Progress Report on Research Responsive to Electronic Records of President George H. W. Bush

William Underwood

Collaborative Expedition Workshop #45

Advancing Information Sharing / Diverse Digital Collections/ Heterogeneous Sensitivities

November 8, 2005

Georgia Tech **Research Institute**

- National Archives and Records Administration

- Bush Presidential Library

- Army Research Laboratory

- Georgia Tech

Georgia Tech Project Staff

William Underwood, PI

Lucja Iwanska, Co-PI

Robert Simpson, Co-PI

Elizabeth Whitaker, Co-PI

Demetrius Campbell

Brian Harris

Sheila Isbell

Jason Kau

Marlit Hayslet-Keck

Sandra Laib

Matthew Underwood

Overview

- Project Objective

- George H W Bush Presidential E-Records

- Archival Processing Tools

- Decision Support for Archival Review

- Summary of Results

PERPOS Project Objective

- Support Presidential Library Archivists in processing personal computer records of Bush Presidential Administration.

File Formats of Bush Presidential E-records

- AMI Professional (SAMNA)
- DCA-RFT
- IBM DisplayWrite 2&3
- IBM DisplayWrite 4&5
- Lotus Manuscript
- MSWord for DOS
- MSWord for Windows 2.0
- MultiMate Advantage 2
- Windows Write
- WordPerfect 4.2
- WordPerfect 5.0
- WordPerfect 5.1/5.2
- WordPerfect Notebook 2.0

- Windows 3.1 Calendar
- WordPerfect Calendar 2-3
- dBase II database
- dBase III database
- dBase IV database
- Advanced Revelation database
- Borland Reflex 2.0 database
- Paradox 4.0 database
- Lotus 123 1.0 and 2.0 Worksheets
- Microsoft Excel 2.0 Worksheet
- PlanPerfect 5.1 Worksheet
- QuatroPro for DOS Worksheet
- QuatroPro for Windows 3.x Workbook
- Harvard Graphics 2.0 Chart
- Harvard Graphics 3.0 Chart

Document Types of Bush Presidential E-Records

Agenda

Attendee List

Bar Chart

Biography

Briefing (Presentation)

Briefing Memo

Decision Memo

Diary

Executive Order

Information Memo

Job Application

Letter

List of Candidates for Federal Office

Mailing List

Memo

Minutes of Meeting

National Security Directive (NSD)

Newsletter

Newswire

Nomination to Federal Office

Notes

Presidential Statement

Press Pool Report

Press Release

Referral Memo

Resume

Schedule

Signature Memo

Situation Report

Summary

Transcript of Speech

Staff Register

Telephone Call Recommendation

Transcript of News Conference

Information Technology & Telecommunications Laboratory

Georgia Tech | Research Institute

- **Archival Repository Tool (ART)**
  - ➤ **Accession**
  - ➤ **Description**
  - ➤ **FOIA Search and Case Management**
- **Archival Processing Tool (APT)**
  - ➤ **Filtering**
  - ➤ **Arrangement**
  - ➤ **Preservation**
  - ➤ **Review**

FOIA Access Exemptions

b(1) national security and foreign policy

b(2) personnel rules and practices of an agency

b(3) exempted by statute

b(4) confidential commercial information

b(5) deliberative process privilege

b(6) personal privacy

b(7) law enforcement investigations

b(8) financial institution reports

b(9) geological information about wells

PRA Access Restrictions

a(1), b(1) national security and foreign policy

a(2) appointments to Federal offices
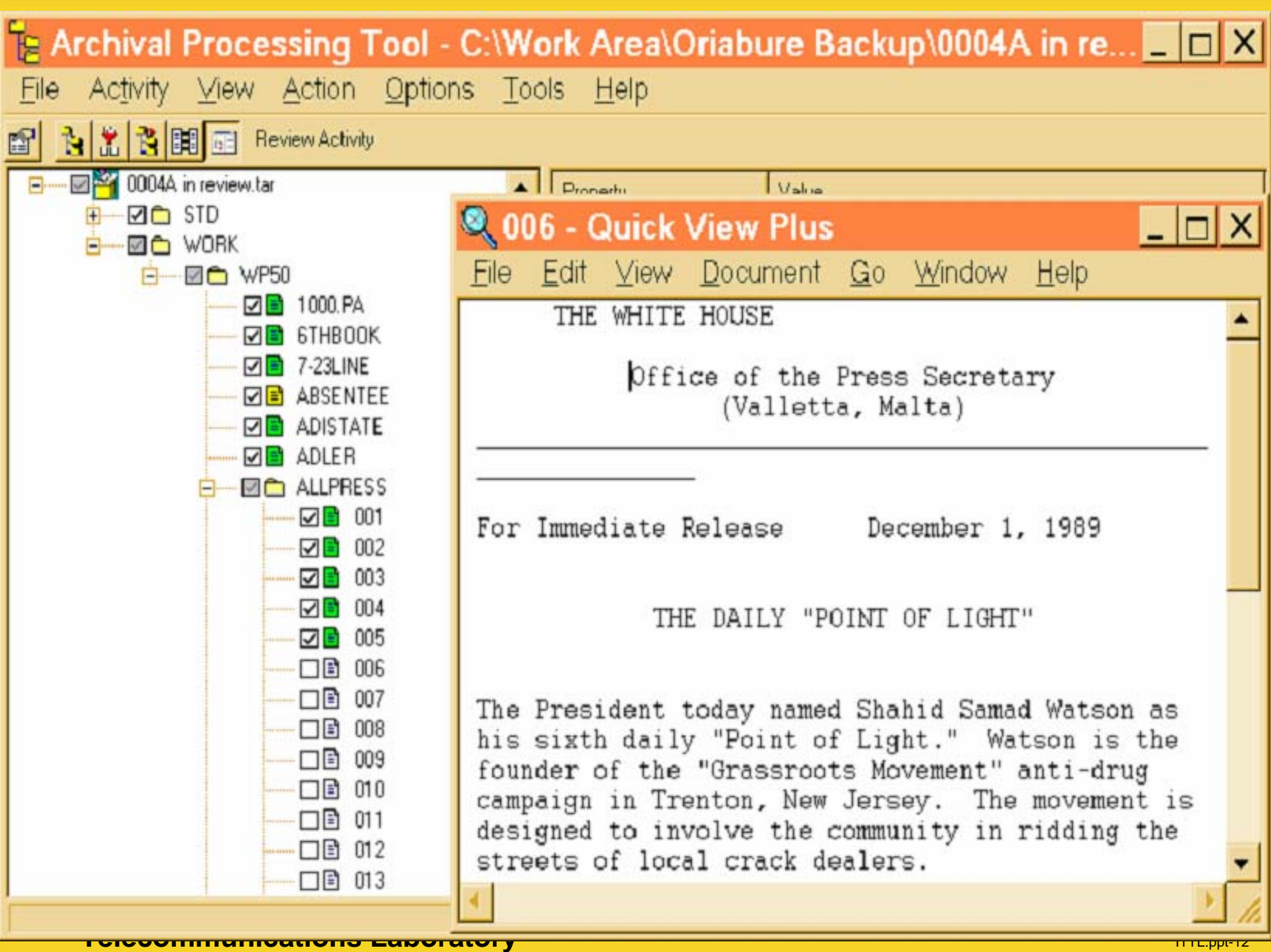
a(3) b(3) exempted by statute

a(4) b(4) confidential commercial information

a(5) confidential advice

a(6) b(6) personal privacy

PRA Restriction a(5) "Confidential Advice"

- "confidential communications requesting or submitting advice, between the President and his advisers, or between such advisers."

- This includes, but is not limited to, policy or legal advice. It includes all documentary forms containing or requesting advice including final memoranda, draft memoranda, notes from meetings, letters, etc.

The FOIA and PRA Review Problem

- Review is an intellectually demanding task.

- Requires page-by-page review.

- Increasing volume of Presidential e-records in a large variety of file formats and document types.

- Limited human resources to be applied.

- The review process is an archival processing bottleneck.

Access Restriction Checker Procedure

1. Convert the record from its original format into an html version of the document.

2. Use factual knowledge and information extraction rules to identify person's names, job titles, organization names, addresses, dates and other relevant information and markup the html version of the record.

3. Identify the document type of the record.

4. Use factual knowledge and template filling rules to fill in templates indicating the kind of communication action the record conveys, the purpose of the action, the author, addressee and its content.

5. Use personal/political record decision rules and access restriction decision rules and subsumption-based reasoning to infer from the filled in template(s) whether there is an access restriction.

6. Display the results to the archivist in the user interface.

Potential Benefits of Such a Tool

- reducing the risk of opening a document or passage of a record whose access should be restricted,

- a tutoring tool during training of review archivists.

- a tool that novice reviewers could use to check their work.

- provision of additional evidence in case a reviewer's judgment was uncertain, or point out uncertainties, where the reviewer thought the decision was certain.

- support estimation of FOIA review workload in terms of the number of restrictions and types of restrictions likely to apply.

- support reviews of Federal Records for FOIA exemptions.

Domain Knowledge

- George H. W. Bush Family Members
- President Bush's Friends
- Campaign Staff
- RNC Staff
- Presidential Nominations and Appointments to Federal Office
- White House Staff Members, Titles and Offices
- Bush Administration Senior Officials (Cabinet Secretaries and Undersecretaries)
- Presidential Advisors
- Members of 101st and 102nd Congresses
- Foreign Heads of State

Method for
Document Type Recognition

- Identify File Format

- Convert file format to ASCII or HTML

- Use Information Extraction Technology to Markup Document.

- Learn Grammatical Form of Document Types

- Use Grammars for Recognizing Document Types of other Records

Information Extraction

- *Information extraction* (IE) is a procedure that selects, extracts and combines data from text in order to produce structured information.

- *Named entity task* is to identify all named persons, organizations, locations, dates, times, numeric monetary amounts and percentages in text.

White House Correspondence

March 27, 1990

Dear Mr. Allen

Thank you very much for your letter of March 15, 1990 which stated your concerns and suggestions regarding the Americans with Disabilities Act.

In order to fulfill President Bush's campaign promise of bringing Americans with handicaps into the mainstream of American life, the Bush Administration supports the objectives of the A.D.A

As you may know, the bill is still in House Committee for consideration and change. You can be sure that your thoughts have been fully noted and are appreciated.

Sincerely,
Doug Wead
Special Assistant to the President
for Public Liaison

Ray Allen, President
American Cultural Traditions
P.O. Box 1895
Washington, D.C.20013

Information Technology & Telecommunications Laboratory

Named Entities Extracted from the Letter

```
<date>March 27, 1990</date>
<greeting>Dear</greeting><person>Mr. Allen</person>
<p>Thank you very much for your letter of <date>March 15, 1990</date>
which stated your concerns and suggestions regarding the Americans with
Disabilities Act.</p>
<p>In order to fulfill <person>President Bush's</name> campaign
promise of bringing Americans with handicaps into the mainstream of
American life, the Bush Administration supports the objectives of the
A.D.A.</p>
<p>As you may know, the bill is still in <organization>House
Committee</organization> for consideration and change. You can be sure
that your thoughts have been fully noted and are appreciated.</p>
<formula of respect>Sincerely,</formula of respect>
<person>Doug Wead</person>
<job title>Special Assistant to the President
for Public Liaison</job title>
<person>Ray Allen</person>, <job title>President</job title>
<organization>American Cultural Traditions</organization>
<postal address>P.O. Box 1895</postal address>
<location>Washington, D.C.</location> <zipcode>20013</zipcode>
```

A -><date> </date> <greeting> <greeting> <p> text </p>

A -> <greeting> <greeting> <p> text </p> <p> text </p>

....

Grammar after Applying Substitution and Recursion Operators

A -> <date> </date> <greeting> </greeting> B
C <salutation> </salutation> B<address> </address>

B ->

C -> <p> text </p> C | ∈

Examples of Speech Acts Carried Out in Presidential Records

resignation - the speech act of giving up a claim or office or possession.

appointment - the speech act of putting a person into a non-elective position.

nomination - the speech act of officially naming a candidate.

advice - the speech act of advising as to an appropriate course of action.

recommendation - the speech act of recommending something as advisable

request - the speech act of requesting

briefing - the speech act of providing detailed instructions, as for a military operation.

report - the speech act of informing by report

Rules for Filling in Speech Act Templates

If sentence is imperative,

and object of sentence is ?z,

then assert (act "request")

  assert (content ?z)

If document is memorandum,

and "From <person> ?x </person>"

and "To <person> ?y </person>"

then assert (author ?x), assert (addressee ?y)

Content Extraction Applied to Recognizing Request for Confidential Advice

December 5, 1990

MEMORANDUM FOR BOYDEN GRAY
THROUGH: BRENT SCOWCROFT
FROM: THE PRESIDENT
Boyden —

Please prepare for me a short analysis of the War Powers Resolution. . . . With out recognizing the constitutional validity of the War Powers Resolution, is there a way for the President to fulfill all his responsibilities to Congress by saying, a few days before any fighting was to begin, "hostilities are imminent—period!!

I am several thousand miles south, but these questions stay on my mind:

1.   How do we fully involve Congress?
2.   If we have to attack from a cold start how does the latest UN Resolution impact on congress?
   1. Is there something short of "declaring" war that satisfies Congress yet doesn't risk tying the President's hands?
   2. As the clock on the UN resolution keeps running toward the time when force has international authority, what possible official requests can/should a President make of Congress?

If you reply to this memorandum before I return, please hand carry your reply to Brent for "Eyes Only" transmission to me.

Please share a copy of this memorandum with John Sununu and Brent only.

Warm regards.

Communication Act Template for the Document

```
(communication_act
    (document Doc-0014)
    (act request)
    (purpose directive)
(author "The President")
(addressee "Boyden Gray")
(date "December 5, 1999)
(content "analysis of War Powers Resolution")
)
```

Decision Rule for PRA restriction a(5), Confidential Advice

- If record is a communication between the President and a presidential advisor, or the record is a communication between presidential advisors, and the purpose of the communication is a request (for action, information) or an order, and the content involves National Security Policy issues, then access is restricted under PRA a(5).

Summary of Results

- Tools for Accessioning, Arranging, Preserving, Reviewing and Describing Presidential e-records.

- Technologies for

  - Acquiring domain knowledge from textual documents

  - Extracting information from text

  - Learning and identifying document types

  - Determining speech act carried out by a document

  - Recognizing documents or passages of documents that whose access might be restricted under FOIA or PRA.

  - Supporting Archival review decisions

Additional Information

- http://perpos.gtri.gatech.edu

- william.underwood@gtri.gatech.edu

- Archival Processing Tools: User Manual

- PERPOS: Results of Laboratory Experiments and Use by Archivists, Nov 2003

- PERPOS II: Annual Technical Status Report July 1, 2004 – June 30, 2005