

**A Corpus of Presidential, Federal and Personal Records for use in
Information Extraction, Description and FOIA/PRA Review
Experiments**

William Underwood
Marlit Hayslett-Keck
CSITD/ITTL
GTRI

PERPOS Technical Report 04-5

June 2004

PERPOS II Project

The Army Research Laboratory (ARL) and the National Archives and Records Administration (NARA) sponsored this research under Army Research Office Cooperative Agreement DAAD19-03-2-0018. The findings in this paper should not be construed as an official ARL or NARA position unless so indicated by other authorized documentation.

Abstract

The Presidential Electronic Records PiLOt System (PERPOS) project is developing a system to assist archivists in processing electronic records created by office applications on personal computers (PCs). The system, called the Archival Processing Tool (APT), supports archivists in accessioning, arranging, preserving, reviewing, and describing record series. Additional tools are envisioned to further assist archivists in describing and reviewing personal computer records. These include tools for automatically identifying document types, creating better folder titles, summarizing the contents of record series, and checking for access restrictions. All of these tools depend on advanced information extraction and text understanding technology.

Records at the Bush Presidential Library are being analyzed to identify the kinds of information that need to be extracted from documents to support description and review. This analysis will also result in the generalization of rules that characterize the conditions under which one of these restrictions might apply to an entire document or to terms or passages in a document. We will also analyze Personal Misfiled Records in the Bush Presidential Library to construct the knowledge needed to discriminate personal/political records from Presidential Records. Using this knowledge we are constructing a prototype Access Restriction Checker, a document type identifier, a summarizer, and a folder titler.

Because we do not have a corpus of unprocessed Presidential, Federal and Personal records for use in our laboratory, we are constructing a corpus of "simulated" records. We are also creating some "simulated" Personal Records. The corpus also includes some records that are not subject to PRA restrictions or FOIA exemptions and have been opened for public access.

After testing of the tools on the "Test Corpus", additional experiments will be conducted at the Bush Presidential Library with actual Presidential and Personal Records from the Bush Administration's Personal Computer Files.

Another benefit of creating this collection of "simulated" Presidential Records is that we will have a collection of records that can be used with the APT for archival training in the processing (arrangement, preservation, review, and description) of digital computer records.

Table of Contents

INTRODUCTION	1
BACKGROUND	1
PURPOSE	3
SCOPE.....	3
DOCUMENT TYPES	3
PERSONAL AND POLITICAL RECORDS	5
RESTRICTION A(1) (B) NATIONAL SECURITY	7
PRA RESTRICTION A(2) "APPOINTMENTS TO FEDERAL OFFICE"	7
A(3) B(3) "EXEMPTED BY STATUTE"	9
A(4) B(4) "CONFIDENTIAL COMMERCIAL INFORMATION"	12
PRA RESTRICTION A(5) "CONFIDENTIAL ADVICE"	13
A(6) B(6) "PERSONAL PRIVACY"	16
EXEMPTION B(2) "INTERNAL AGENCY RULES AND PRACTICES"	17
FOIA EXEMPTION B(7) "LAW ENFORCEMENT INVESTIGATIONS"	18
PRESIDENTIAL RECORDS THAT HAVE NO RESTRICTIONS	20
FACTUAL KNOWLEDGE	21
FILE FORMATS	21
OFFICES, RECORD SERIES, AND FOLDERS	22
SUMMARY	23
REFERENCES	25

Introduction

Background

The Presidential Electronic Records Pilot System (PERPOS) project is developing a system to assist archivists in processing electronic records created by office applications on personal computers. The system, called the Archival Processing Tool (APT), supports archivists in accessioning, arranging, preserving, reviewing, and describing record series. During the review activity, a reviewer can view records in a file system and review them for access restrictions. The records (files) can be opened or closed in their entirety, redacted, marked as a Personal Record Misfile, or transferred to a software library, because they are misfiled software that was used to create the records, rather than being a record [1].

Additional tools are envisioned to further assist archivists in describing and reviewing personal computer records. These include tools for automatically identifying document types, creating better folder titles, summarizing the contents of record series, and checking for access restrictions. All of these tools depend on advanced information extraction and text understanding technology.

A review tool is envisioned that could further assist the archivists by checking for access restrictions. The first step in making this vision a reality was to study the Presidential Records Act (PRA) and Freedom of Information Act (FOIA) statutes, Department of Justice guidance regarding FOIA review, and NARA guidance regarding review of Presidential records [2].

Four of the six PRA restrictions are identical to four of the nine FOIA exemptions. Those PRA restrictions that are different from FOIA exemptions are (a)(2), Appointments to Federal Office, and (a)(5), Confidential Advice. The identifiers a(n) refer to the paragraphs in the PRA that specify the restrictions and the identifiers b(n) refer to paragraphs in the FOIA that specify exemptions. Taken together there are eleven different restrictions/exemptions.

- 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,
- b(2) internal Agency rules and practices,
- b(5) agency communications regarding litigation,
- b(7) law enforcement investigations,
- b(8) financial institution reports, and
- b(9) geological information about wells.

In our research, we are not addressing the following restrictions and exemptions:

- PRA a(1) (FOIA exemption b(1)) because the collection of electronic Presidential records on which we will eventually test the prototype "Access Restriction Checker" is presumed to exclude classified national security records.
- FOIA exemption b(5) because this exemption only applies to Federal records.
- FOIA exemptions b(8) and b(9) because Presidential Library reviewers seldom if ever encounter Presidential Records containing these kinds of information.

To determine the kinds of information that will need to be extracted from digital records and to construct rules that can be used with the extracted information to infer whether a restriction applies, we are examining some records at the Bush Presidential Library that have FOIA exemptions and PRA restrictions.¹

A software prototype "Access Restriction Checker" is being constructed that reads an electronic document, extracts information, applies the rules characterizing access restrictions, and displays the results. Because we do not have a corpus of unprocessed Presidential, Federal and Personal records for use in our laboratory, we are constructing a "Test Corpus" of "simulated" records that have fictitious information corresponding to access restrictions. We are also creating "simulated" Personal Records. The corpus also includes some records that are not subject to PRA restrictions or FOIA exemptions and have been opened for public access.

The George W. Bush Presidential Library is our primary source of opened Presidential records. The book, *All the Best*, by George Bush [3], is one of our sources of personal/political records.

We are also defining the types of documents that occur in the White House staff member's electronic files. A document type is a class of documents having the same form; for example, a letter, memo, or meeting agenda.

Experiments will be conducted to determine whether the patterns defining these document types can be matched against automatically marked up documents to determine an electronic document's document type. Automatic identification of the document type of an electronic record enables identification of such information as the date of the document, the parties involved and the subject of the document.

After testing the "Access Restriction Checker" on the "Test Corpus", experiments will be conducted at the Bush Presidential Library with actual Presidential Records and Personal

¹ The project staff members who analyze these records have signed a nondisclosure agreement and will examine the records with the help of Bush Presidential Library Staff. The PERPOS project staff also have security clearances, but will not be examining classified National Security records.

Records from the Bush PC Files. The accuracy of the "Access Restriction Checker" will be assessed for each of the categories that it is engineered to check.

Purpose

The kinds of documents subject to various PRA restrictions and FOIA exemptions were identified in a previous study [2]. The purpose of this paper is to report on examples of fictitious documents that have been created and that simulate information that would be subject to these restrictions. A second purpose is to report on sample documents that have been collected of personal/political papers and of Presidential records that do not have restrictions.

Scope

In the next section, we describe the document types that we have collected and the number of each type. Next, we list the kinds of records that would be considered Personal/Political records as opposed to Presidential Records, and the document identifiers for the examples of these kinds that we have collected. In subsequent sections, we go through each of the PRA restrictions and FOIA exemptions, indicating the kinds of documents collected or created. Next, we describe that examples of records that do not have access restrictions. Then we describe the types of factual knowledge that will be needed by the Access Restriction Checker, and that we have begun to collect. We then describe how the corpus is structured into aggregations by office, record series, and folders (directories) Finally, we describe the kinds of file formats in which these sample documents are represented.

Document Types

Whether a document is a personal record or a Presidential record, or whether it might be restricted is often dependent on document type. For example, a diary or recipe is likely to be personal record, while a memo or press release is more likely to be Presidential record. Furthermore, if a document is a memo, knowing whom it is to or from, and its subject will be factors in determining whether it might involve confidential advice or be related to an appointment to Federal Office. Thus the Access Restriction Checker will need to have the capability to recognize document types. This means that our "Test Corpus" should contain the types of documents that occur in the Bush Administrations Personal Computer files. The "Test Corpus" currently includes documents of the following types.

Document Type	Number of Documents
Agenda	0
Attendee List	4
Bar Chart	0
Biography	1
Briefing (Presentation)	0
Briefing Memo	0
Decision Memo	3
Diary	1
Executive order	3
Grocery list	0
Information Memo	3
Job Application	0
Journal	0
Letter	54
List of Candidates	1
Mailing List	0
Memo	16
Memo from President	6
Minutes	0
National Security Directive (NSD)	1
Newsletter	1
Nomination to Federal Office	1
Notes	2
Presidential Statement	1
Press Pool Report	3
Press Release	8
Recipe	1
Referral Memo	1
Resume	4
Schedule	2
Signature Memo	2
Situation Report	1
Summary	1
Transcript of Speech	0
Staff Register	0
Telephone Call Recommendation`	1
Transcript of News Conference	3
Total	125

Some document types have standard names. For example, materials for the President should be in the form of Briefing Memoranda, Decision Memoranda, Signature Memoranda, Information Memoranda and Telephone Call Recommendations.² The US Government Correspondence Manual defines the preferred form of internal memoranda between staff members or between the White House Staff and government agencies.³ We are creating idiosyncratic names for some of the document types, for example, attendee lists, and list of candidates.

² *White House Correspondence Manual*, Bush Presidential Library and Museum.

³ *US Government Correspondence Manual*, US General Services Administration, 1992.
<http://www.va.gov/publ/gsa/corresp/>

While a Press Release is a document type, there are types of Press Releases, which can also be considered a document type or subtype, for example, Announcement of Appointment, Presidential Statement, Statement by the Press Secretary.

Personal and Political Records

The Presidential Records Act defines personal records as follows:

“The term "personal records" means all documentary materials, or any reasonable segregatable portion thereof, of a purely private or nonpublic character which do not relate to or have any effect upon the carrying out of the constitutional, statutory, or other official or ceremonial duties of the President. Such term includes

- a) Diaries, journals, or other personal notes serving as the functional equivalent of a diary or journal, which are not prepared or utilized for, or, circulated or communicated in the course of, transacting Government business.
- b) Materials relating to private political associations, and having no relation to or direct effect upon the carrying out of constitutional duties of the President; and
- c) Materials relating exclusively to the President’s own election to the office of the Presidency; and materials directly relating to the election of a particular individual or individuals to Federal, State or local office which have no relation to or direct effect upon the carrying out of constitutional, statutory, or other official or ceremonial duties of the President."

To represent the knowledge needed and to test the Access Restriction Checker's ability to distinguish personal/political records from Presidential Records, we have begun to collect or simulate personal/political records of the following types.

Reason for Withdrawal (Personal Record)	Document Id(s)	Source or Simulated
personal diary, journal or personal notes	004	All the Best pp 411-428
staff resume for a post-administration job		
job application for a position outside the White House Offices		
recipe	122	Barbara Bush ⁴
Christmas card mailing list		
grocery list,		
Correspondence with friends expressing sentiments, e.g., "Congratulations," "Happy Birthday," "Thanks for the Gift."		
Bush Presidential Campaign Material or correspondence	044	Simulated(personal/political)

⁴ <http://cookie.allrecipes.com/az/BarbrBshsChcChps.asp>

Document relating to an individual's campaign for local, state, or federal office.		
Letter from the Chairman of the RNC.	043	Simulated(personal/political)
Document sent to the RNC or to Lee Atwater.	016	All the Best p. 417
Document that contains phrases such as "I was saying to Lee Atwater, ..." "I sent so and so to Lee Atwater."		
RNC staff register		
record concerning the President's plans for attending a funeral of a personal friend, rather being a state occasion.		
Material pertaining to a personal appointment, rather than an appointment for Presidential business.		
Letter from RNC to First Lady	103	Simulated
communication with a family member that does not involve a constitutional duty.	005 006	All the Best p. 413 All the Best p. 457
First Lady's records - personal correspondence.	001 106 107	⁵ 6 7
Poll on political issues by private pollsters, e.g., "How many Republicans in a predominately Democratic District?" "What can we do to get Women or Blacks to vote for you?"		

The kinds of documents that are personal/political papers were identified in a previous study [2]. A unique document identifier is assigned to each file representing a type of Personal/Political record. If the record is an actual record that has been made available by the one of the persons involved, its source will be indicated. If the record is "simulated", that will be indicated.

At least one document will be collected (or simulated) for each of the types of Personal/Political records in the table, but if there are significant differences within a type of record, additional subtypes will be created.

Presidential Records that are in contrast to the Personal/Political Papers

Some of the types of Personal/Political records shown in the prssssevious table, could be Presidential records if they involved carrying out the President's constitutional duty. The following table shows examples of such records.

Reason that Record is not Personal but Presidential	Document Id(s)	Source or Simulated
Letter from Assistant to President in thanking a citizen for a letter to the President	112	Bush Library, Opened
Letter from First Lady relating to carrying on the President's constitutional responsibility.		

⁵ Special Collections and University Archives of Rutgers University/Mary G. Roebling Papers. <http://www.scc.rutgers.edu/wild/>

⁶ http://www.dvdfile.com/software/review/dvd-video_8/simpsons_s4.html

⁷ <http://www.aasrcleveland.org/aasr/charity.htm>

Poll on national/nonpolitical issues carried out by a Govt. pollster.	POSsurvey_briefing_oct15-18_1991_p1.txt	Bush Library, Opened
A document concerning the President or First Lady attending a state funeral.		
A communication to or from family members, that involves a constitutional duty	027, 028, 031	⁸
Communication from ambassador to President with personal info, but Presidential Record	035	Bush Library, Opened
Communication with congressman who is a friend regarding Political/Presidential duties	009	All the Best, p 481
Personal letter to Former President that is also Presidential	015	All the Best, p 546

Restriction a(1) (b) National Security

We are not focusing on identifying documents that should be withdrawn for reasons of national security, but have collected examples of a few documents that were previously so restricted, but are now open.

Reason for Withdrawal	Document Id(s)	Source or Simulated
Correspondence with Head of State	007 008	All the Best p. 433 All the Best pp 433-434
National Security Directive	014	⁹
Memo to Chief of staff regarding foreign policy	011	All the Best, p 441

PRA Restriction a(2) "Appointments to Federal Office"

This category restricts from release records “relating to appointments to Federal office.” Disclosure of these kinds of records is prohibited for 12 years after the end of a Presidential Administration. For purposes of testing of the Access Restriction Checker, we have to create "simulated" documents with fictitious persons and facts that nevertheless reflect the kind of information that make such a document restricted.

Currently, there are about 3000 Presidential appointments, with about 1000 requiring nomination and Senate approval.¹⁰ Those requiring Senate approval are nominations of Supreme Court Justices, appellate court judges and district court judges. Other Presidential nominations requiring Senate approval for appointment include the Cabinet and subcabinet, members of regulatory commissions, US Attorneys, ambassadorships and members of numerous advisory boards. There are an additional 2500 positions that fall under White House Staff.

⁸ Bush Presidential Library, Bush Presidential Records, Counsel's Office, C. Boyden Gray's Files. Correspondence

⁹ Bush Presidential Library, Bush Presidential Records, National Security Council. www.bushlibrary.tamu

¹⁰ <http://www.opm.gov/feddata/html/POL0901.asp>

The table below reflects types of records that are subject to restriction a(2).

Reason for Withdrawal	Document Id(s)	Source or Simulated
Recommendations for appointments to Federal Office from personal advisors. ¹¹	054 ¹²	¹³
Recommendations for appointments to Federal Office from persons who are not personal advisors, but have been asked for advice.		
Internal memos that have lists of candidates who may not be the final choice including biographical information and resumes.	091 ¹⁴ , 108, 111, 092, 098 ¹⁵	Simulated(closed)

The kinds of documents subject to PRA restriction a(2) were identified in a previous study [2]. The following table reflects types of records that would have been closed under a(2), but for the fact that President Bush waived his restriction rights under a(2).

Waiver	Document Id(s)	Source or Simulated
Routine unsolicited request for position or job application	034, 037	¹⁶ Bush Library, opened
Resume associated with an appointment to Federal office (excepting personal data such as social security number and home telephone number per restriction a(6)b(6)).	055	¹⁷
Routine ministerial materials conveying a final action dealing with appointments to office		
a. A formal recommendation in which the final outcome is evident.	002, 053	¹⁸ ¹⁹
b. Letter of Nomination from the President to the Senate	056	²⁰
Announcement of appointment to office and accompanying press release.	116, 117, 118	Bush Public Papers
Routine letters of congratulations, resignations and acceptances of resignations		
a. Routine letter of congratulations on appointment to Federal office.		
b. Resignation from Federal office.	046 057	Bush Lib (open) ²¹

¹¹ Also restricted under a(5)

¹² However, this document is also a recommendation when the final outcome is evident.

¹³ Bush Presidential Library, President Bush Records-White House Staff and Office Files, Counsels Office, Lee Liberman Nominee Files, Leighton, Ronald B. (FOIA 98-0205-F).

¹⁴ Also restricted under a(5)

¹⁵ Also restricted under a(5)

¹⁶ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#314933.

¹⁷ Bush Presidential Library, President Bush Records-White House Staff and Office Files, Counsels Office, Lee Liberman Nominee Files, Leighton, Ronald B. (FOIA 98-0205-F).

¹⁸ Bush Presidential Library, President Bush Records-White House Staff and Office Files, Counsels Office, Lee Liberman Nominee Files, Leighton, Ronald B. (FOIA 98-0205-F).

¹⁹ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, President, Personal Representative of the President (FG001-02) document #10864255

²⁰ Bush Presidential Library, President Bush Records-White House Staff and Office Files, Counsels Office, Lee Liberman Nominee Files, Leighton, Ronald B. (FOIA 98-0205-F).

c. Acceptance of resignation from Federal Office.	047, 032	Bush Library, opened
d. Routine Letter from Nominee	024	Bush Library, opened
Routine letters of recommendations for positions		
a. External recommendations for appointment to Federal office from Federal Judges or congressmen who are not personal advisors and have not been asked for advice.	030, 029	²² ²³
b. External letter of recommendation that was not requested and is not from a Federal Judge or congressmen.	048	²⁴
c. Unsolicited recommendation from organizations including the American Bar Association (ABA).		
d. External recommendation for appointment to Task Force	072, 073	²⁵
e. Recommendation from a Staff Member for a Staff position (inc. resume)	087, 088	²⁶
f. Routine Memorandum of recommendation for Delegations	049	²⁷

a(3) b(3) "Exempted by Statute"

An attempt will be made to find and analyze documents whose contents are protected from disclosure by the statutes shown in the following table.

Reason for Withdrawal		Document Id(s)	Source or Simulated
Statute	Paragraph		
Rule 6(e)(6) FRCP ²⁸	exempts matters occurring before a Grand Jury from disclosure.		
Rule 32(e)(1) FRCP	governs the disclosure of pre-sentence reports.		
5 USC 7114(b)(4))	restricting disclosure of materials that constitute guidance, advice, counsel, or training provided for management officials or supervisors, relating to collective bargaining		

²¹ Bush Presidential Library, Bush Presidential Records, White House Staff and Office Files, Counsel's Office, Lee Lieberman's Files, W D Washington.

²² Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#315353.

²³ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#315353.

²⁴ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#347712

²⁵ Bush Presidential Library, Bush Presidential Records, WHORM Subject Files, Commodities (CM), Transportation Equipment, ID#359681.

²⁶ Bush Presidential Library, Quayle - Vice Presidential Records, VP Office, William Kristol's Chron Files, July 1990.

²⁷ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, President, Personal Representative of the President (FG001-02) document #10864255 (could also be in Office of Personnel, Chase Untermeyer's Files, Cop[ies] of Memos to President (1990)

²⁸ Federal Rules of Criminal Procedure

7 USC 12	restricting disclosure of any data or information concerning or obtained in connection with any pending investigation of any person and any information or data obtained from a foreign futures authority.		
8 U.S.C. 1202 (f)	authorizes the withholding of information pertaining to the issuance or refusal of visas and permits to enter the United States. These are actually restricted under (a)(6) and (b)(6).		
10 USC 130	restricts the release of unclassified military technical data		
10 USC 130b	restricts the disclosure of duty station address information regarding defense employees in sensitive, routinely deployable, and/or overseas units.		
10 USC 130c	Restricts access to information from a foreign government that the foreign government consider confidential.		
10 USC 1102	protects confidentiality of military medical quality assurance records.		
13 USC 8(b) and 9(a)	prohibits the use of population, agriculture and housing information furnished under the Census Act for any purpose other than the statistical purpose for which it was supplied.		
13 USC 301(g)	restricts disclosure of shipper's export declarations.	New_7525v.pdf	²⁹
15 USC 2055(a)(2)	protects from release confidential information such as trade secrets, processes, operations, style of work, or apparatus, or to the identity, confidential statistical data, amount or source of any income, profits, losses, or expenditures of any person, firm, partnership, corporation, or association. This information is also protected under a(4) b(4).		
15 USC 2055(b)(1)	requires prior notice to manufacturers or labelers of information submitted to the Consumer Product Safety Commission that is to be publicly disclosed that might disclose the identity of the manufacturer or labeler.		
18 USC 2510-20	exempts from disclosure the		

²⁹ www.census.gov/foreign-trade/regulations/forms/

	contents of electronic surveillance (wiretap intercepts) conducted in violation of the Act.		
19 USC 1677f	restricts disclosure of proprietary information submitted to the Tariff Commission.		
22 USC 1461(a)	limits the dissemination of USIA materials domestically.		
22 USC 3104(c)	restricts disclosure of Investment Survey information.		
26 USC 6103	governs disclosure of tax return information. This is actually exempted under a(6) or b(6).		
35 USC 122	prohibits the disclosure of patent applications and information concerning them.	pto-sb01.pdf	www.uspto.gov/web/forms/
41 U.S.C. 253(b)(m)	prohibits the release or disclosure of a technical, management, or cost proposal submitted by a contractor in response to the requirements of a solicitation for a competitive proposal.		
42 U.S.C. 2011	prohibits the disclosure of information concerning Department programs that is classified as Restricted Data and Formerly Restricted Data.		
42 USC 2161-66	protects Restricted Data and Formerly Restricted Data concerning transport of spent nuclear fuel.		
45 USC 12(d)	restricts the disclosure of personal identifiable material associated with Railroad Unemployment Insurance		
49 USC 40115	requires nondisclosure of information that would prejudice formulation and presentation of positions of United States in international negotiations or adversely affect the competitive position of any air carrier in foreign air transportation.		
50 USC 402 Notes	exempts the disclosure of the organization or function of the National Security Agency or any information with respect to the activities of the NSA, or of names, titles, salaries or numbers of persons employed by such agency.		
50 USC 403g	requires the protection of information about the internal structure of the CIA, including disclosures related to the organizations, functions, names, official titles, salaries, or numbers of personnel employed by the agency.		
50 USC 403-3	protects CIA "intelligence sources		

(c)(7)	and methods" from disclosure. Intelligence sources and methods are actually closed under a(1) b(1).		
50 USC App. 2411(c)(1)	governs the disclosure of export licenses and applications.		

There are 28 nondisclosure paragraphs listed in this table. In FY 2002, under FOIA exemption b(3), Federal agencies withheld information pursuant of 142 different nondisclosure statutes. So there are many more nondisclosures statutes than shown here. Depending on which statutory nondisclosure paragraphs tend to be applied in reviewing Presidential Records, we may expand or contract the list of those for which we create simulated records. We may be able to simulate many of these documents from the cases referred to in the *DOJ Freedom of Information Act Guide & Privacy Act Overview*.

a(4) b(4) "Confidential Commercial Information"

This restriction applies to two broad categories of information: (1) trade secrets, and (2) information which is (a) commercial or financial, (b) obtained from a person, and (c) privileged or confidential.

Reason for Withdrawal	Document Id(s)	Source or Simulated
Trade secrets and commercial or financial information contained in loans, bids, contracts, and/or proposals. ³⁰		
Statistical data concerning contract performance, income, profit, losses, and/or expenditures.		
Personal statements in inspections, investigations, or audits that reveal company trade secrets or proprietary information.		
Financial data in local wage surveys used to adjust pay schedules.		
Scientific and manufacturing data submitted with a research grant application or with a report while research is still in progress.		
Technical or scientific information developed at private expense.		
Personal Financial Information		
Cue phrases occurring in the document that may indicate a possible restriction: "privileged", "confidential", "trade secret," "confidential information," "proprietary."		

This restriction does not apply to the following type of information

Reason for Opening	Document ID(s)	Source or Simulated
Routine business information provided voluntarily by persons or corporations.		

³⁰ See also a(3)b(3) 41 USC 253(b)(m)

Information that corporations by law must file with the Securities and Exchange Commission that is publicly available, such as corporate registration statements, periodic reports to shareholders, tender offers, and ownership reports.		
---	--	--

PRA Restriction a(5) "Confidential Advice"

This restriction applies to "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 President's advisors include counselors and assistants to the President, Deputy Assistants, Special Assistants to the President, and the Director of Media Affairs. It could include a Senator or Congressman who writes to the President as a personal friend or trusted adviser, rather than in his or her official capacity. It could also include anyone in the Executive Branch providing advice, including interagency groups and committees generating options or advice. PRA restriction a(5) applies for twelve years after the expiration of the President's term in office.

The table below reflects those types of records that are subject to restriction a(5). Most examples of documents closed under a(5) are simulated.

Reason for Withdrawal	Document ID(s)	Source or Simulated
Presidential request for advice or information from an advisor	045 096 010 013	Simulated(Closed) Simulated(closed) All the best 491-492, All the Best, p490-491
Request from a Presidential advisor to another Advisor	095, 091 ³¹ , 097	Simulated(closed)
Response by a Presidential advisor to request from another Presidential Advisor	102	Simulated (closed)
Policy advice	100	Simulated(closed)
To President from Counsel's Office	101, 104, 109	Simulated(closed)
Advice from the Director of Media Affairs		
Advice from a Senator who is an advisor		
Advice from a State Representative who is an advisor		
Advice from a Personal friend who is advisor		
Advice from a member of the Cabinet		
Advice from an Interagency group		
Advice from a Committee		
Opinions about candidates for Appointment to Federal Office	098 ³²	Simulated(closed)

³¹ Also a(2).

³² Also a(2).

Facts inextricably intertwined with opinion and recommendation	093, 094	Simulated(closed)
Proposed guidance for answering press questions on foreign policy issues	099	Simulated(closed)

Examples of documents that contain confidential advice often contain administrative markings such as Administratively Confidential, Eyes Only, Limited Official use only, Personal and Confidential, Official Use Only, FOUO, Limit to Official Use Only, Limited Distribution, or Limited Official Use.

The following Table reflects types of records that would have been closed, but for the fact that President's Reagan or Bush waived their restriction rights under PRA a(5).

Waiver	Document Id(s)	Source or Simulated
Routine administrative requests for information, studies, or action, as well as information or studies submitted from a staff member or agency official to another staff member or agency official.		
Memorandums from staff members to other staff members that present proposals on routine administrative matters for approval or disapproval by a staff member.	032	Bush Library - Opened
Documents reflecting conversations of staff members with other staff members, agency officials, or members of the public concerning administrative or factual matters.	089, 090	³³
Presidential scheduling files including daily schedules, drafts, and proposals; routine background materials relating to schedules; movement logs, telephone logs; and other routine briefing materials.		
Speech-related materials (speech drafts, background, research, clearance process and speech cards).		
Draft press releases		
"Fact sheets," including drafts of and background for fact sheets.		
Memoranda and reports provided to the President and his staff which are purely informational or factual in content, including position papers, reports and studies that examine an issue.	040, 050, 033 025, 026	³⁴ ³⁵
Memoranda between staff members which are purely informational or factual in content	017	All the Best, p427
Requests and referrals from White House staff to other staff or agency officials for purely factual information and guidance in the preparation of		

³³ Bush Presidential Library, Quayle - Vice Presidential Records, VP Office, William Kristol's Chron Files, July 1990.

³⁴ Bush Presidential Library, Bush Presidential Records, White House Staff Member and Office Files, Office of Policy Development, Roger B Porters Files

³⁵ Bush Presidential Library, Bush Presidential Records, Council of Economic Advisors, Michael J. Boskin's Files

routine responses to inquiries and the information received.		
Memoranda and recommendations regarding routine congressional resolutions		
Routine memoranda analyzing and/or making recommendations regarding Administration positions on legislation or recommending Presidential approval or disapproval of legislation		
Routine White House Staffing Memorandums and comments		
White House Correspondence Tracking Worksheets and routine annotations		
Background materials, preparatory materials and drafts regarding routine Presidential proclamations and Special Messages.		
Background materials, preparatory materials and drafts regarding declarations, executive orders and other exercises of Presidential legal authority, prepared by the White House Executive Clerk's Office.		

Some Communications that May Contain Advice but that is not Confidential

Not all advice between the President and his advisors, and between advisors, is confidential.

Reason for Opening	Document ID(s)	Source or Simulated
Recommendations from Agency WRT legislation and correspondence with the Agency	058, 059, 060, 061	³⁶
Not confidential advice, e.g., "We think you should serve chicken." "I don't think we need to meet with them this week."		
Topics for meetings conducted by Think Tanks	063, 064, 065	³⁷
Fact Sheets about proposed legislation	062	³⁸
Correspondence from interest groups expressing concern about legislation and domestic and economic policy	084, 085, 086	³⁹
External Suggestions to help Small Business	080, 081	⁴⁰
Urging Disaster Relief	074, 075, 076, 077	⁴¹
Natural Disaster Situation Reports	078, 079	⁴²

³⁶ Bush Presidential Library, Bush Presidential Records, Staff Member and Office Files, Legislative Affairs, Richard Breeden's Files, S&L Rescue: FSLIC Suggestions

³⁷ Bush Presidential Library, Bush Presidential Records, Office of Policy Development, Charles E. M. Kolb's Files, AEI: American Enterprise Institute.

³⁸ Bush Presidential Library, Bush Presidential Records, Office of Policy Development, Charles E. M. Kolb's Files, AEI: American Enterprise Institute.

³⁹ Bush Presidential Library, Bush Presidential Records, Public Liaison, Doug Weed's Files, [Alpha Correspondence File 2/90-6/90 "A" "B"]

⁴⁰ Bush Presidential Library, Bush Presidential Records, WHORM Subject File, Business-Economics (BE), Small Business. ID#022987

⁴¹ Bush Presidential Library, Bush Presidential Records, WHORM Subject File, Disasters- Natural, ID#315058

External Facts and offer to express views	082, 083	⁴³
Humorous memoranda amongst staff or with President	012	All the Best, p479
Routine memoranda from President to Department and Agency Heads	036	⁴⁴
Letter from Senator who is not an advisor	023	Bush Library, opened

a(6) b(6) "Personal Privacy"

This restriction on release concerns records that contain "personal information" whose disclosure would "constitute a clearly unwarranted invasions of personal privacy." Fictitious documents are being constructed that contain the kinds of information that are subject to this restriction.

Reason for Withdrawal	Document ID(s)	Source or Simulated
Home addresses, home telephone numbers, and home e-mail addresses	042	⁴⁵
Birthdate	042	⁴⁶
Official duty email addresses and duty telephone numbers for personnel assigned to overseas, sensitive, and/or routinely deployable units		
Personnel Files: social security numbers, citizenship, religious affiliation, criminal history (criminal convictions)		
Marital status (e.g., separated, divorced), legitimacy of children	042	⁴⁷
Medical files: medical condition (e.g., AIDS, HIV, gonorrhea), mental condition (e.g., psychologist, psychiatrist)	105, 110	Simulated
Employment Application data (resumes, SF-171s)		
Security clearance records		
Personnel administrative or disciplinary actions		
Welfare payments		
Information about unsuccessful job applicants		
Information on the personal details of a Federal employee's service is also restricted, e.g., employees home address, reasons for leaving previous positions, contents of an FBI investigation concerning federal		

⁴² Bush Presidential Library, Bush Presidential Records, WKHORM Subject File, Disasters-Natural, ID#324869.

⁴³ Bush Presidential Library, Bush Presidential Records, WHORM Subject File, Business-Economics, Small Business, ID#022987

⁴⁴ Bush Presidential Library, Bush Presidential Records, Office of the President, President's Memoranda File.

⁴⁵ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#315353.

⁴⁶ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#315353.

⁴⁷ Bush Presidential Library, Bush Presidential Records, WHORM - Subject Files, Federal Government, Judicial Branch, Court of Appeals, ID#315353.

judicial appointment or appointment to the Executive Branch		
job performance evaluations or appraisals, e.g., Staff ratings such as 1, 2, 3 where 3 is negative, phrases such as "Sam isn't as good as ..."		
Salaries of non-Federal employees		
Information that is practically obscure (once public but currently hard to locate)		
Documents having terms describing personal situations or attitudes, e.g., Alcoholic, drunk, intoxicated, mental breakdown, nervous breakdown, idiot, hate, personal, affair		
External recommendation for appointment to federal office that is negative.		
Telephone number on a resume that is part of an appointment to Federal Office.		

Exemption b(2) "Internal Agency Rules and Practices"

There are two subcategories of records related "solely to the internal personnel rules and practices of an agency."

High 2: Substantial internal matters, the disclosure of which would risk circumvention of a statute or agency regulation, operating rules, and guidelines.

Low 2: internal matters of a relatively trivial nature.

Records in the High 2 subcategory must be withheld. The following table shows the types of records in that subcategory.

Reason for Withdrawal: High 2	Document ID(s)	Source or Simulated
Law enforcement manuals setting out guidelines for criminal investigations, such as when to pursue an investigation, prison security techniques, and hijacker profiles.		
Information that would reveal the nature and extent of a particular investigation.		
Agreements between state and federal agencies concerning when to exchange information relevant to potential tax law violations.		
Computer codes and sensitive computer programs used by law enforcement agencies.		
Hiring plans and job applicant testing materials.		
Examination questions and answers used for training, employment, and/or promotion.		
Vulnerability assessments, such as computer security programs, critical systems, facilities, stockpiles, security plans, and training manuals.		
Sensitive administrative notations, such confidential sources numbers.		
Identifying information relating to sources or undercover agents.		

Information regarding agency audit or examination guidelines, such as cut off criteria. Operating rules and manuals for auditors and or examiners.		
Security classification guidelines.		

The information in records in the Low 2 subcategory may be disclosed. The following tables lists examples of the types of records in that subcategory.

Reason for Opening: Low 2	Document ID(s)	Source or Simulated
parking facility rules		
lunch hour rules	127	⁴⁸
sick leave policy	126	⁴⁹
file numbers		
mail routing stamps		
Initials		
data processing notations.		

FOIA Exemption b(7) "Law Enforcement Investigations"

This category relates to "records or information compiled for law enforcement purposes, but only to the extent that the production of such law enforcement records or information:

- a) Could reasonably be expected to interfere with enforcement proceedings,
- b) Would deprive a person of a right to a fair trial or an impartial adjudication,
- c) Could reasonably be expected to constitute an unwarranted invasion of personal privacy,
- d) Could reasonably be expected to disclose the identity of a confidential source, including a State, local, or foreign agency or authority or any private institution which furnished information on a confidential basis, and, in the case of a record or information compiled by a criminal law enforcement authority in the course of a criminal investigation or by an agency conducting a lawful national security intelligence investigation, information furnished by a confidential source,
- e) Would disclose techniques and procedures for law enforcement investigations or prosecutions, or would disclose guidelines for law enforcement investigations or prosecutions if such disclosure could reasonably be expected to risk circumvention of the law, or
- f) Could reasonably be expected to endanger the life or physical safety of any individual."

If an archivist reviewing documents discovers information in one of the subcategories of this restriction, they must not only indicate the restriction, but the subcategory as well, e.g., b(7)C. Examples of records in each of these subcategories must be analyzed to determine the information that needs to be extracted, and to formulate rules that would enable inferring that a subcategory of the restriction applied.

⁴⁸ <http://www.opm.gov/oca/worksch/HTML/LUNCH.HTM>

⁴⁹ <http://www.opm.gov/oca/leave/html/fampres.htm>

Reason for Withdrawal	Document ID(s)	Source or Simulated
A.		
Material of a current or prospective (Federal, state or foreign) law enforcement investigation		
Material of a current or prospective (Federal, state or foreign) law enforcement prosecution		
Material of a current or prospective (Federal, state or foreign) administrative or regulatory proceedings (of a law enforcement nature)		
Material of a current or prospective Independent Counsel investigation.		
B.		
information the disclosure of which would deprive an individual of the right to a fair trial or impartial adjudication.		
C.		
law enforcement records--release of which could reasonably be expected to cause an unwarranted invasion of personal privacy, e.g., mention of individual's name, mention of Federal, state or local law enforcement personnel's names, names of individuals who provide information to law enforcement agencies., or the names of witnesses, their home and business addresses, and their telephone numbers.		
Document was opened or opened in part because it was in the public interest, i.e., "shed light on an agency's performance of its statutory duties		
D.		
Materials containing the identities of confidential sources and in the case of a criminal investigation conducted by a criminal law enforcement authority or a lawful national security intelligence investigation, all information provided by those sources. The exemption applies to intelligence sources, criminal sources, state, local and foreign agencies and to private institutions (e.g., credit bureaus) that furnished information on a confidential basis. Sources include citizens providing unsolicited allegations of misconduct.		
Document containing Intelligence sources of the US Secret Service		
E.		
Methods of law enforcement such as stakeouts		
The FBI record regarding the relocation of witnesses under the Witness Relocation Program, because disclosure would jeopardize the effectiveness of the program		
Leads, spares (tape backups of computerized data, spare patrol cars left in reserve, or an Emergency Operations Center (EOC) in another location that is used only if a power failure or natural disaster renders the primary communications facility inoperable)		
Movements in diagrams or schedules		
Methods of the US Secret Service (USS).		

Documents containing methods of agents of the US Secret Service		
F.		
Document containing names of FBI agents, or other law enforcement personnel, as well as information that might lead to their identities, where their safety would be endangered by disclosure.		
Names of agents of the US Secret Service		

Presidential Records that have No Restrictions

There are Presidential Records that have no access restrictions, for example, White House press releases, newspaper articles, press feeds, speeches, and transcriptions of press conferences. They have no access restrictions because they have already been published or otherwise released to the public.

Reason for Opening	Document Id(s)	Source
Press Release: Transcript of Presidential Speech or remarks	123, 124, 125	Bush Public Papers
Press Release: Presidential announcement of nomination or appointment	051	⁵⁰
Press Release: Presidential Executive Order ⁵¹	119, 120, 121	⁵²
Press Release: Presidential Bill signing		
Press Release: Fact Sheet		
Press Release: Presidential Proclamation	052	⁵³
Press Release: Statement by the Press Secretary	003	⁵⁴
Press Pool Report	038, 039, 041	Bush Library
Schedules	066, 067	⁵⁵
Transcript of President's News Conference	113, 114, 115	Bush Public Papers
Copy of Newspaper Article (or News wire)		
Daily Diary including Appendices (names of Intelligence Personnel or Secret Service may be redacted)	018, 019, 020, 021, 022	⁵⁶
Letters of Appreciation to the President for attending/speaking at a meeting	068, 069	⁵⁷
Suggestions for a Presidential Campaign Visit and Response	070, 071	⁵⁸

⁵⁰ Bush Presidential Library, Bush Presidential Records, White House Press Office, Press Office Files - Daily Press Releases 1989-1993.

⁵¹ Executive orders are filed in the Federal Register by the President and after 30 days if there is no objection from Congress become codified law, as valid as a statute, passed by Congress.

⁵² Bush Public Papers, <http://bushlibrary.tamu.edu/research/paper.html>

⁵³ Bush Presidential Library, Bush Presidential Records, White House Press Office, Press Office Files - Daily Press Releases 1989-1993.

⁵⁴ Bush Presidential Library, Bush Presidential Records, White House Press Office, Press Office Files - Daily Press Releases 1989-1993.

⁵⁵ Bush-Office of Vice President, Advance, John Keller's Files, Atlanta, Georgia Debate, February, 26-28.

⁵⁶ Bush Presidential Library, The Daily Diary of President George Bush, May 11, 1989

⁵⁷ Bush Presidential Library, Bush Presidential Records, WHORM Subject File, Education, ID#287437

Factual Knowledge

To distinguish personal/political records from Presidential records, lists of the Bush Family members, personal friends, campaign staff, and RNC staff will be needed. To determine whether a document concerns an appointment to Federal office that is subject to Senate approval, or an appointment that is not subject to Senate approval, we need a list of such offices. To determine whether PRA restriction a(5) might apply to a Presidential record, we will need a list of Presidential advisors,

In summary, the Access Restriction Checker will require the following factual knowledge.

- Names of George H. W. Bush Family Members
- President's Friends
- Campaign Staff
- RNC Staff Register
- Federal Offices to which the President appoints or nominates individuals that are subject to Senate approval, and those that are not.⁵⁹
- Presidential Advisors
- Names of members of Cabinet and other Bush Administration Senior Officials
- White House Staff Members, their titles and office names⁶⁰
- Members of 101st (1989-90) and 102nd 1991-92) Congresses
- Legislation of the 101st and 102nd Congress⁶¹
- Names of Foreign Heads of State, and their titles (1989-1992)⁶²

File Formats

One of this challenges of this research is dealing with legacy file formats. The Bush Administrations personal computer files were created using a variety of legacy word processors, for example,

- AMI Pro (SAMNA)
- DCA-RFT
- IBM DisplayWrite 2&3
- IBM DisplayWrite 4&5
- Lotus Manuscript
- MSWord for DOS
- MSWord for Windows 2.0
- MultiMate Advantage 2

⁵⁸ Bush Presidential Library, Bush Presidential Records, WHORM Subject File, Education, ID#287437

⁵⁹ *Federal Yellow Book, Judicial Yellow Book*, Leadership Directories, Inc.

⁶⁰ White House Telephone Directory

⁶¹ THOMAS <http://thomas.loc.gov/bss/d101query.html>

⁶² 1990, 1991, 1992, 1993 CIA World Fact Books <http://manybooks.net/titles/usciaetext93world192.html>

- Windows Write
- WordPerfect 4.2
- WordPerfect 5.0
- WordPerfect 5.1/5.2
- WordPerfect Notebook

Some of the documents in the corpus are created in these file formats so that we can test the capability of the Stellent Filters in Oracle Text 9.2 [4] to convert documents in these formats to ASCII text or HTML, which is a format needed by the information extraction routines.

Furthermore there are other types of legacy digital objects than text documents that occur in the Bush Administration's personal computer files. For example

- 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
- QuatroPro for DOS Worksheet
- QuatroPro for Windows 3.x Workbook
- WordPerfect Planner
- Harvard Graphics 2.0 Chart
- Harvard Graphics 3.0 Chart

These objects can contain textual information. Hence, we will attempt to create some of these object types and test the capability to convert to ASCII text.

Offices, Record Series, and Folders

The Bush Administration's PC Files are aggregated in directories (or folders) from a particular office and are usually created, received or used by a single person. At the time an archivist reviews these records, they have described the aggregation of files by the Office Name, the record series, and the folder titles, for example

- White House Press Office
- Press Office Files
- Daily Press Releases

Similarly, the documents in the "Test Corpus" are grouped into folders (Daily Press Releases) that are a part of a record series (Press Office Files) created in an Office (White

House Press Office). This provides contextual information that aids in determining whether documents are personal records or presidential records, and which type of restrictions might or might not apply.

This aggregation of documents into directories is also needed to test the directory/folder titler and filing system summarizer.

Summary

We have collected or created about 125 sample records. This includes personal/political records that are available to the public or are simulated. We have also created fictitious records that simulate documents with access restrictions. Finally, we have collected records from the Bush Presidential Library that have no access restrictions, and so have been opened for public access. These paper records have been scanned and converted to machine-readable form.

Fifty documents from the corpus have already been used in an experiment to determine the initial performance of our information extraction tools. After improvement of the knowledge base of these tools, an additional fifty documents from the corpus are being used to test the performance after the improvements [5]. The corpus will also be used to test the performance of a software agent that learns document types and then to recognize document types. Additional examples of document types will be needed for machine learning of document type, and then testing the document type identifier.

The corpus will also be used in an attempt to automatically create better folder titles. It will also be used in experiments to summarize the contents of a file system, that is, to generate the content notes for a record series.

The "Test Corpus" will also be used to formulate the knowledge needed by a prototype Access Restriction Checker to distinguish personal/political records from Presidential Records and to determine whether documents contain information that is exempt from release under FOIA, or which is restricted under a provision of the Presidential Records Act.

We will initially analyze records in the "Test Corpus" that have one of the following three types of access restriction.

- a(2) appointments to Federal offices
- a(5) confidential advice
- a(6) b(6) personal privacy

This analysis will identify the kinds of information that need to be extracted from documents and result in the generalization of rules that characterize the conditions under which one of these restrictions might apply to an entire document or to terms or passages in a document.

After testing of the information extraction tools, summarizer, the folder titler, document type identifier, and "Access Restriction Checker" using the "Test Corpus," experiments will be conducted at the Bush Presidential Library with actual Presidential and Personal Records from the Bush PC Files.

Another benefit of creating this collection of "simulated" Presidential Records is that we will have a collection of records that can be used with the Archival Processing Tool (APT) for training archivists in the use of the tool for arrangement, preservation, review, and description of digital computer records.

An Access database table has been constructed for describing the electronic documents in the corpus. The table below shows the design.

Attribute	Value	Example
Document ID	3 char	
Record_Type ⁶³	12 char	
P1_F1 ⁶⁴	Yes/No	
P2	Yes/No	
P3_F3	Yes/No	
P4_F4	Yes/No	
P5	Yes/No	
P6_F6	Yes/No	
F2	Yes/No	
F7A	Yes/No	
F7B	Yes/No	
F7C	Yes/No	
F7D	Yes/No	
F7E	Yes/No	
F7F	Yes/No	
Reason_for_Withdrawal_or_Opening	255 char	
Document Type	16 char	
Office	50 char	
Series_Title	50 char	
Folder_Title (possible value is null)	50 char	
File Format	20 char	
Source	255 char ⁶⁵	
Url	50 char	

⁶³ Presidential, Federal, Personal or Personal/Political

⁶⁴ P1 is the same as a(1). F1 is the same as b(1). If the record is Personal (Misfiled), then none of the restrictions will have a value. Mark the restrictions with an X. There may be multiple restrictions for a document. We included P1_F1, i.e., a(1)b(1) even though we said we would not look at national security records, because we have such good examples of communications between heads of state that have been declassified.

⁶⁵ One of the possible values of Source is "Simulated." Actually all of the records are simulated because all will have been keyed from paper or web documents. However, we use the term "Simulated" in the cases that we have analyzed a Bush Library document that has one or more restrictions on release to the public and we have replaced the date, author addressee, and restricted information with fictitious information. This is done so that we don't provide any information that could lead to inferring the actual content of the record. On the other hand, if we key in a record from the Bush Library that has been opened, we will site the Bush Presidential Library, the collection name, the Series Title and possibly the Folder Title.

References

1. W. Underwood, M. Hayslett-Keck and S. Laib. The Archival Processing Tool: User's Guide, Version 2.06. PERPOS Technical Report ITTL/CSITD 03-2, Information Technology and Telecommunications Laboratory, Georgia Tech Research Institute, Revised December, 2003
2. M. Hayslett-Keck and W. Underwood. An Analysis of the Knowledge Required to Perform FOIA and PRA Review. PERPOS Technical Report ITTL/CSITD 04-1, January 2004, Revised March 2004. Georgia Tech Research Institute, Atlanta Georgia.
3. G. Bush. *All the Best: My Life in Letters and Other Writings*. Scribner, New York, 1999.
4. Oracle Text Reference Release 9.2
5. M. G. Underwood. Evaluation of Named-Entity Recognition Technologies Applied to Presidential Electronic Records. PERPOS Technical Report 04-4, Information Technology and Telecommunications Laboratory, Georgia Tech Research Institute, June 2004.