIS PARTICIPANTS)
MAY 20th (FRIDAY) 9:00 am to 11:00 am (AGENCY EMPLOYEES)

It is expected other visitors may visit your exhibit during these periods.

In addition to the contractors in attendance, on the 18th and 19th, Mr. Kennedy Wilson, who is a staff member of the House Appropriations Committee, will be attending TSIS, and a probable visitor to the demos/exhibits as well. Mr. Wilson is on the Surveys and Investigations team.

We have asked the TSIS staff to provide appropriate name tags. For

that
reason, please identify any persons other than the addresses to this
e-mail,
who will be staffing your exhibit. Please also ensure you bring a
supply of
your business cards, if you have them, and any unclassified handouts
about your
technology, which have been approved for release. Please contact [redacted]
[redacted] by

COB Thursday (May 12th) at 961-6050s or e-mail at [redacted] to confirm
your
staffing arrangements.

You assistance and cooperation are needed to showcase your technology
in the
most favorable light. We are entering the hectic part of the
execution phase and
you may see some changes, so please bear with us. This is the first
time we
are undertaking this level of participation and we will all learn from
the
experience. I hope participation will be rewarding and I thank you in
advanced for your contribution.

[redacted]
(b)(1)
(b)(2)
(b)(3)-P.L. 86-36

Sender: [redacted]
Reply-To: [redacted]
Date: Wed, 11 May 94 14:11:15 EDT
From: [redacted]
To: mdamash [redacted] nsa
Subject: CALL MICROSOFT
Cc: [redacted]
Classification: U

(b)(6)

Marc:

[redacted] from MICROSOFT called here for you but I couldn't find
your outside number. Please call her at 206 [redacted]