Far More Than You Ever Wanted To Tell

Hidden Data In Document Formats

2004-09-29

Summerschool Applied IT Security 2004

Maximillian Dornseif

See http://md.hudora.de/presentations/summerschool/

Laboratory for Dependable Distributed Systems

RHEINISCH-
WESTFÄLISCHE
TECHNISCHE
HOCHSCHULE
AACHEN

RWTH

UNIVERSITY
OF TECHNOLOGY
AACHEN

The Problem

- Complex Dataformats
  - We are not supposed to understand
  - or we are not willing to understand
- Covert channels everywhere!

Tools of the Trade

Word Converters

- Antiword
  - Word 2, 6, 7, 97, 2000 and 2002
  - http://www.winfield.demon.nl/
- catdoc & xls2csv
  - no support for OLE streams
  - http://www.45.free.net/~vitus/ice/catdoc/
- word2x
  - http://word2x.sourceforge.net/

Laola

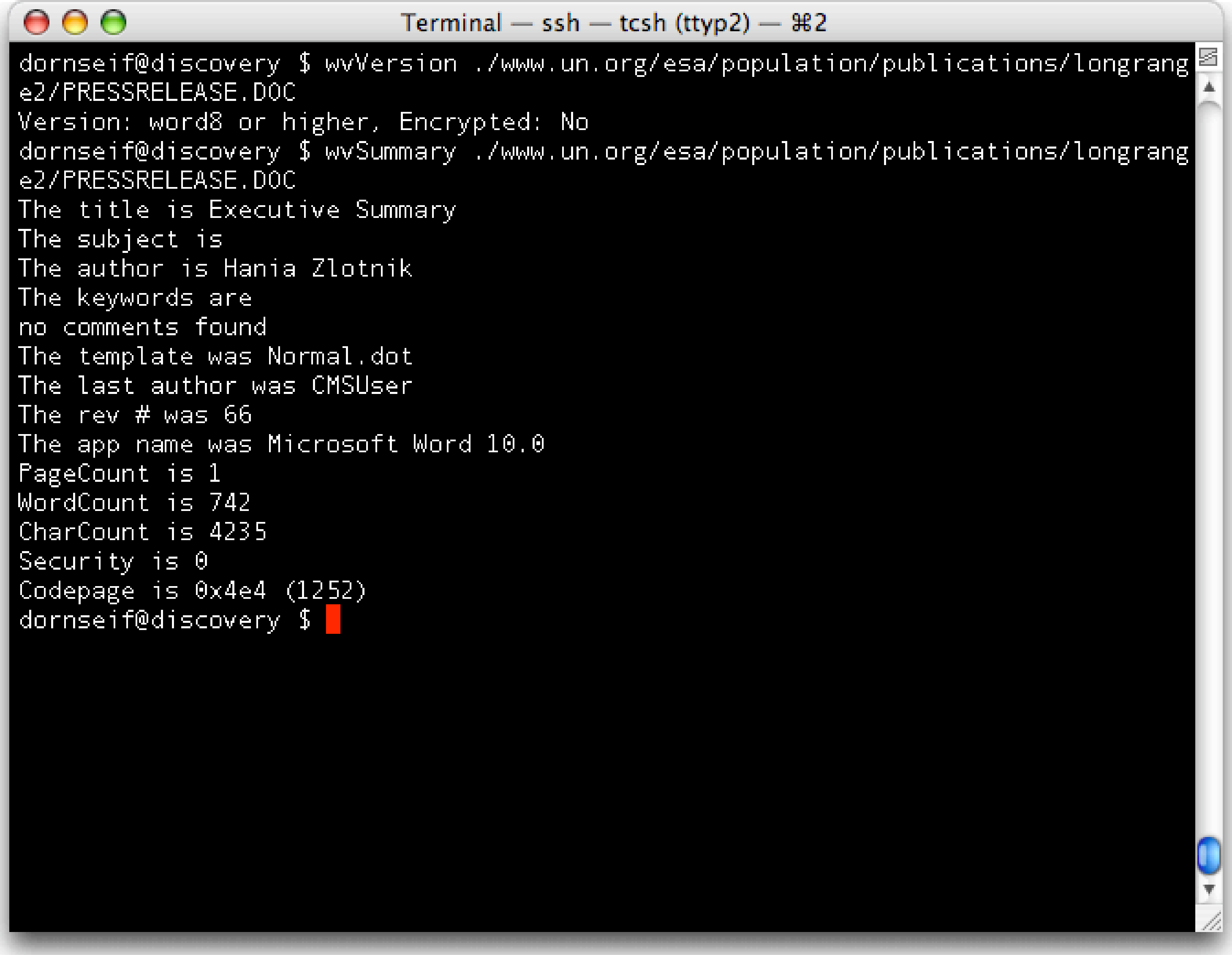
- Laola "is a collection of documentations and perl programs dealing with binary file formats of Windows program documents."
- Contains
  - **lclean** - Laola Clean: "Saves the trash sections of e.g. Word 6, Word 7 or Excel documents to own files."
  - **ldat** -Laola Display Authress Title: "Lists author, title, creation date and some other information sticked in a laola file. Gets printer information from Excel and Word files."
  - **lls** - Laola List: "Lists the structure of a Laola document."
  - **Elser** - "password resolving, macro decoding".
- Development ceased for 5 years.
- http://www.cs.tu-berlin.de/~schwartz/pmh/index.html

wvWare

- used by abiword
- tested by kword
- actively developed, but development lines are hard to understand: WordView, wv, wv2, wvWare ...
- Tools
  - wvText, wvHtml
  - wvSummary, wvVersion

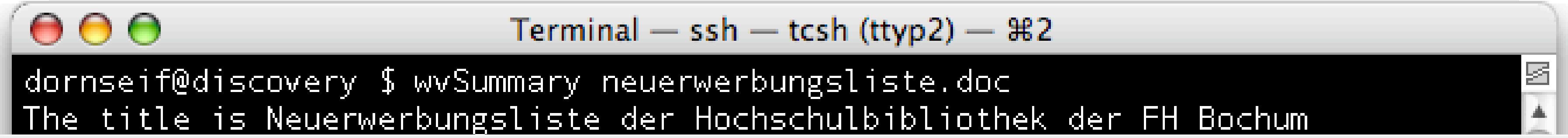
http://wvware.sourceforge.net/

wvVersion, wvSummary

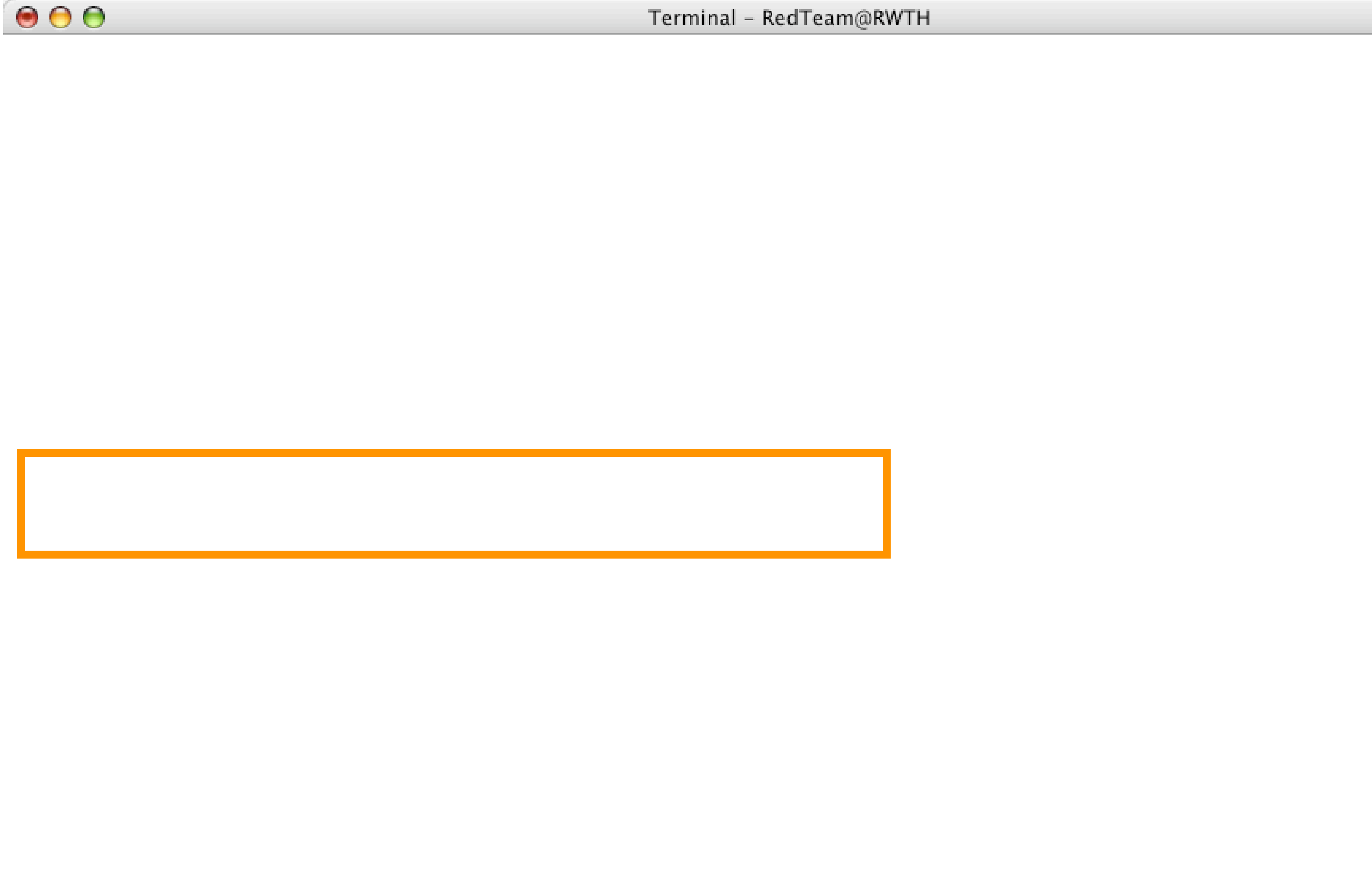


```
dornseif@discovery $ wvVersion ./www.un.org/esa/population/publications/longrang
e2/PRESSRELEASE.DOC
Version: word8 or higher, Encrypted: No
dornseif@discovery $ wvSummary ./www.un.org/esa/population/publications/longrang
e2/PRESSRELEASE.DOC
The title is Executive Summary
The subject is
The author is Hania Zlotnik
The keywords are
no comments found
The template was Normal.dot
The last author was CMSUser
The rev # was 66
The app name was Microsoft Word 10.0
PageCount is 1
WordCount is 742
CharCount is 4235
Security is 0
Codepage is 0x4e4 (1252)
dornseif@discovery $
```

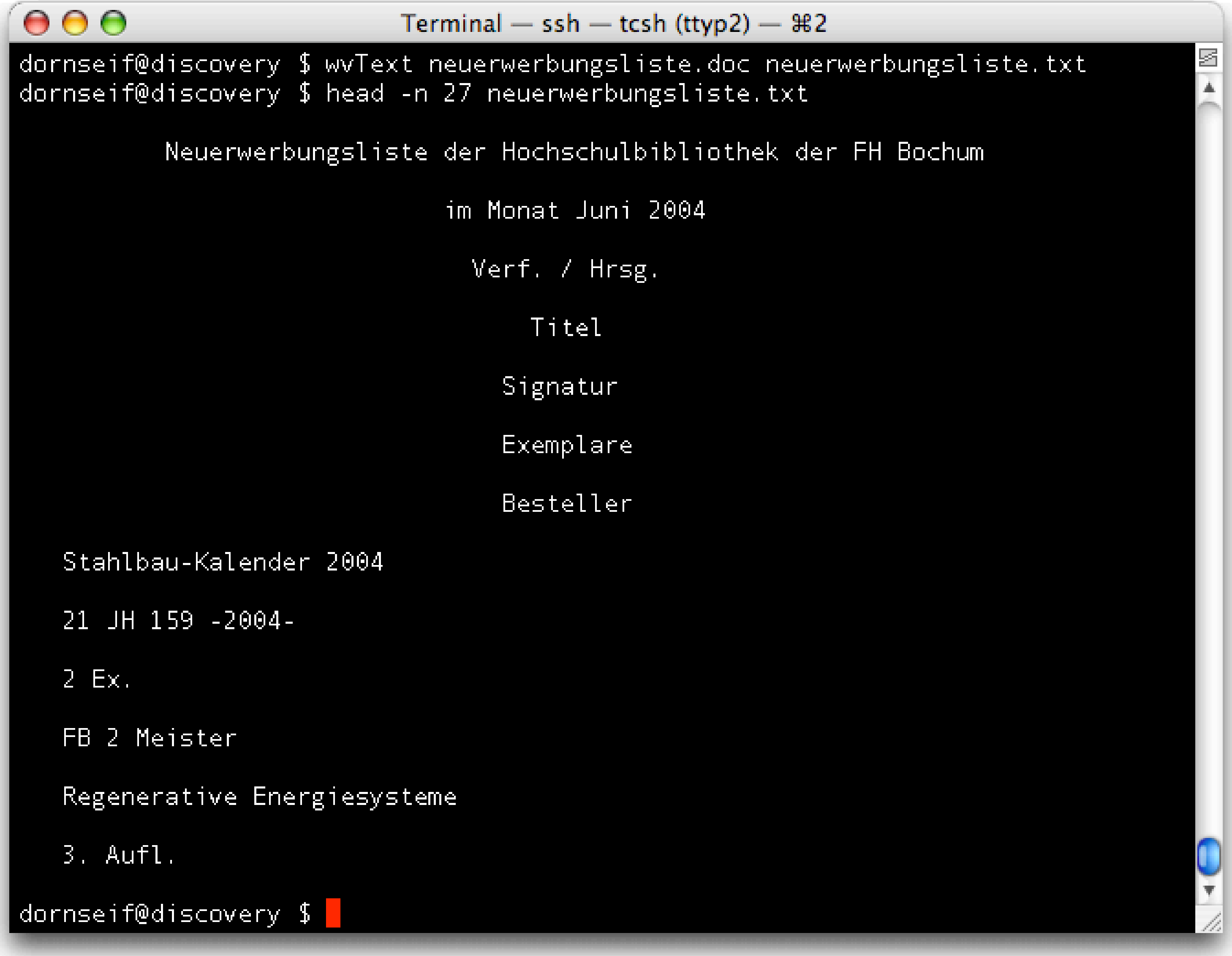
wvToolkit



```
dornseif@discovery $ wvSummary neuerwerbungsliste.doc
The title is Neuerwerbungsliste der Hochschulbibliothek der FH Bochum
```

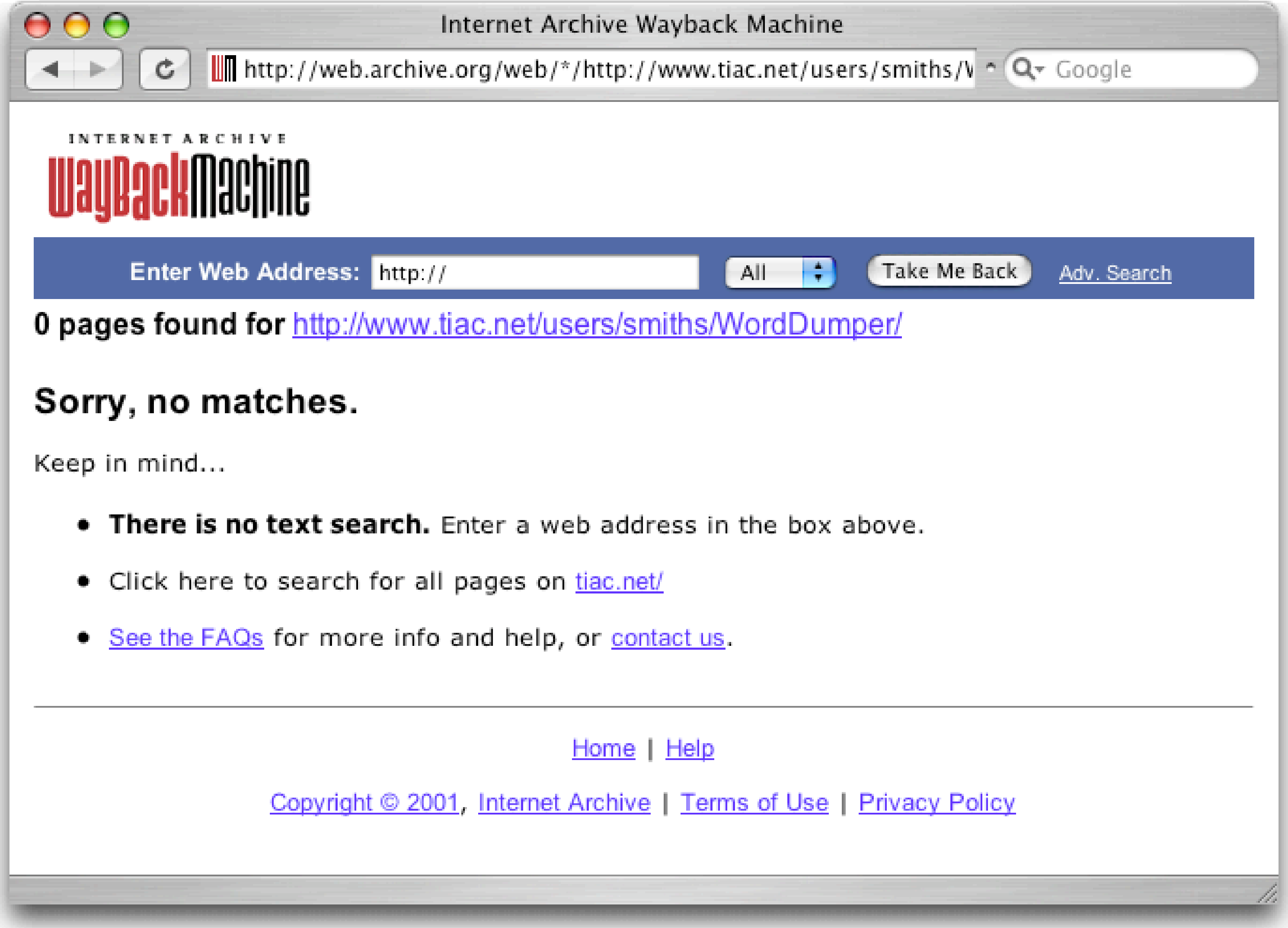


wvText



```
dornseif@discovery $ wvText neuerwerbungsliste.doc neuerwerbungsliste.txt
dornseif@discovery $ head -n 27 neuerwerbungsliste.txt

          Neuerwerbungsliste der Hochschulbibliothek der FH Bochum

                             im Monat Juni 2004

                               Verf. / Hrsg.

                                   Titel

                                 Signatur

                                 Exemplare

                                 Besteller

   Stahlbau-Kalender 2004

   21 JH 159 -2004-

   2 Ex.

   FB 2 Meister

   Regenerative Energiesysteme

   3. Aufl.

dornseif@discovery $ 
```

WordDumper



Internet Archive Wayback Machine

http://web.archive.org/web/*/http://www.tiac.net/users/smiths/W

Enter Web Address: http://

0 pages found for http://www.tiac.net/users/smiths/WordDumper/

Sorry, no matches.

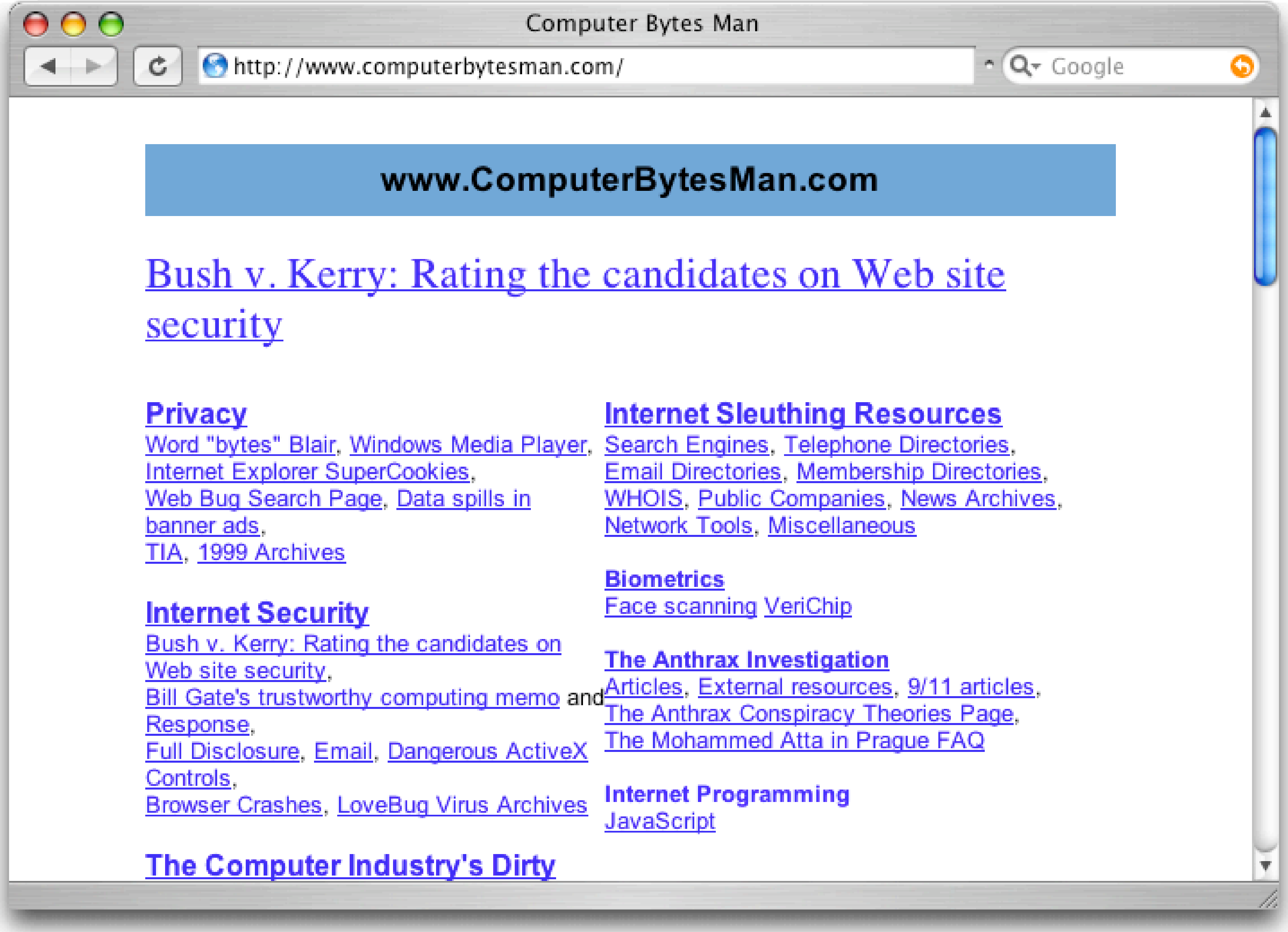
Keep in mind...

- **There is no text search.** Enter a web address in the box above.
- Click here to search for all pages on tiac.net/
- See the FAQs for more info and help, or contact us.

Home | Help

Copyright © 2001, Internet Archive | Terms of Use | Privacy Policy

WordDumper



www.ComputerBytesMan.com

Bush v. Kerry: Rating the candidates on Web site security

Privacy

Word "bytes" Blair, Windows Media Player, Internet Explorer SuperCookies, Web Bug Search Page, Data spills in banner ads, TIA, 1999 Archives

Internet Security

Bush v. Kerry: Rating the candidates on Web site security, Bill Gate's trustworthy computing memo and Response, Full Disclosure, Email, Dangerous ActiveX Controls, Browser Crashes, LoveBug Virus Archives

The Computer Industry's Dirty

Internet Sleuthing Resources

Search Engines, Telephone Directories, Email Directories, Membership Directories, WHOIS, Public Companies, News Archives, Network Tools, Miscellaneous

Biometrics

Face scanning VeriChip

The Anthrax Investigation

Articles, External resources, 9/11 articles, The Anthrax Conspiracy Theories Page, The Mohammed Atta in Prague FAQ

Internet Programming

JavaScript

Examples

Laboratory for Dependable Distributed Systems

RHEINISCH-
WESTFÄLISCHE
TECHNISCHE
HOCHSCHULE
AACHEN

RWTH

UNIVERSITY
OF TECHNOLOGY
AACHEN

Mail- & News-Headers

- RfC 822 and friends are well known in the techie community but a mystery to everybody else.
- Data in there possibly include: OS, IP, server, software and their versions, organisation, time, customer number at isp / telephone number (!), etc.

Incidents

- T-Online

Config Files

- Config files which are not well understood are a security issue...
- ... but also can result in disclosure of information which is not to be disclosed

Incidents

- Apache
- BitchX

HTML

- Complex programs generate complex HTML
- Most obvious:
  - META generator
  - Paths to local files

Incidents

- <META Creator>
- <!-- comments -->
- Defaced web pages (attrition.org)

```
<img src=”c:\...\Jon Doe\My Documents\coolpix.jpg”>
```

PDF

- Looks like an “open standard” ...
- ... but very hard to decode in depth
- The Problem of [redacted] / redaction.

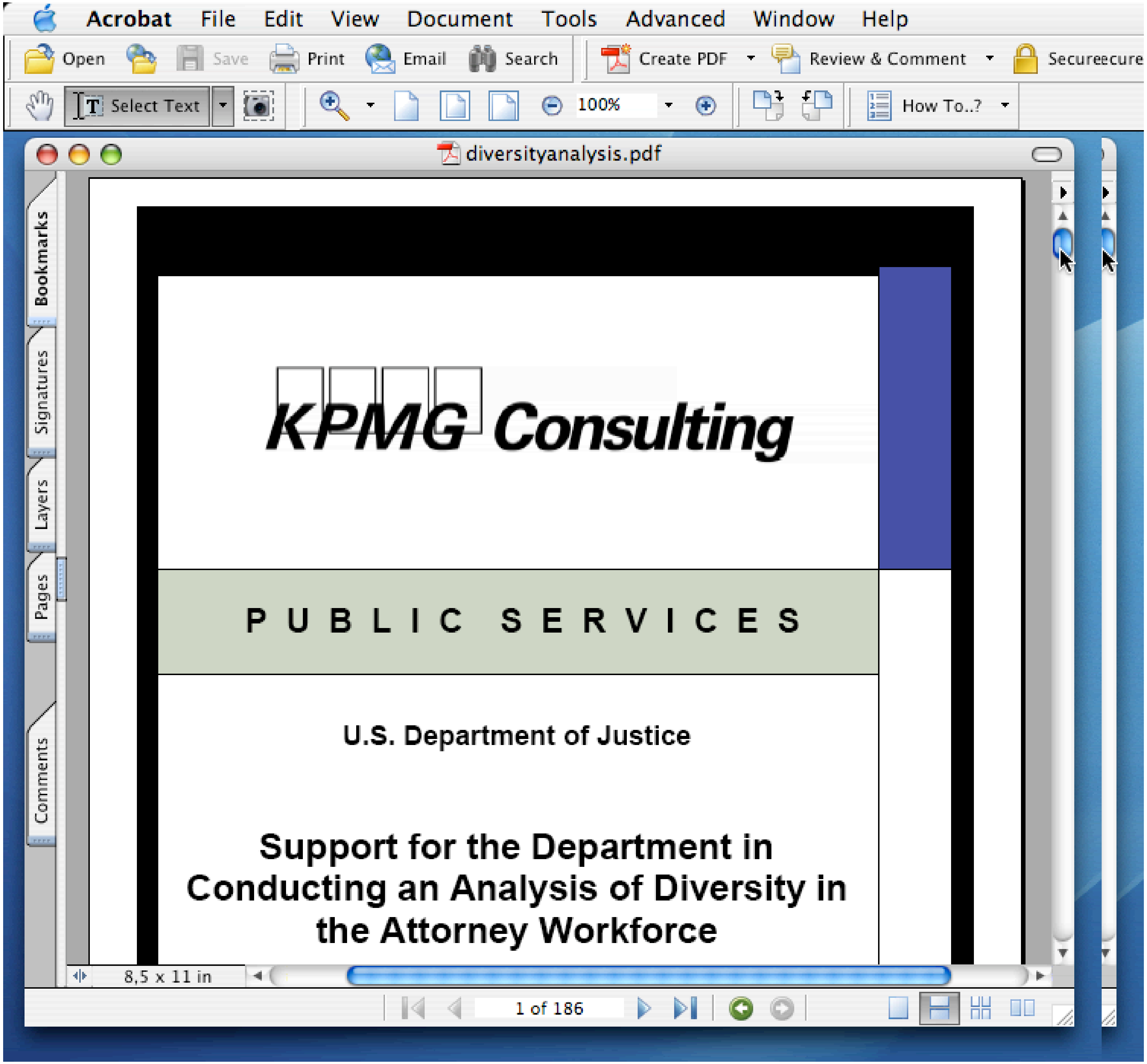
Incidents

- Sniper Letter
- The Justice Dept's Attorney Workforce Diversity Study
- "Secrets of History: The C.I.A. in Iran"

Exploiting hidden data

- Methods
  - Copy black text on black ground
  - Copy underlying graphics
  - Remove overlaying graphics
- Very dependant on the amount of Adobe software you have at hand.

black text on black ground

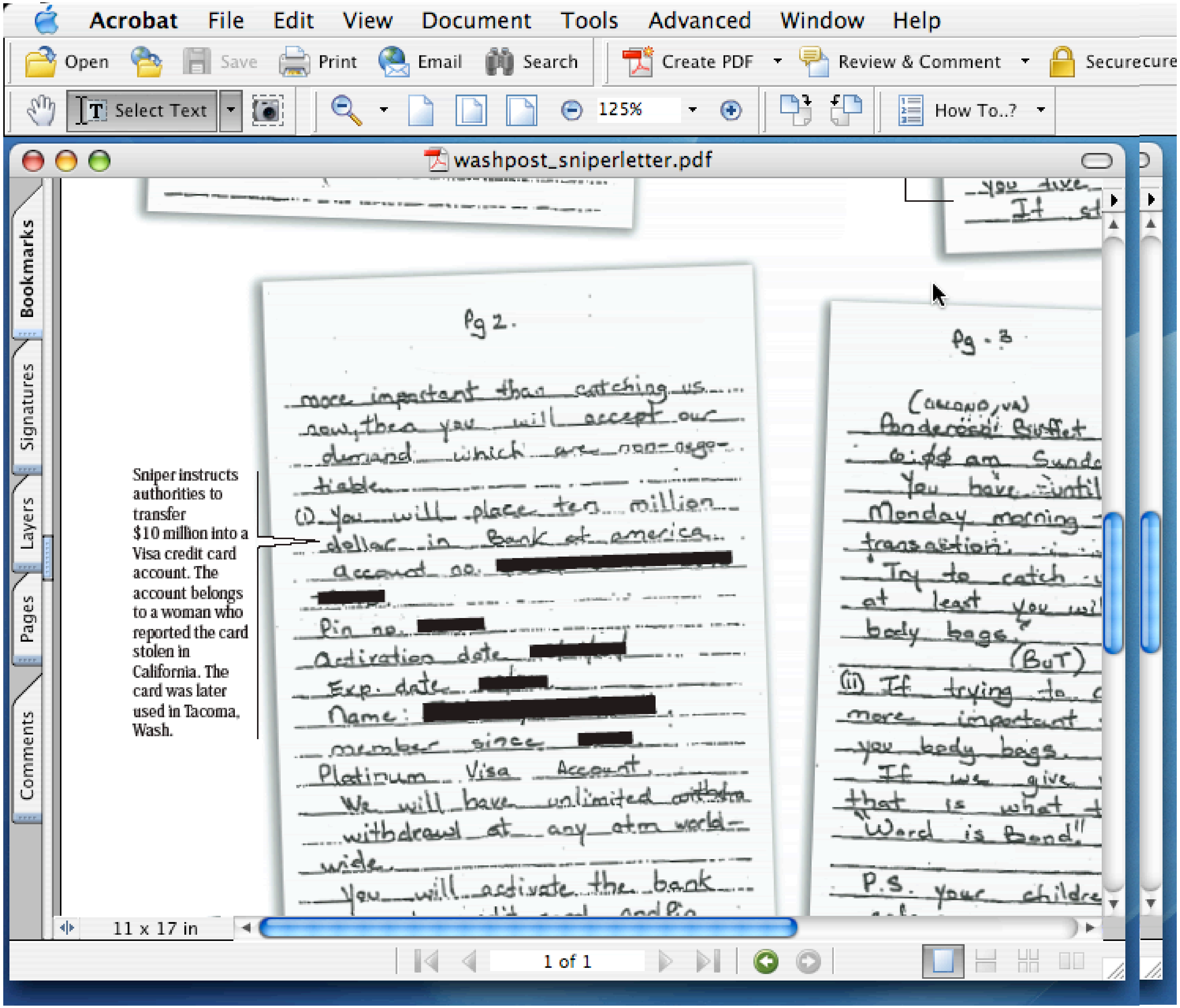


KPMG Consulting

PUBLIC SERVICES

U.S. Department of Justice

Support for the Department in Conducting an Analysis of Diversity in the Attorney Workforce copy underlying graphics



Sniper instructs authorities to transfer $10 million into a Visa credit card account. The account belongs to a woman who reported the card stolen in California. The card was later used in Tacoma, Wash.

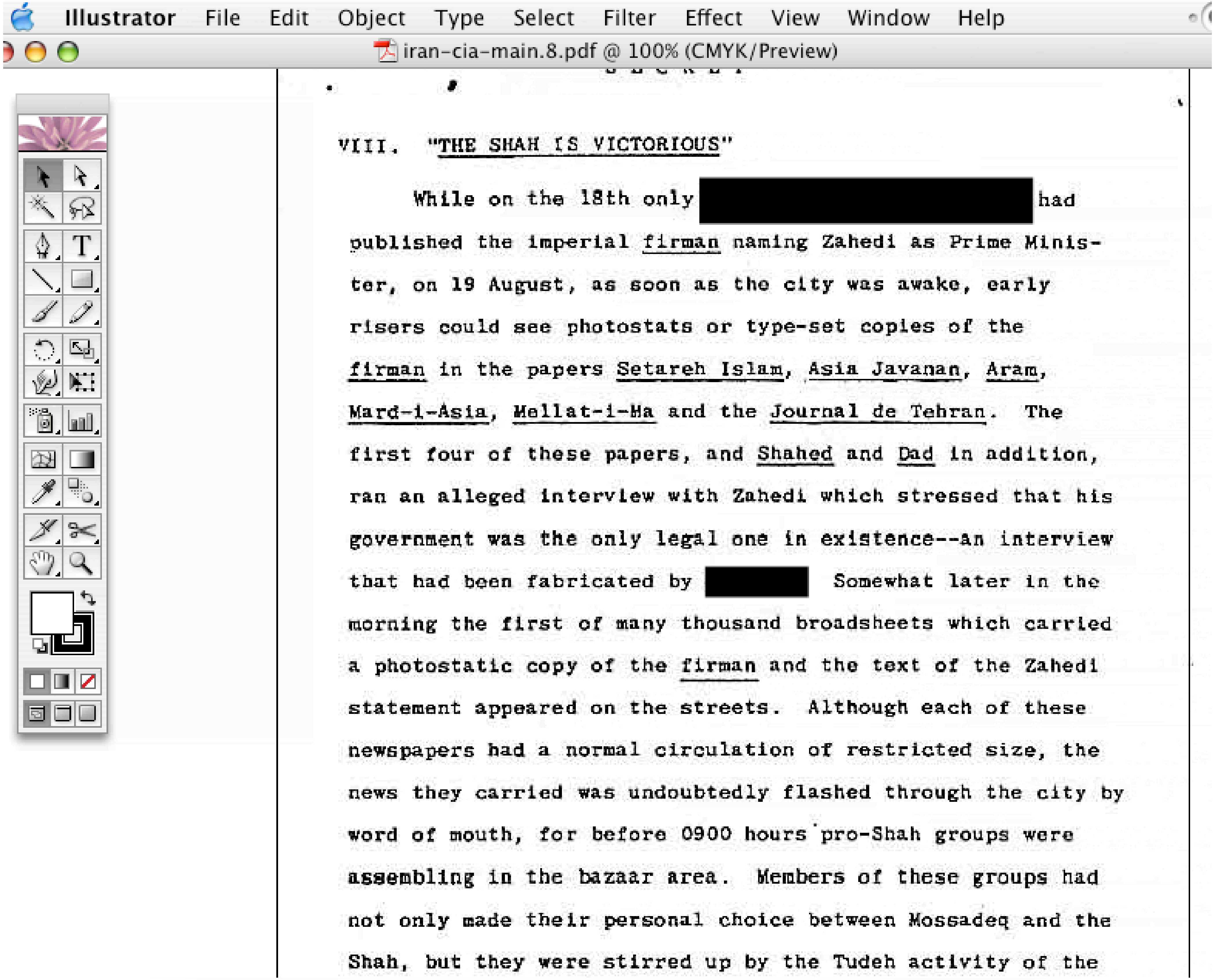
Pg 2.

more important than catching us.
now, then you will accept our
demand which are non-nego-
tiable.
(i) You will place ten million
dollar in Bank of america.
account no. [redacted]
Pin no. [redacted]
Activation date [redacted]
Exp. date [redacted]
Name: [redacted]
member since [redacted]
Platinum Visa Account.
We will have unlimited ~~withdra~~
withdrawl at any atm world-
wide.
You will activate the bank

Pg. 3

(Ashland, VA)
Ponderosa Buffet
6:00 am Sunda
You have until
Monday morning
transaction.
Try to catch
at least you wil
body bags.
(BuT)
(ii) If trying to c
more important
you body bags.
If we give
that is what
"Word is Bond!"
P.S. your childre

remove overlay graphics



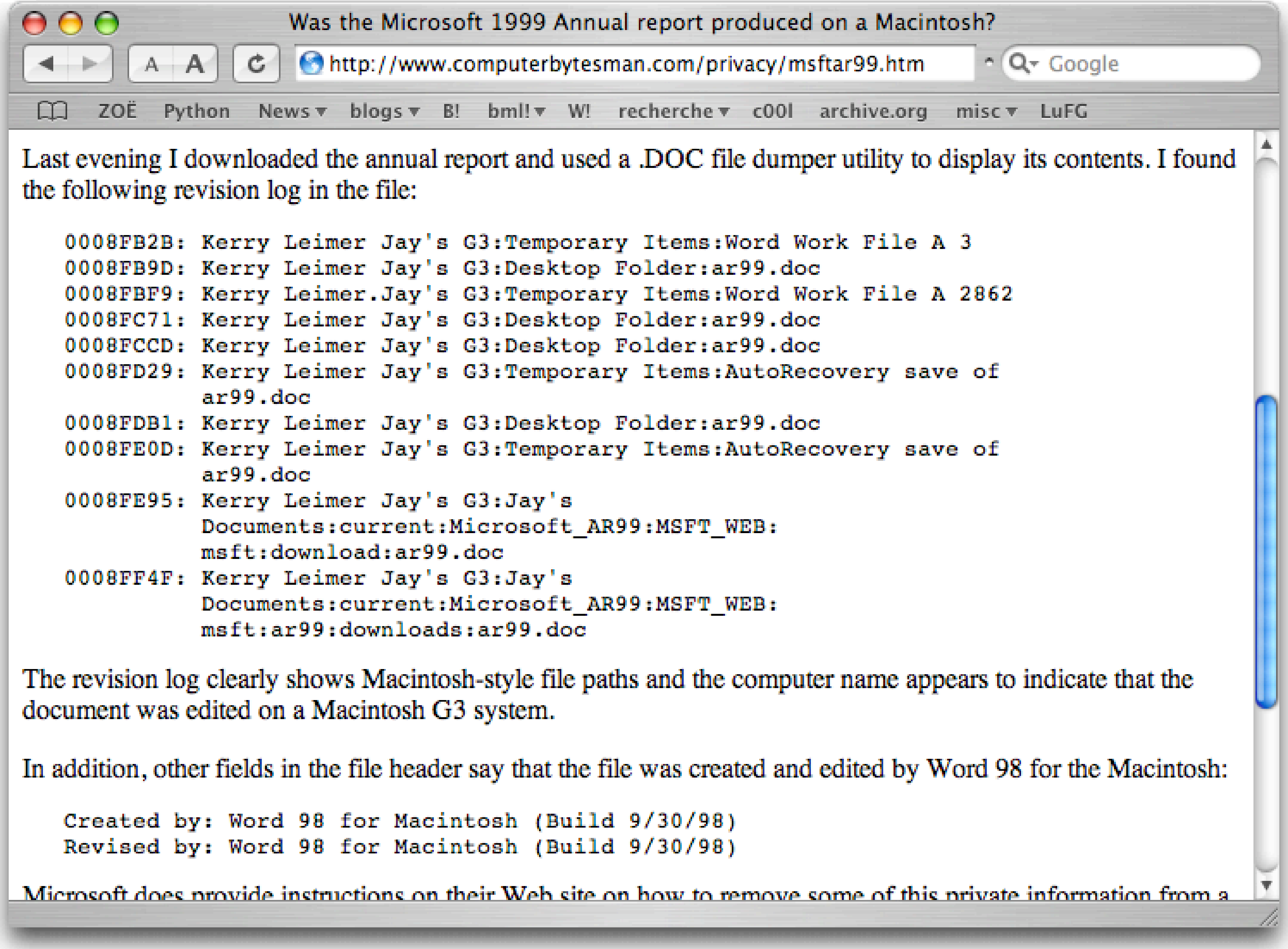
VIII. "THE SHAH IS VICTORIOUS"

While on the 18th only [redacted] had published the imperial firman naming Zahedi as Prime Minister, on 19 August, as soon as the city was awake, early risers could see photostats or type-set copies of the firman in the papers Setareh Islam, Asia Javanan, Aram, Mard-i-Asia, Mellat-i-Ma and the Journal de Tehran. The first four of these papers, and Shahed and Dad in addition, ran an alleged interview with Zahedi which stressed that his government was the only legal one in existence--an interview that had been fabricated by [redacted] Somewhat later in the morning the first of many thousand broadsheets which carried a photostatic copy of the firman and the text of the Zahedi statement appeared on the streets. Although each of these newspapers had a normal circulation of restricted size, the news they carried was undoubtedly flashed through the city by word of mouth, for before 0900 hours pro-Shah groups were assembling in the bazaar area. Members of these groups had not only made their personal choice between Mossadeq and the Shah, but they were stirred up by the Tudeh activity of the

MS Office

The MS Office document format is incredibly complex,
undocumented and ever changing.
It is well known that’s full of unwanted data.

History


Was the Microsoft 1999 Annual report produced on a Macintosh?
http://www.computerbytesman.com/privacy/msftar99.htm
Google
ZOË
Python
News
blogs
B!
bml!
W!
recherche
c00l
archive.org
misc
LuFG

Last evening I downloaded the annual report and used a .DOC file dumper utility to display its contents. I found the following revision log in the file:

```
    0008FB2B: Kerry Leimer Jay's G3:Temporary Items:Word Work File A 3
    0008FB9D: Kerry Leimer Jay's G3:Desktop Folder:ar99.doc
    0008FBF9: Kerry Leimer.Jay's G3:Temporary Items:Word Work File A 2862
    0008FC71: Kerry Leimer Jay's G3:Desktop Folder:ar99.doc
    0008FCCD: Kerry Leimer Jay's G3:Desktop Folder:ar99.doc
    0008FD29: Kerry Leimer Jay's G3:Temporary Items:AutoRecovery save of
              ar99.doc
    0008FDB1: Kerry Leimer Jay's G3:Desktop Folder:ar99.doc
    0008FE0D: Kerry Leimer Jay's G3:Temporary Items:AutoRecovery save of
              ar99.doc
    0008FE95: Kerry Leimer Jay's G3:Jay's
              Documents:current:Microsoft_AR99:MSFT_WEB:
              msft:download:ar99.doc
    0008FF4F: Kerry Leimer Jay's G3:Jay's
              Documents:current:Microsoft_AR99:MSFT_WEB:
              msft:ar99:downloads:ar99.doc
```

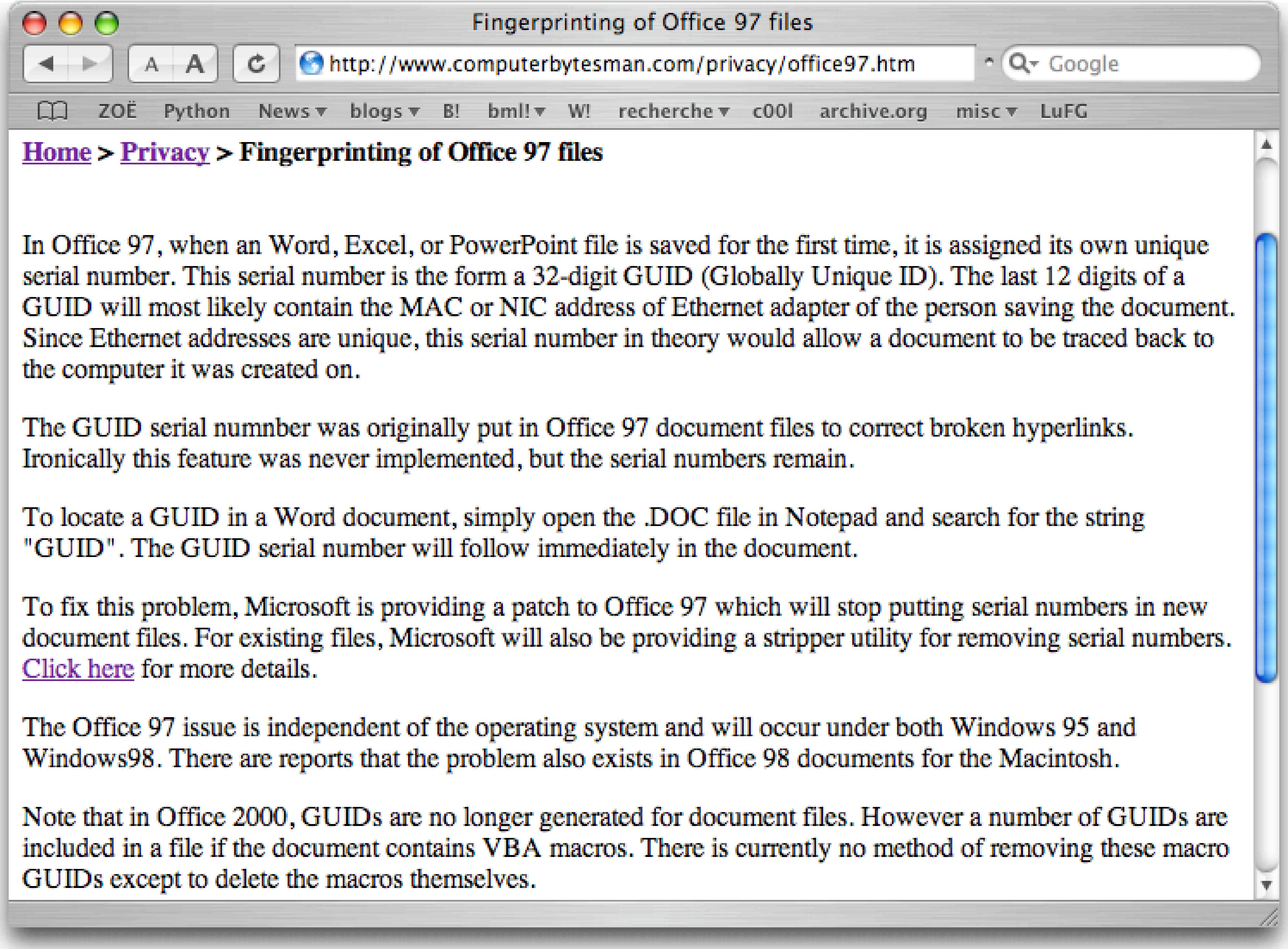
The revision log clearly shows Macintosh-style file paths and the computer name appears to indicate that the document was edited on a Macintosh G3 system.

In addition, other fields in the file header say that the file was created and edited by Word 98 for the Macintosh:

```
    Created by: Word 98 for Macintosh (Build 9/30/98)
    Revised by: Word 98 for Macintosh (Build 9/30/98)
```

Microsoft does provide instructions on their Web site on how to remove some of this private information from a

History


Fingerprinting of Office 97 files
http://www.computerbytesman.com/privacy/office97.htm
Google
ZOË
Python
News
blogs
B!
bml!
W!
recherche
c00l
archive.org
misc
LuFG

Home > Privacy > Fingerprinting of Office 97 files

In Office 97, when an Word, Excel, or PowerPoint file is saved for the first time, it is assigned its own unique serial number. This serial number is the form a 32-digit GUID (Globally Unique ID). The last 12 digits of a GUID will most likely contain the MAC or NIC address of Ethernet adapter of the person saving the document. Since Ethernet addresses are unique, this serial number in theory would allow a document to be traced back to the computer it was created on.

The GUID serial numnber was originally put in Office 97 document files to correct broken hyperlinks. Ironically this feature was never implemented, but the serial numbers remain.

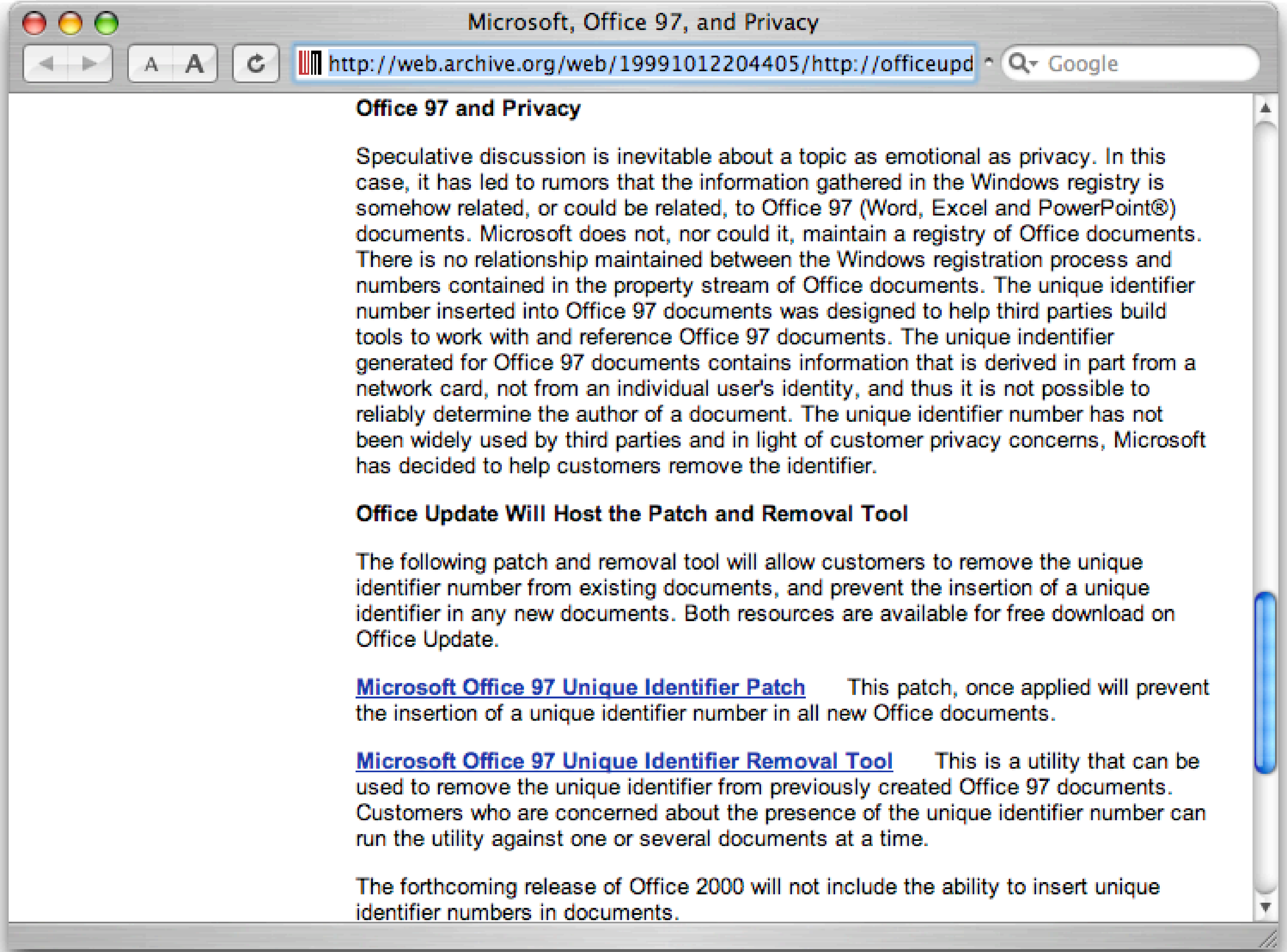
To locate a GUID in a Word document, simply open the .DOC file in Notepad and search for the string "GUID". The GUID serial number will follow immediately in the document.

To fix this problem, Microsoft is providing a patch to Office 97 which will stop putting serial numbers in new document files. For existing files, Microsoft will also be providing a stripper utility for removing serial numbers. Click here for more details.

The Office 97 issue is independent of the operating system and will occur under both Windows 95 and Windows98. There are reports that the problem also exists in Office 98 documents for the Macintosh.

Note that in Office 2000, GUIDs are no longer generated for document files. However a number of GUIDs are included in a file if the document contains VBA macros. There is currently no method of removing these macro GUIDs except to delete the macros themselves.

History



Microsoft, Office 97, and Privacy
http://web.archive.org/web/19991012204405/http://officeupd
Google

Office 97 and Privacy

Speculative discussion is inevitable about a topic as emotional as privacy. In this case, it has led to rumors that the information gathered in the Windows registry is somehow related, or could be related, to Office 97 (Word, Excel and PowerPoint®) documents. Microsoft does not, nor could it, maintain a registry of Office documents. There is no relationship maintained between the Windows registration process and numbers contained in the property stream of Office documents. The unique identifier number inserted into Office 97 documents was designed to help third parties build tools to work with and reference Office 97 documents. The unique indentifier generated for Office 97 documents contains information that is derived in part from a network card, not from an individual user's identity, and thus it is not possible to reliably determine the author of a document. The unique identifier number has not been widely used by third parties and in light of customer privacy concerns, Microsoft has decided to help customers remove the identifier.

Office Update Will Host the Patch and Removal Tool

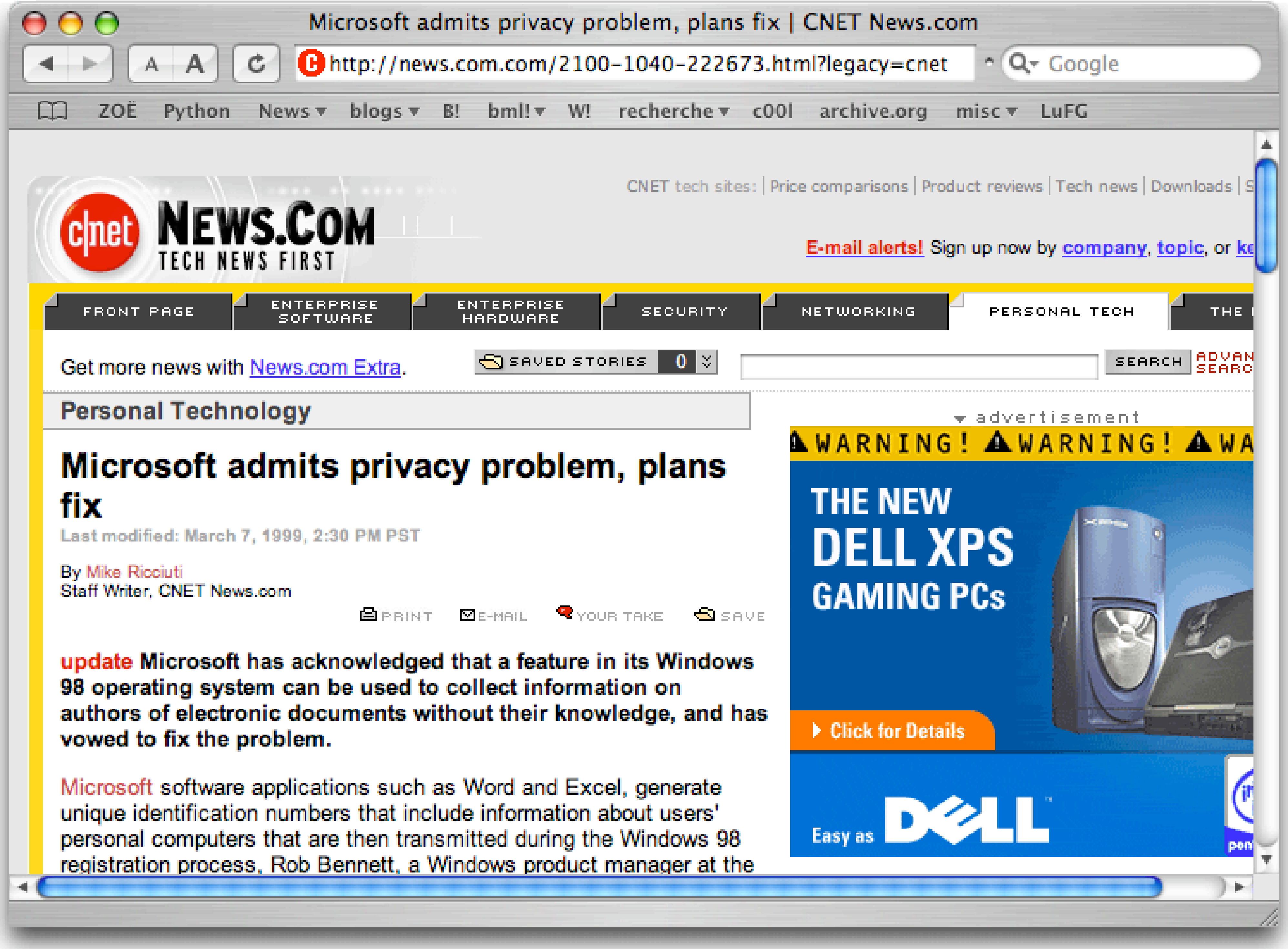
The following patch and removal tool will allow customers to remove the unique identifier number from existing documents, and prevent the insertion of a unique identifier in any new documents. Both resources are available for free download on Office Update.

Microsoft Office 97 Unique Identifier Patch This patch, once applied will prevent the insertion of a unique identifier number in all new Office documents.

Microsoft Office 97 Unique Identifier Removal Tool This is a utility that can be used to remove the unique identifier from previously created Office 97 documents. Customers who are concerned about the presence of the unique identifier number can run the utility against one or several documents at a time.

The forthcoming release of Office 2000 will not include the ability to insert unique identifier numbers in documents.

History



E-mail alerts! Sign up now by company, topic, or ke

Get more news with News.com Extra.

Personal Technology

Microsoft admits privacy problem, plans fix

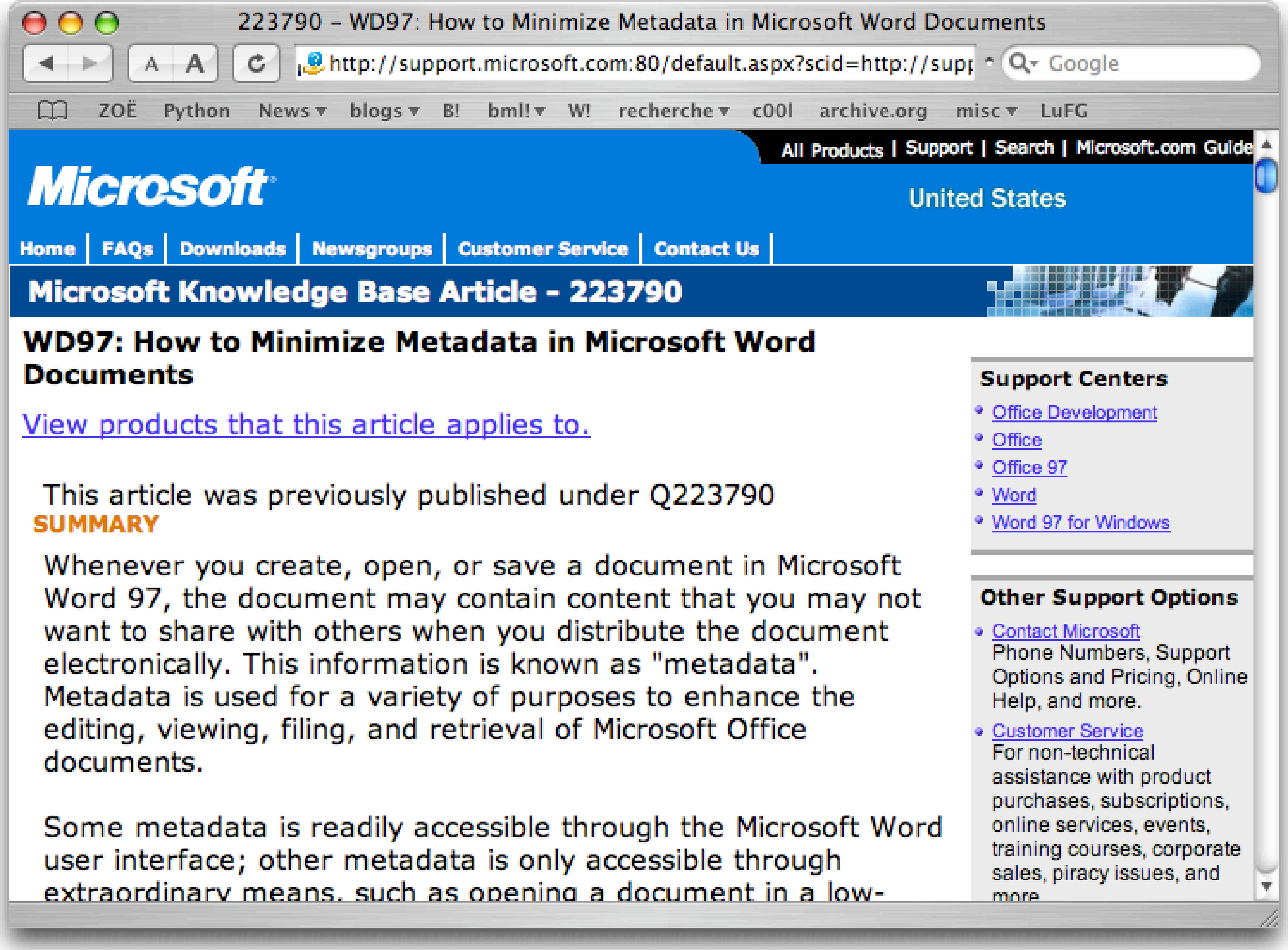
Last modified: March 7, 1999, 2:30 PM PST

By Mike Ricciuti
Staff Writer, CNET News.com

update Microsoft has acknowledged that a feature in its Windows 98 operating system can be used to collect information on authors of electronic documents without their knowledge, and has vowed to fix the problem.

Microsoft software applications such as Word and Excel, generate unique identification numbers that include information about users' personal computers that are then transmitted during the Windows 98 registration process. Rob Bennett, a Windows product manager at the

History



223790 – WD97: How to Minimize Metadata in Microsoft Word Documents
http://support.microsoft.com:80/default.aspx?scid=http://supp
Google
ZOË Python News blogs B! bml! W! recherche c00l archive.org misc LuFG
All Products | Support | Search | Microsoft.com Guide
Microsoft
United States
Home | FAQs | Downloads | Newsgroups | Customer Service | Contact Us

Microsoft Knowledge Base Article - 223790

WD97: How to Minimize Metadata in Microsoft Word Documents

View products that this article applies to.

This article was previously published under Q223790

SUMMARY

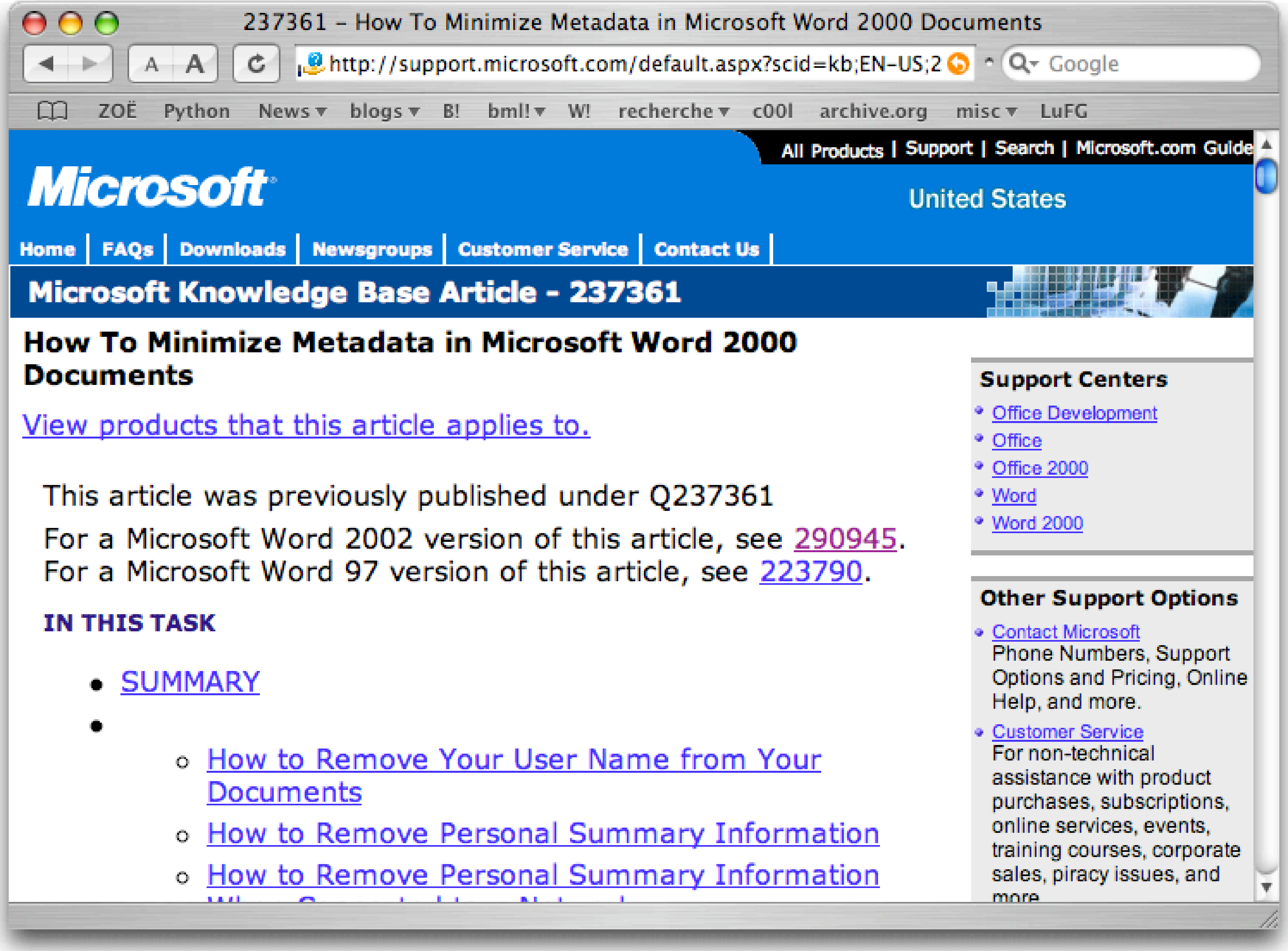
Whenever you create, open, or save a document in Microsoft Word 97, the document may contain content that you may not want to share with others when you distribute the document electronically. This information is known as "metadata". Metadata is used for a variety of purposes to enhance the editing, viewing, filing, and retrieval of Microsoft Office documents.

Some metadata is readily accessible through the Microsoft Word user interface; other metadata is only accessible through extraordinary means, such as opening a document in a low-

Support Centers
- Office Development
- Office
- Office 97
- Word
- Word 97 for Windows

Other Support Options
- Contact Microsoft
  Phone Numbers, Support Options and Pricing, Online Help, and more.
- Customer Service
  For non-technical assistance with product purchases, subscriptions, online services, events, training courses, corporate sales, piracy issues, and more

History


237361 – How To Minimize Metadata in Microsoft Word 2000 Documents
http://support.microsoft.com/default.aspx?scid=kb;EN-US;2
Google
ZOË Python News blogs B! bml! W! recherche c00l archive.org misc LuFG
All Products | Support | Search | Microsoft.com Guide
Microsoft
United States
Home | FAQs | Downloads | Newsgroups | Customer Service | Contact Us

Microsoft Knowledge Base Article - 237361

How To Minimize Metadata in Microsoft Word 2000 Documents

View products that this article applies to.

This article was previously published under Q237361

For a Microsoft Word 2002 version of this article, see 290945.
For a Microsoft Word 97 version of this article, see 223790.

IN THIS TASK

- SUMMARY
-
  - How to Remove Your User Name from Your Documents
  - How to Remove Personal Summary Information
  - How to Remove Personal Summary Information

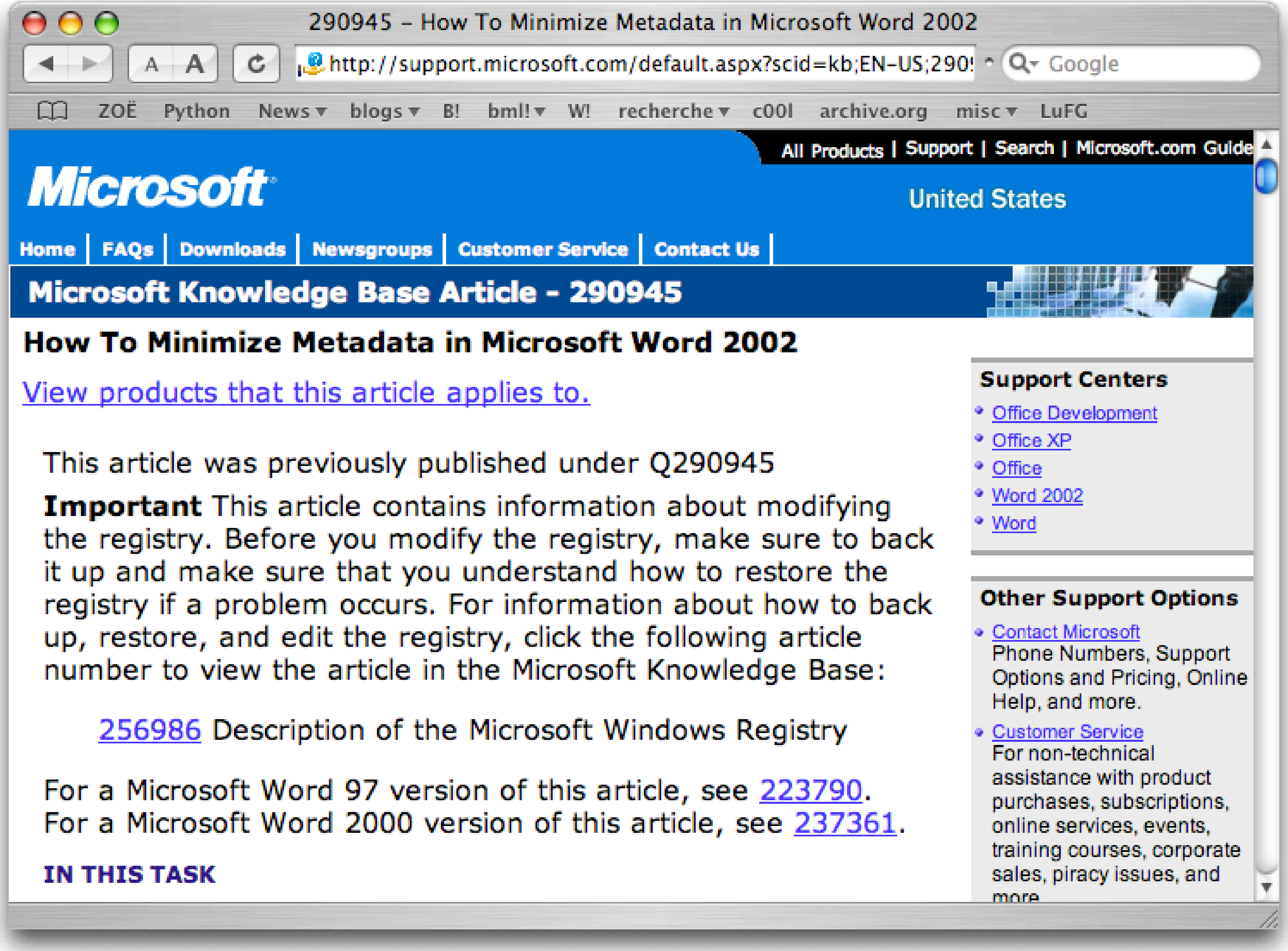
Support Centers

- Office Development
- Office
- Office 2000
- Word
- Word 2000

Other Support Options

- Contact Microsoft
  Phone Numbers, Support Options and Pricing, Online Help, and more.
- Customer Service
  For non-technical assistance with product purchases, subscriptions, online services, events, training courses, corporate sales, piracy issues, and more

History



290945 – How To Minimize Metadata in Microsoft Word 2002
http://support.microsoft.com/default.aspx?scid=kb;EN-US;290!
Google
ZOË Python News blogs B! bml! W! recherche c00l archive.org misc LuFG
All Products | Support | Search | Microsoft.com Guide
Microsoft
United States
Home | FAQs | Downloads | Newsgroups | Customer Service | Contact Us

Microsoft Knowledge Base Article - 290945

How To Minimize Metadata in Microsoft Word 2002

View products that this article applies to.

This article was previously published under Q290945

Important This article contains information about modifying the registry. Before you modify the registry, make sure to back it up and make sure that you understand how to restore the registry if a problem occurs. For information about how to back up, restore, and edit the registry, click the following article number to view the article in the Microsoft Knowledge Base:

256986 Description of the Microsoft Windows Registry

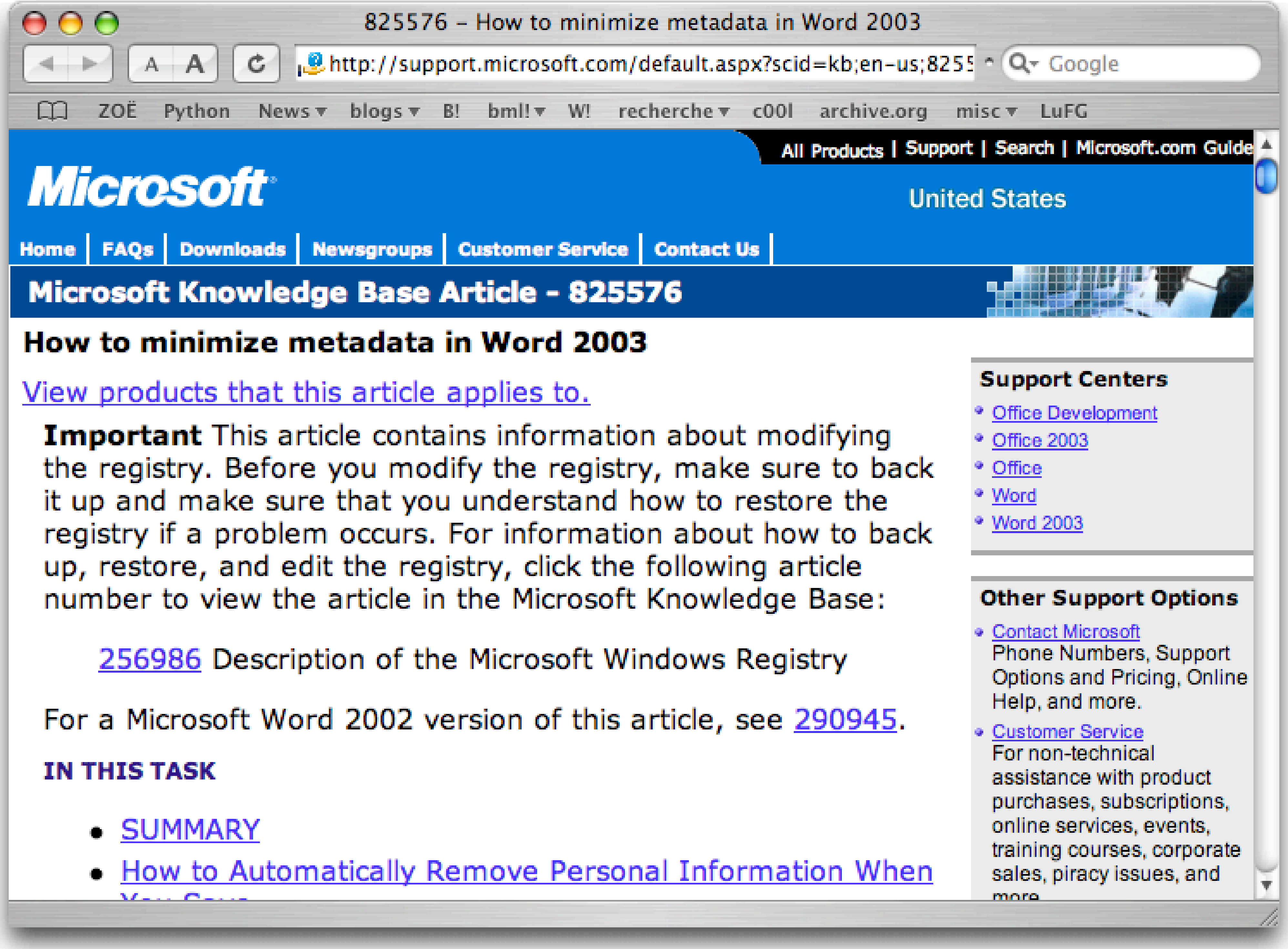
For a Microsoft Word 97 version of this article, see 223790.
For a Microsoft Word 2000 version of this article, see 237361.

IN THIS TASK

Support Centers
- Office Development
- Office XP
- Office
- Word 2002
- Word

Other Support Options
- Contact Microsoft
Phone Numbers, Support Options and Pricing, Online Help, and more.
- Customer Service
For non-technical assistance with product purchases, subscriptions, online services, events, training courses, corporate sales, piracy issues, and more

History


825576 – How to minimize metadata in Word 2003
http://support.microsoft.com/default.aspx?scid=kb;en-us;8255
Google
ZOË
Python
News
blogs
B!
bml!
W!
recherche
c00l
archive.org
misc
LuFG
All Products | Support | Search | Microsoft.com Guide
Microsoft
United States
Home | FAQs | Downloads | Newsgroups | Customer Service | Contact Us

Microsoft Knowledge Base Article - 825576

How to minimize metadata in Word 2003

View products that this article applies to.

Important This article contains information about modifying the registry. Before you modify the registry, make sure to back it up and make sure that you understand how to restore the registry if a problem occurs. For information about how to back up, restore, and edit the registry, click the following article number to view the article in the Microsoft Knowledge Base:

256986 Description of the Microsoft Windows Registry

For a Microsoft Word 2002 version of this article, see 290945.

IN THIS TASK

- SUMMARY
- How to Automatically Remove Personal Information When You Save

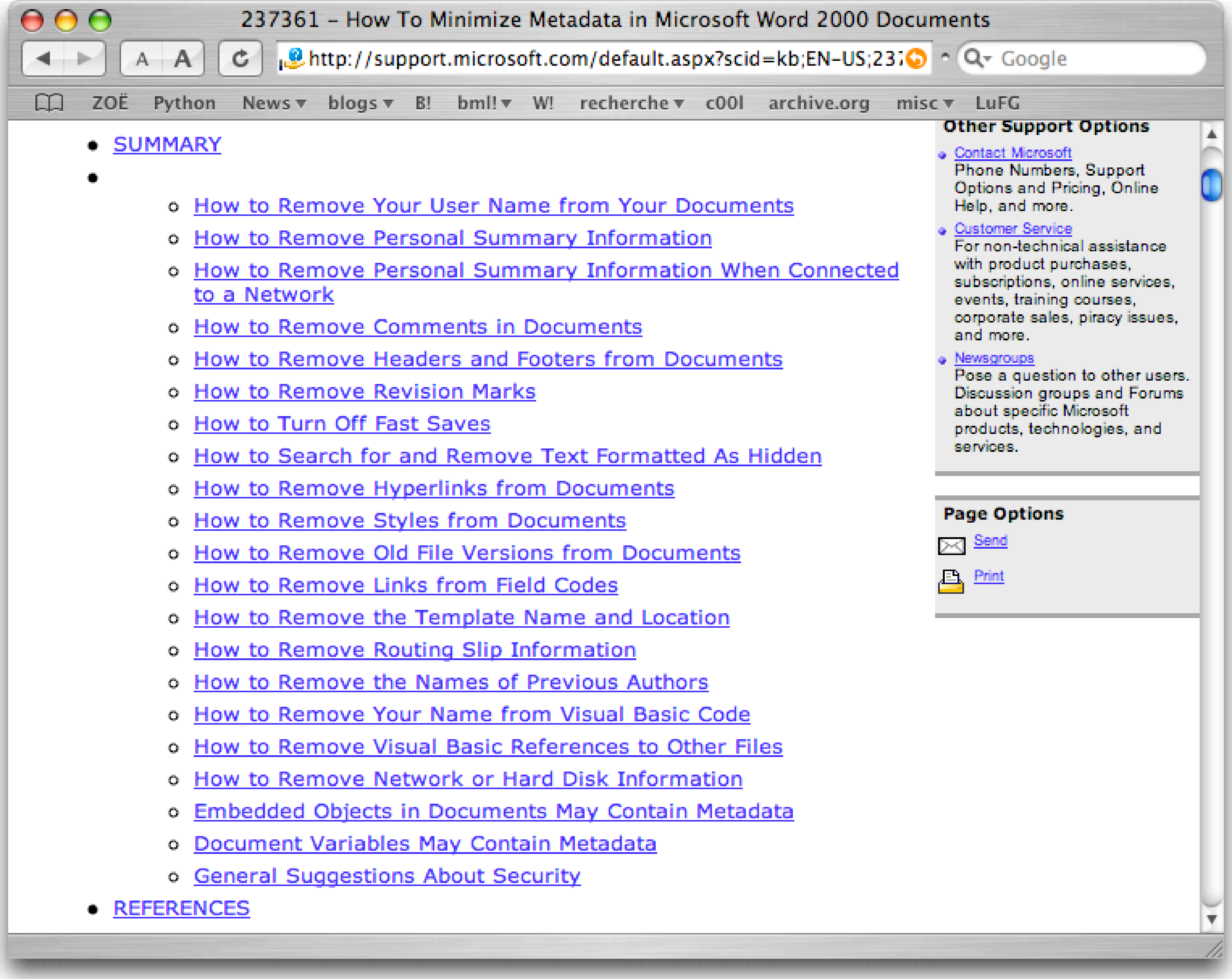
Support Centers

- Office Development
- Office 2003
- Office
- Word
- Word 2003

Other Support Options

- Contact Microsoft
  Phone Numbers, Support Options and Pricing, Online Help, and more.
- Customer Service
  For non-technical assistance with product purchases, subscriptions, online services, events, training courses, corporate sales, piracy issues, and more

What's in there?



237361 – How To Minimize Metadata in Microsoft Word 2000 Documents

http://support.microsoft.com/default.aspx?scid=kb;EN-US;237

- SUMMARY
-
  - How to Remove Your User Name from Your Documents
  - How to Remove Personal Summary Information
  - How to Remove Personal Summary Information When Connected to a Network
  - How to Remove Comments in Documents
  - How to Remove Headers and Footers from Documents
  - How to Remove Revision Marks
  - How to Turn Off Fast Saves
  - How to Search for and Remove Text Formatted As Hidden
  - How to Remove Hyperlinks from Documents
  - How to Remove Styles from Documents
  - How to Remove Old File Versions from Documents
  - How to Remove Links from Field Codes
  - How to Remove the Template Name and Location
  - How to Remove Routing Slip Information
  - How to Remove the Names of Previous Authors
  - How to Remove Your Name from Visual Basic Code
  - How to Remove Visual Basic References to Other Files
  - How to Remove Network or Hard Disk Information
  - Embedded Objects in Documents May Contain Metadata
  - Document Variables May Contain Metadata
  - General Suggestions About Security
- REFERENCES

Other Support Options

- Contact Microsoft
  Phone Numbers, Support Options and Pricing, Online Help, and more.
- Customer Service
  For non-technical assistance with product purchases, subscriptions, online services, events, training courses, corporate sales, piracy issues, and more.
- Newsgroups
  Pose a question to other users. Discussion groups and Forums about specific Microsoft products, technologies, and services.

Page Options

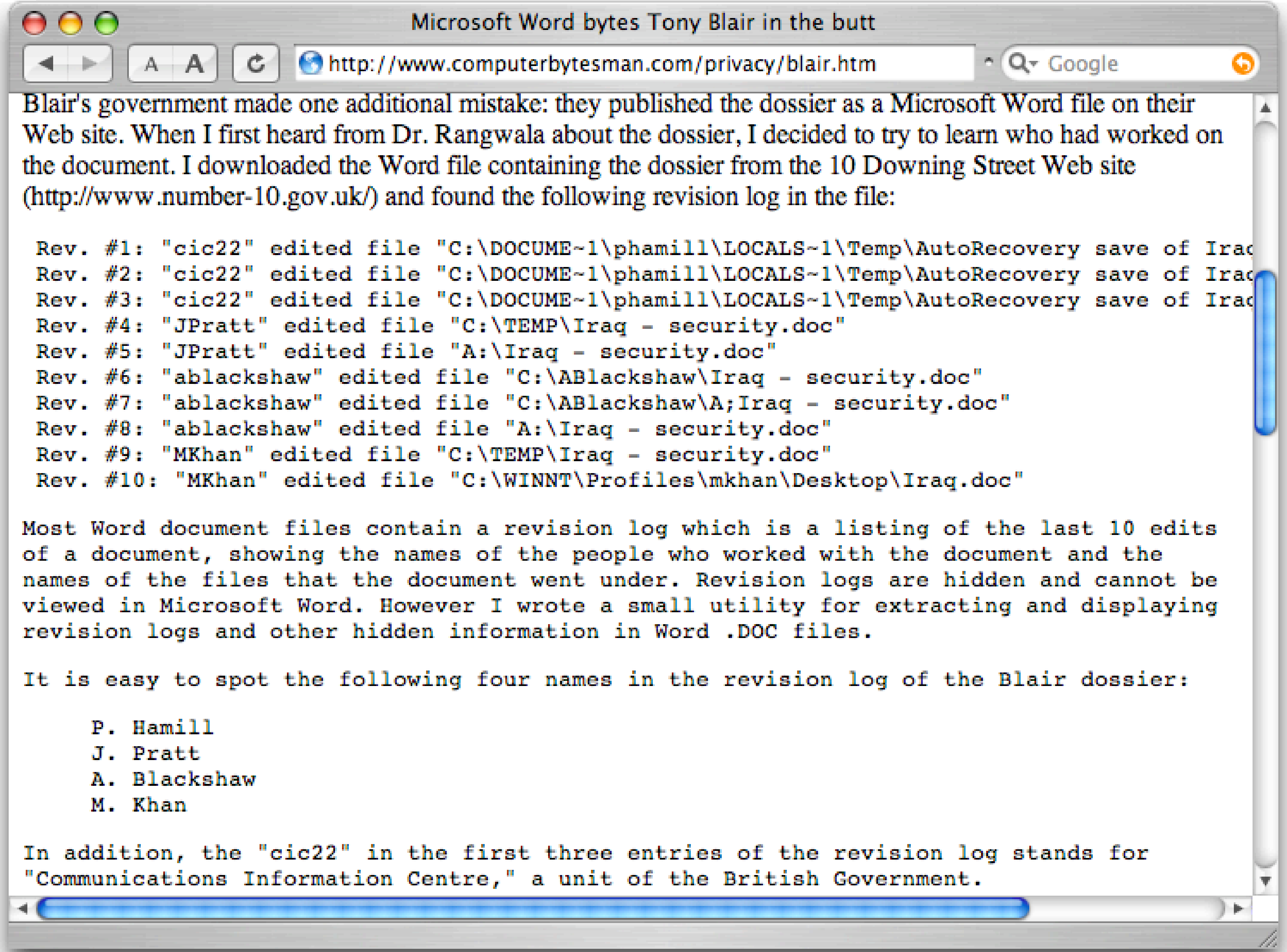
Send

Print

Incidents

- UK Irak Dossier
- Transrapid / Rheinbraun / Managment-Machbarkeitsstudie
- Melissa

Irak Dossier


Microsoft Word bytes Tony Blair in the butt
http://www.computerbytesman.com/privacy/blair.htm
Google

Blair's government made one additional mistake: they published the dossier as a Microsoft Word file on their Web site. When I first heard from Dr. Rangwala about the dossier, I decided to try to learn who had worked on the document. I downloaded the Word file containing the dossier from the 10 Downing Street Web site (http://www.number-10.gov.uk/) and found the following revision log in the file:

```
 Rev. #1: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Irac
 Rev. #2: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Irac
 Rev. #3: "cic22" edited file "C:\DOCUME~1\phamill\LOCALS~1\Temp\AutoRecovery save of Irac
 Rev. #4: "JPratt" edited file "C:\TEMP\Iraq - security.doc"
 Rev. #5: "JPratt" edited file "A:\Iraq - security.doc"
 Rev. #6: "ablackshaw" edited file "C:\ABlackshaw\Iraq - security.doc"
 Rev. #7: "ablackshaw" edited file "C:\ABlackshaw\A;Iraq - security.doc"
 Rev. #8: "ablackshaw" edited file "A:\Iraq - security.doc"
 Rev. #9: "MKhan" edited file "C:\TEMP\Iraq - security.doc"
 Rev. #10: "MKhan" edited file "C:\WINNT\Profiles\mkhan\Desktop\Iraq.doc"
```

Most Word document files contain a revision log which is a listing of the last 10 edits of a document, showing the names of the people who worked with the document and the names of the files that the document went under. Revision logs are hidden and cannot be viewed in Microsoft Word. However I wrote a small utility for extracting and displaying revision logs and other hidden information in Word .DOC files.

It is easy to spot the following four names in the revision log of the Blair dossier:

```
     P. Hamill
     J. Pratt
     A. Blackshaw
     M. Khan
```

In addition, the "cic22" in the first three entries of the revision log stands for "Communications Information Centre," a unit of the British Government.

Incidents

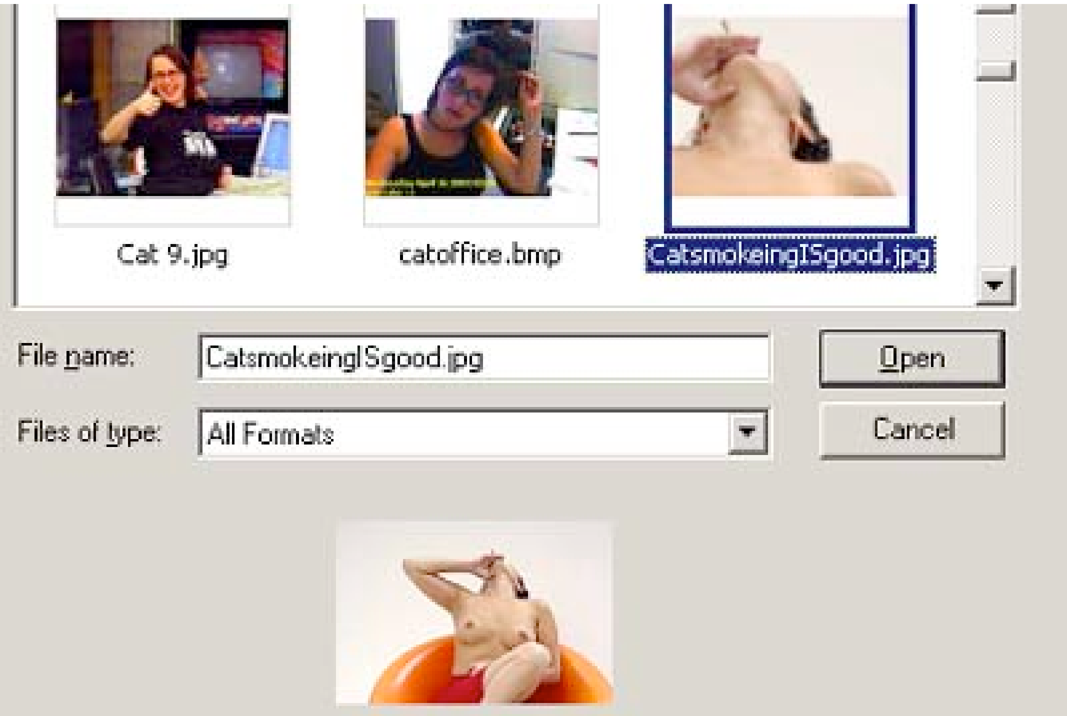
- Documented incidents include:
  - Text from a completely unrelated document edited before appears in the file.
  - Data deleted from the document or overwritten is appears in the file.

JPEG/EXIF

- Many image formats contain comment fields which might disclose unwanted data.
- JPEG has the extensible EXIF format for meta data.
- There was a remarkable incident with EXIF thumbnails:

techTV incident

techTV incident



File name: CatsmokeingISgood.jpg

Files of type: All Formats

Misc

- Starr Report
- Embedded Serials / GUIDS
- “unregistered” marks
- ...

An Experiment

Laboratory for Dependable Distributed Systems

RHEINISCH-
WESTFÄLISCHE
TECHNISCHE
HOCHSCHULE
AACHEN

RWTH

UNIVERSITY
OF TECHNOLOGY
AACHEN

An Experiment

- Idea
  - Crawl the Web
  - Download Documents
  - Find the ones with hidden data.
- Problem:
  - How to detect hidden data?

The Byers Experiment

- Scalable Exploitation of, and Responses to Information Leakage Through Hidden Data in Published Documents, Simon Byers, IEEE Security & Privacy pp. 23-27, March / April 2004

My Experiment

- Crawling the Web <u>fast</u> is fun and hard
- lenz3 is my own experimental crawler
  - testbed for "directed crawling"
  - If it's breaks, you can keep the parts: http://c0re.23.nu/c0de/lenz2/

“Crawl”

- Brute-Force approach: Nils Provos’ crawl -0.4
- Needed a patch to l33ch MS Office documents instead of pictures: http://c0re.23.nu/c0de/misc/crawl–0.4–doc.patch
- Got a sustained transferrate of about 1000 kb/s on a very fast Link.

Execution

- Used searchengines as seed
- Selected downlodes Documents if
  - MIME-type application/(msexel | mspowerpoint | msword | vnd.ms-excel | vnd.ms-powerpoint | vnd.ms-word | xls)
  - MIME-type application/octet-stream and filename ended in .doc, .ppt, .xls
- Collected about 150.000 Documents

Analysis

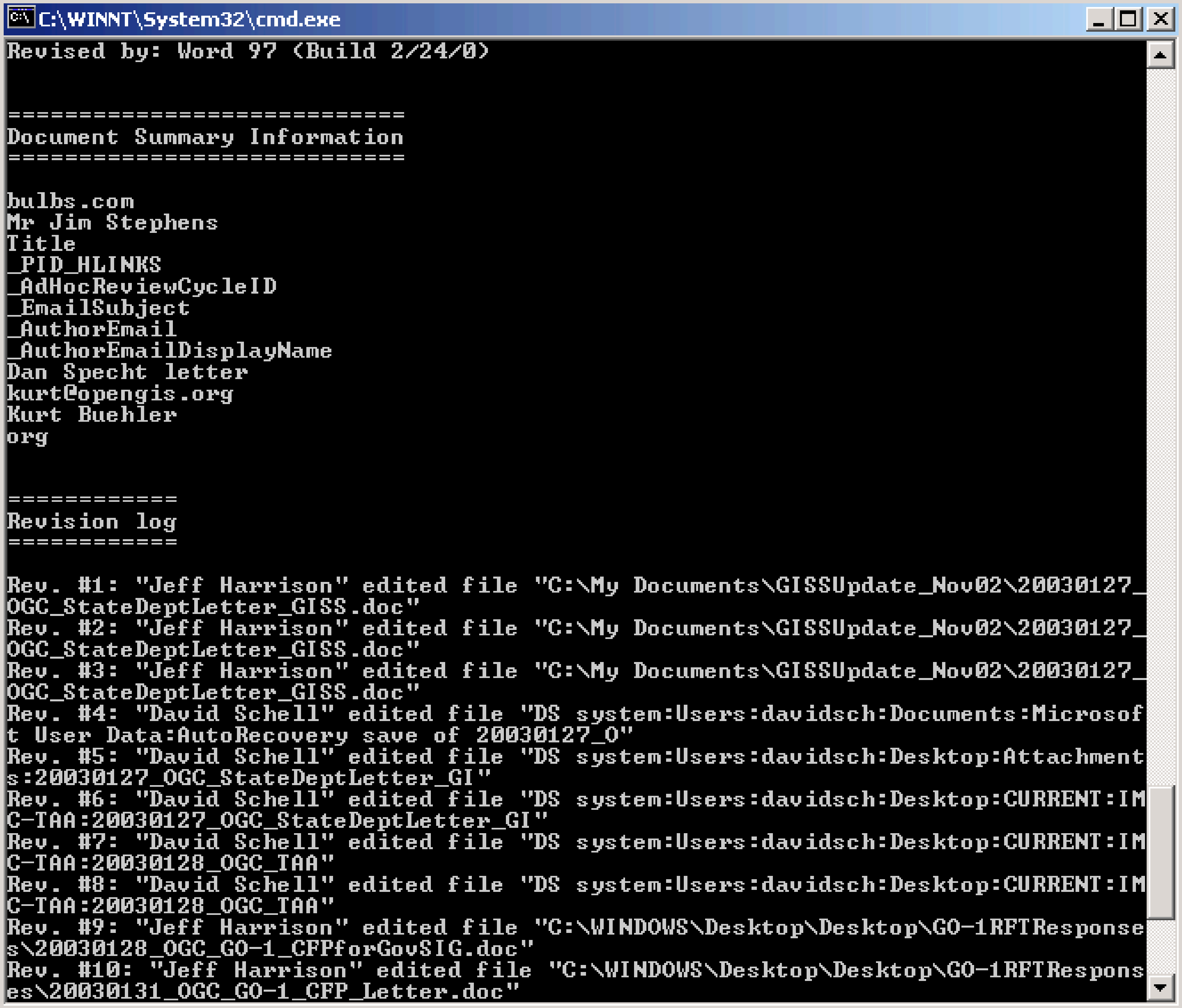
- Ran strings, antiword, catdoc, wv* on all files and dumped the output to files.
- Skimmed the resulds by hand / eye.

Results

- Seen unbelievable misconfigurations of Web servers.
    - E.g. status code 226
- Seen nothing breathtaking but many things i cosider interesting.

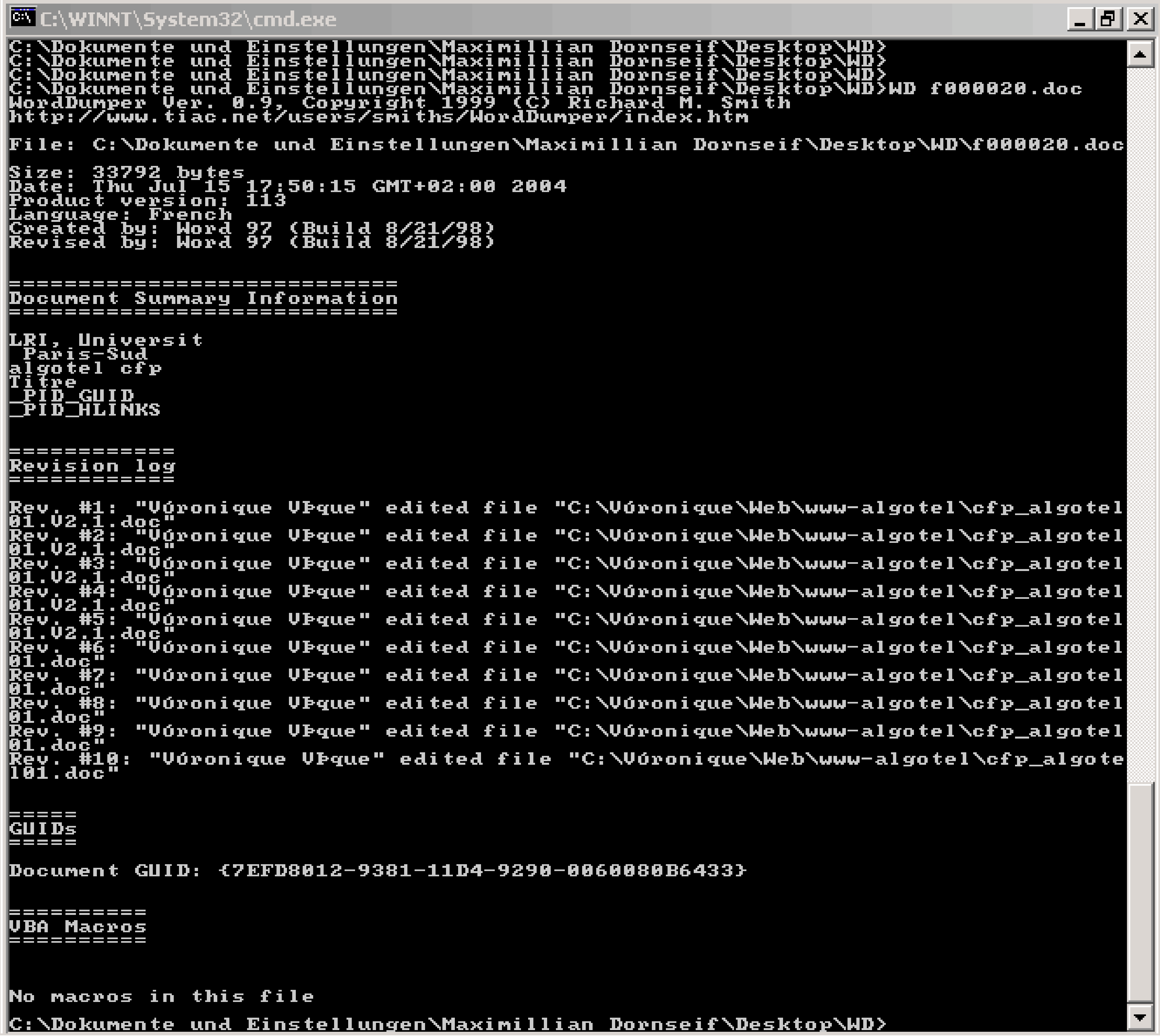
MS Office 2003

MS Office 2000



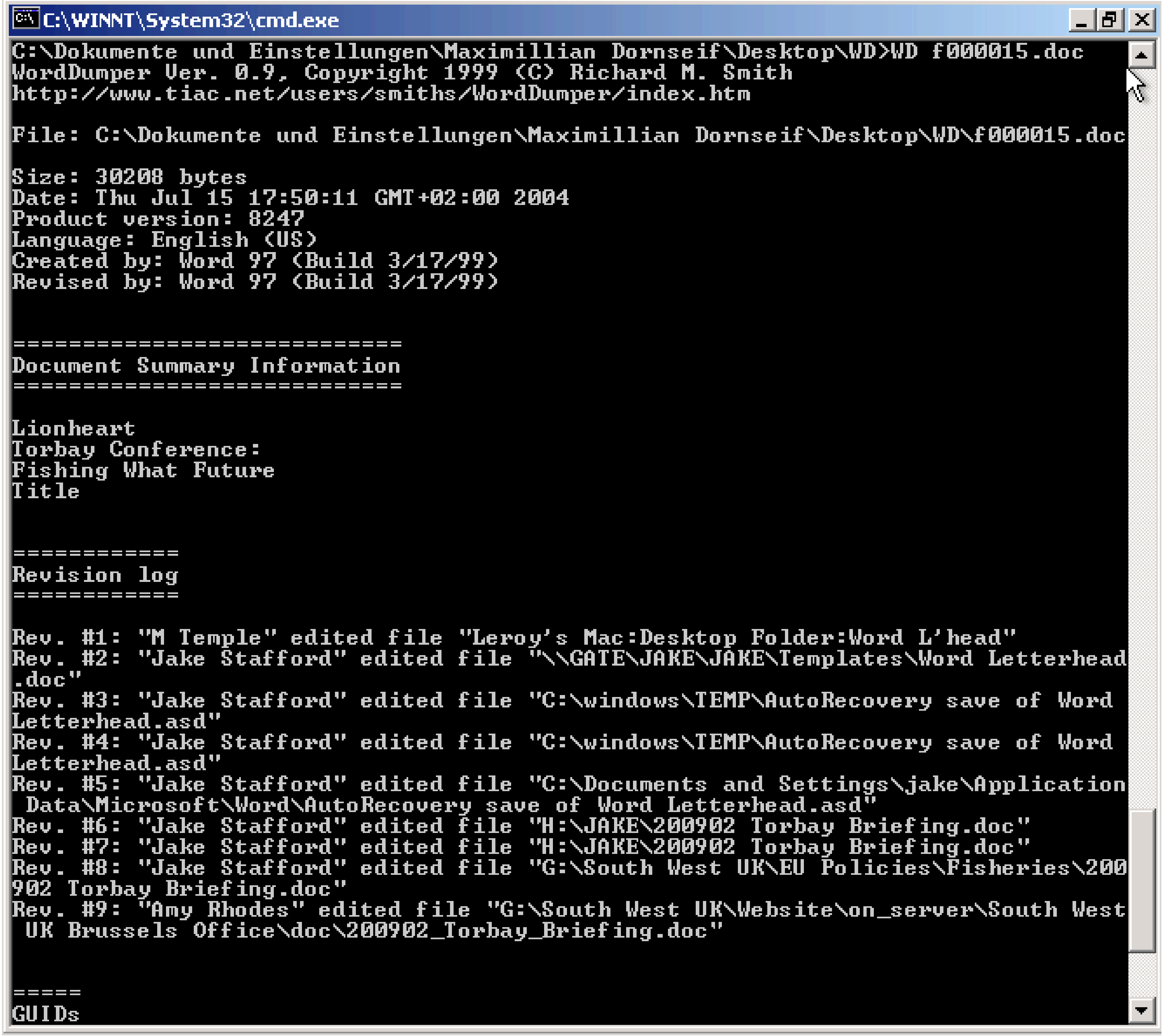
```
Revised by: Word 97 (Build 2/24/0)


=============================
Document Summary Information
=============================

bulbs.com
Mr Jim Stephens
Title
_PID_HLINKS
_AdHocReviewCycleID
_EmailSubject
_AuthorEmail
_AuthorEmailDisplayName
Dan Specht letter
kurt@opengis.org
Kurt Buehler
org


============
Revision log
============

Rev. #1: "Jeff Harrison" edited file "C:\My Documents\GISSUpdate_Nov02\20030127_
OGC_StateDeptLetter_GISS.doc"
Rev. #2: "Jeff Harrison" edited file "C:\My Documents\GISSUpdate_Nov02\20030127_
OGC_StateDeptLetter_GISS.doc"
Rev. #3: "Jeff Harrison" edited file "C:\My Documents\GISSUpdate_Nov02\20030127_
OGC_StateDeptLetter_GISS.doc"
Rev. #4: "David Schell" edited file "DS system:Users:davidsch:Documents:Microsof
t User Data:AutoRecovery save of 20030127_O"
Rev. #5: "David Schell" edited file "DS system:Users:davidsch:Desktop:Attachment
s:20030127_OGC_StateDeptLetter_GI"
Rev. #6: "David Schell" edited file "DS system:Users:davidsch:Desktop:CURRENT:IM
C-TAA:20030127_OGC_StateDeptLetter_GI"
Rev. #7: "David Schell" edited file "DS system:Users:davidsch:Desktop:CURRENT:IM
C-TAA:20030128_OGC_TAA"
Rev. #8: "David Schell" edited file "DS system:Users:davidsch:Desktop:CURRENT:IM
C-TAA:20030128_OGC_TAA"
Rev. #9: "Jeff Harrison" edited file "C:\WINDOWS\Desktop\Desktop\GO-1RFTResponse
s\20030128_OGC_GO-1_CFPforGovSIG.doc"
Rev. #10: "Jeff Harrison" edited file "C:\WINDOWS\Desktop\Desktop\GO-1RFTRespons
es\20030131_OGC_GO-1_CFP_Letter.doc"
```

GUID



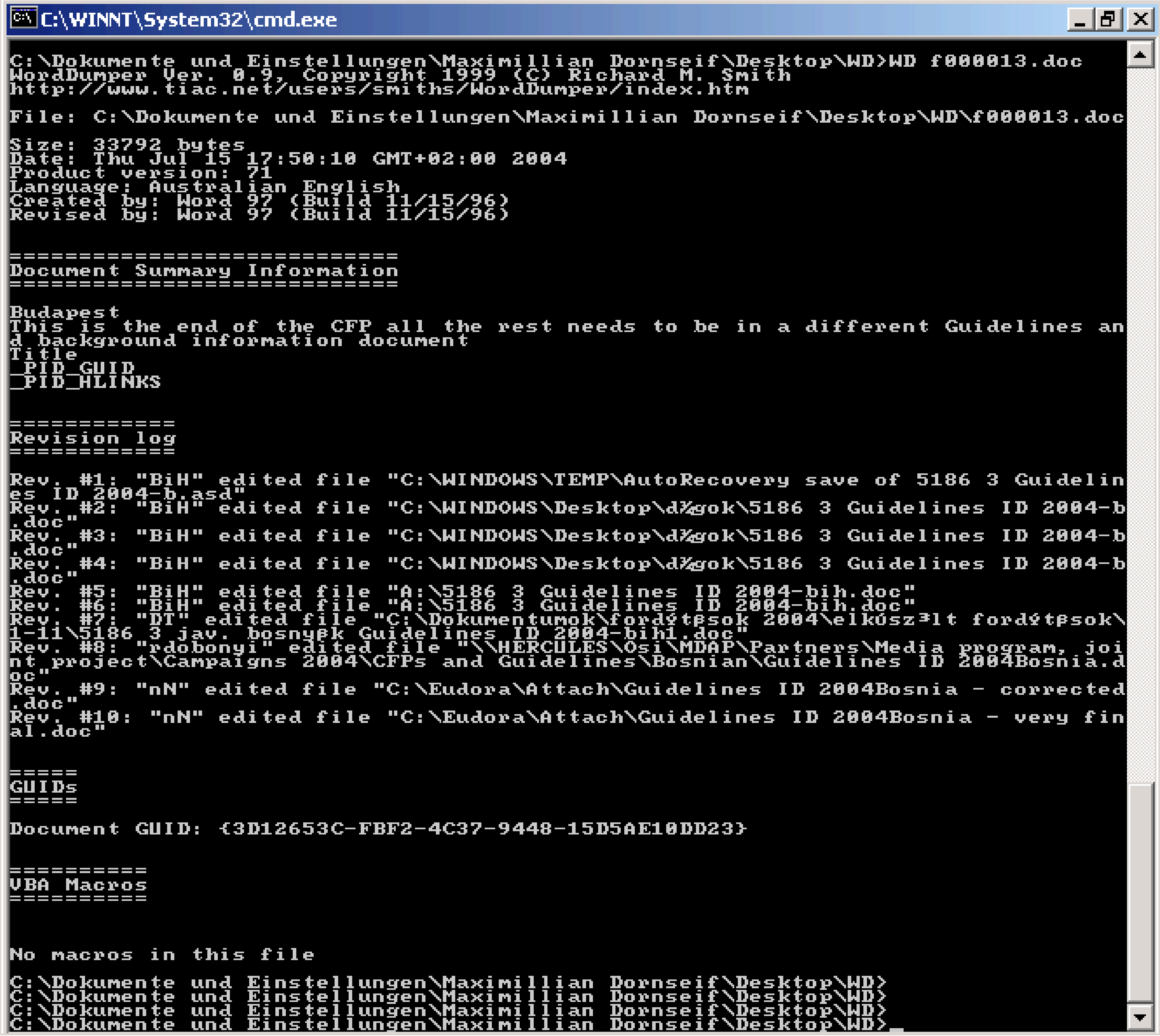
```
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>WD f000020.doc
WordDumper Ver. 0.9, Copyright 1999 (C) Richard M. Smith
http://www.tiac.net/users/smiths/WordDumper/index.htm

File: C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD\f000020.doc

Size: 33792 bytes
Date: Thu Jul 15 17:50:15 GMT+02:00 2004
Product version: 113
Language: French
Created by: Word 97 (Build 8/21/98)
Revised by: Word 97 (Build 8/21/98)


============================
Document Summary Information
============================

LRI, Universit
 Paris-Sud
algotel cfp
Titre
_PID_GUID
_PID_HLINKS


============
Revision log
============

Rev. #1: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.V2.1.doc"
Rev. #2: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.V2.1.doc"
Rev. #3: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.V2.1.doc"
Rev. #4: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.V2.1.doc"
Rev. #5: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.V2.1.doc"
Rev. #6: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.doc"
Rev. #7: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.doc"
Rev. #8: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.doc"
Rev. #9: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algotel
01.doc"
Rev. #10: "Vóronique Vþque" edited file "C:\Vóronique\Web\www-algotel\cfp_algote
l01.doc"


=====
GUIDs
=====

Document GUID: {7EFD8012-9381-11D4-9290-0060080B6433}


==========
VBA Macros
==========


No macros in this file

C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
```

History Odyssey



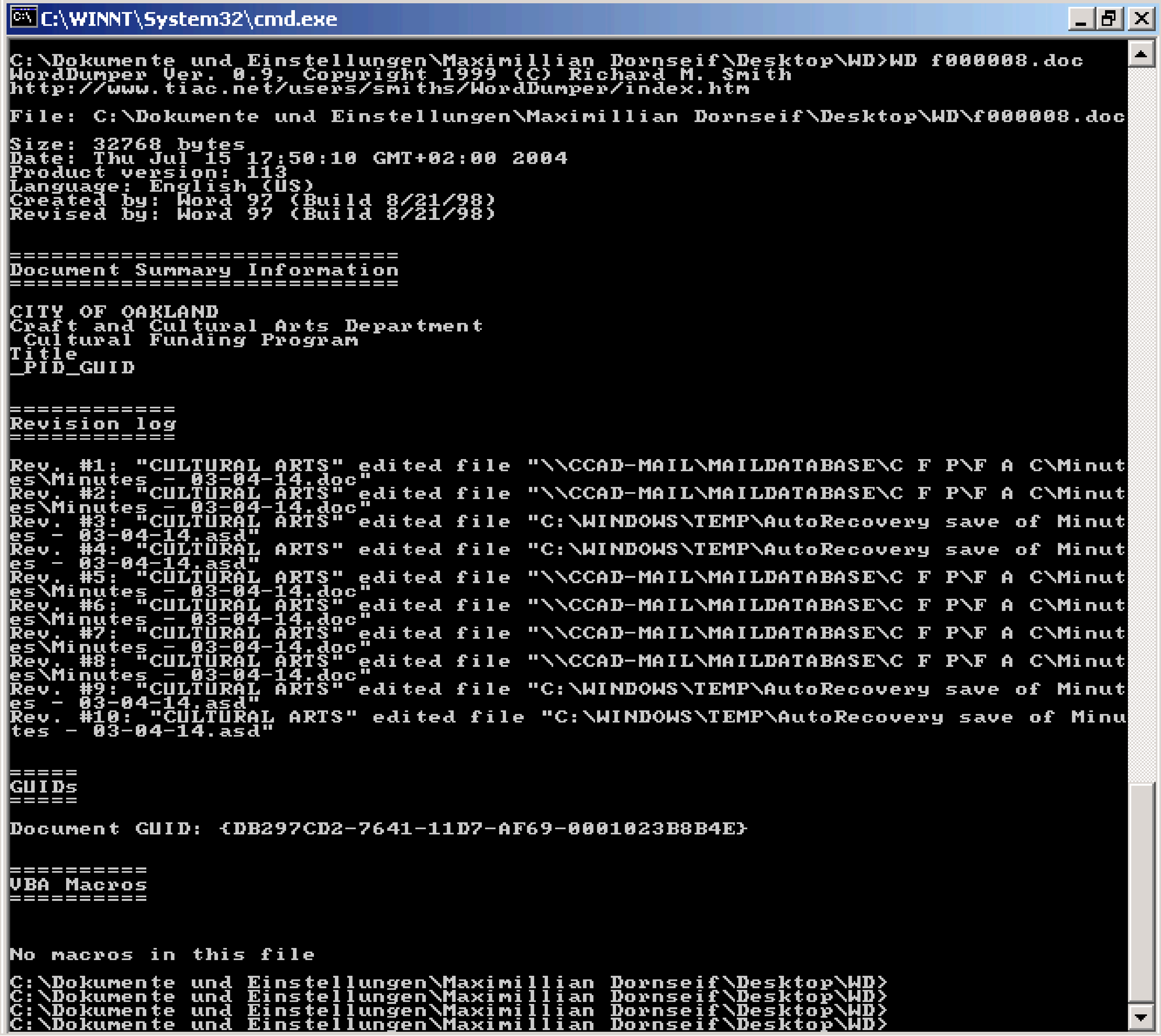
```
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>WD f000015.doc
WordDumper Ver. 0.9, Copyright 1999 (C) Richard M. Smith
http://www.tiac.net/users/smiths/WordDumper/index.htm

File: C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD\f000015.doc

Size: 30208 bytes
Date: Thu Jul 15 17:50:11 GMT+02:00 2004
Product version: 8247
Language: English (US)
Created by: Word 97 (Build 3/17/99)
Revised by: Word 97 (Build 3/17/99)


============================
Document Summary Information
============================

Lionheart
Torbay Conference:
Fishing What Future
Title


============
Revision log
============

Rev. #1: "M Temple" edited file "Leroy's Mac:Desktop Folder:Word L'head"
Rev. #2: "Jake Stafford" edited file "\\GATE\JAKE\JAKE\Templates\Word Letterhead
.doc"
Rev. #3: "Jake Stafford" edited file "C:\windows\TEMP\AutoRecovery save of Word
Letterhead.asd"
Rev. #4: "Jake Stafford" edited file "C:\windows\TEMP\AutoRecovery save of Word
Letterhead.asd"
Rev. #5: "Jake Stafford" edited file "C:\Documents and Settings\jake\Application
 Data\Microsoft\Word\AutoRecovery save of Word Letterhead.asd"
Rev. #6: "Jake Stafford" edited file "H:\JAKE\200902 Torbay Briefing.doc"
Rev. #7: "Jake Stafford" edited file "H:\JAKE\200902 Torbay Briefing.doc"
Rev. #8: "Jake Stafford" edited file "G:\South West UK\EU Policies\Fisheries\200
902 Torbay Briefing.doc"
Rev. #9: "Amy Rhodes" edited file "G:\South West UK\Website\on_server\South West
 UK Brussels Office\doc\200902_Torbay_Briefing.doc"


=====
GUIDs
```

71 Revisions

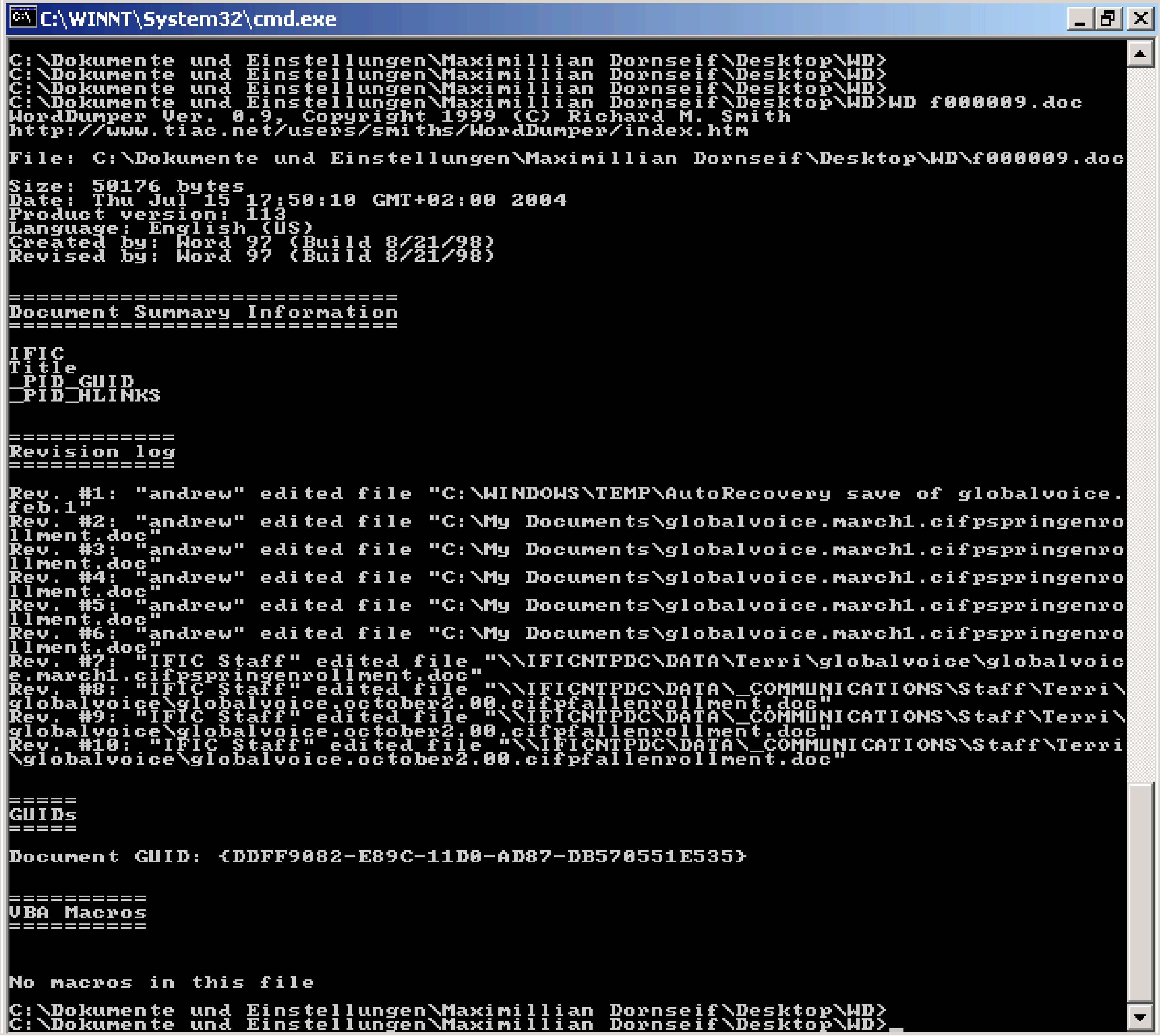


```
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>WD f000013.doc
WordDumper Ver. 0.9, Copyright 1999 (C) Richard M. Smith
http://www.tiac.net/users/smiths/WordDumper/index.htm

File: C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD\f000013.doc

Size: 33792 bytes
Date: Thu Jul 15 17:50:10 GMT+02:00 2004
Product version: 71
Language: Australian English
Created by: Word 97 (Build 11/15/96)
Revised by: Word 97 (Build 11/15/96)


============================
Document Summary Information
============================

Budapest
This is the end of the CFP all the rest needs to be in a different Guidelines an
d background information document
Title
_PID_GUID
_PID_HLINKS


============
Revision log
============

Rev. #1: "BiH" edited file "C:\WINDOWS\TEMP\AutoRecovery save of 5186 3 Guidelin
es ID 2004-b.asd"
Rev. #2: "BiH" edited file "C:\WINDOWS\Desktop\d½gok\5186 3 Guidelines ID 2004-b
.doc"
Rev. #3: "BiH" edited file "C:\WINDOWS\Desktop\d½gok\5186 3 Guidelines ID 2004-b
.doc"
Rev. #4: "BiH" edited file "C:\WINDOWS\Desktop\d½gok\5186 3 Guidelines ID 2004-b
.doc"
Rev. #5: "BiH" edited file "A:\5186 3 Guidelines ID 2004-bih.doc"
Rev. #6: "BiH" edited file "A:\5186 3 Guidelines ID 2004-bih.doc"
Rev. #7: "DT" edited file "C:\Dokumentumok\fordýtßsok 2004\elkósz³lt fordýtßsok\
1-11\5186 3 jav. bosnyßk Guidelines ID 2004-bih1.doc"
Rev. #8: "rdobonyi" edited file "\\HERCULES\Osi\MDAP\Partners\Media program, joi
nt project\Campaigns 2004\CFPs and Guidelines\Bosnian\Guidelines ID 2004Bosnia.d
oc"
Rev. #9: "nN" edited file "C:\Eudora\Attach\Guidelines ID 2004Bosnia - corrected
.doc"
Rev. #10: "nN" edited file "C:\Eudora\Attach\Guidelines ID 2004Bosnia - very fin
al.doc"


=====
GUIDs
=====

Document GUID: {3D12653C-FBF2-4C37-9448-15D5AE10DD23}


==========
VBA Macros
==========


No macros in this file

C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>_
```

Auto Recovey



```
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>WD f000008.doc
WordDumper Ver. 0.9, Copyright 1999 (C) Richard M. Smith
http://www.tiac.net/users/smiths/WordDumper/index.htm

File: C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD\f000008.doc

Size: 32768 bytes
Date: Thu Jul 15 17:50:10 GMT+02:00 2004
Product version: 113
Language: English (US)
Created by: Word 97 (Build 8/21/98)
Revised by: Word 97 (Build 8/21/98)


============================
Document Summary Information
============================

CITY OF OAKLAND
Craft and Cultural Arts Department
 Cultural Funding Program
Title
_PID_GUID


============
Revision log
============

Rev. #1: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #2: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #3: "CULTURAL ARTS" edited file "C:\WINDOWS\TEMP\AutoRecovery save of Minut
es - 03-04-14.asd"
Rev. #4: "CULTURAL ARTS" edited file "C:\WINDOWS\TEMP\AutoRecovery save of Minut
es - 03-04-14.asd"
Rev. #5: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #6: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #7: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #8: "CULTURAL ARTS" edited file "\\CCAD-MAIL\MAILDATABASE\C F P\F A C\Minut
es\Minutes - 03-04-14.doc"
Rev. #9: "CULTURAL ARTS" edited file "C:\WINDOWS\TEMP\AutoRecovery save of Minut
es - 03-04-14.asd"
Rev. #10: "CULTURAL ARTS" edited file "C:\WINDOWS\TEMP\AutoRecovery save of Minu
tes - 03-04-14.asd"


=====
GUIDs
=====

Document GUID: {DB297CD2-7641-11D7-AF69-0001023B8B4E}


==========
VBA Macros
==========


No macros in this file

C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
```



```
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>WD f000009.doc
WordDumper Ver. 0.9, Copyright 1999 (C) Richard M. Smith
http://www.tiac.net/users/smiths/WordDumper/index.htm

File: C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD\f000009.doc

Size: 50176 bytes
Date: Thu Jul 15 17:50:10 GMT+02:00 2004
Product version: 113
Language: English (US)
Created by: Word 97 (Build 8/21/98)
Revised by: Word 97 (Build 8/21/98)


============================
Document Summary Information
============================

IFIC
Title
_PID_GUID
_PID_HLINKS


============
Revision log
============

Rev. #1: "andrew" edited file "C:\WINDOWS\TEMP\AutoRecovery save of globalvoice.
feb.1"
Rev. #2: "andrew" edited file "C:\My Documents\globalvoice.march1.cifpspringenro
llment.doc"
Rev. #3: "andrew" edited file "C:\My Documents\globalvoice.march1.cifpspringenro
llment.doc"
Rev. #4: "andrew" edited file "C:\My Documents\globalvoice.march1.cifpspringenro
llment.doc"
Rev. #5: "andrew" edited file "C:\My Documents\globalvoice.march1.cifpspringenro
llment.doc"
Rev. #6: "andrew" edited file "C:\My Documents\globalvoice.march1.cifpspringenro
llment.doc"
Rev. #7: "IFIC Staff" edited file "\\IFICNTPDC\DATA\Terri\globalvoice\globalvoic
e.march1.cifpspringenrollment.doc"
Rev. #8: "IFIC Staff" edited file "\\IFICNTPDC\DATA\_COMMUNICATIONS\Staff\Terri\
globalvoice\globalvoice.october2.00.cifpfallenrollment.doc"
Rev. #9: "IFIC Staff" edited file "\\IFICNTPDC\DATA\_COMMUNICATIONS\Staff\Terri\
globalvoice\globalvoice.october2.00.cifpfallenrollment.doc"
Rev. #10: "IFIC Staff" edited file "\\IFICNTPDC\DATA\_COMMUNICATIONS\Staff\Terri
\globalvoice\globalvoice.october2.00.cifpfallenrollment.doc"


=====
GUIDs
=====

Document GUID: {DDFF9082-E89C-11D0-AD87-DB570551E535}


==========
VBA Macros
==========


No macros in this file

C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>
C:\Dokumente und Einstellungen\Maximillian Dornseif\Desktop\WD>_
```

Conclusions

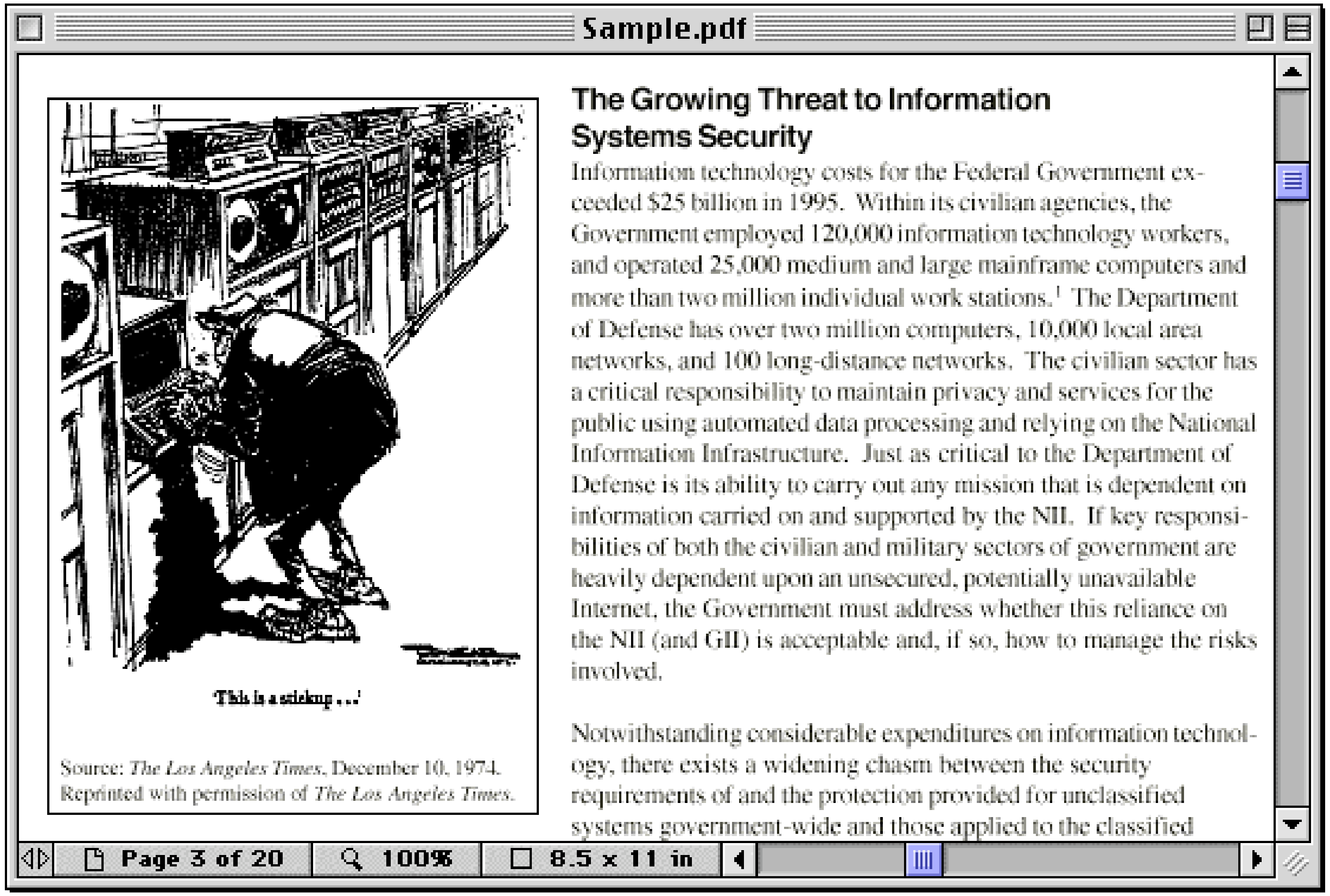
- You never know what properitary formats carry
- Open formats are only part of a solution
- Spider the web and enjoy

Bonus Track

Laboratory for Dependable Distributed Systems

RHEINISCH-
WESTFÄLISCHE
TECHNISCHE
HOCHSCHULE
AACHEN
RWTH
UNIVERSITY
OF TECHNOLOGY
AACHEN

PDF Scrubbing



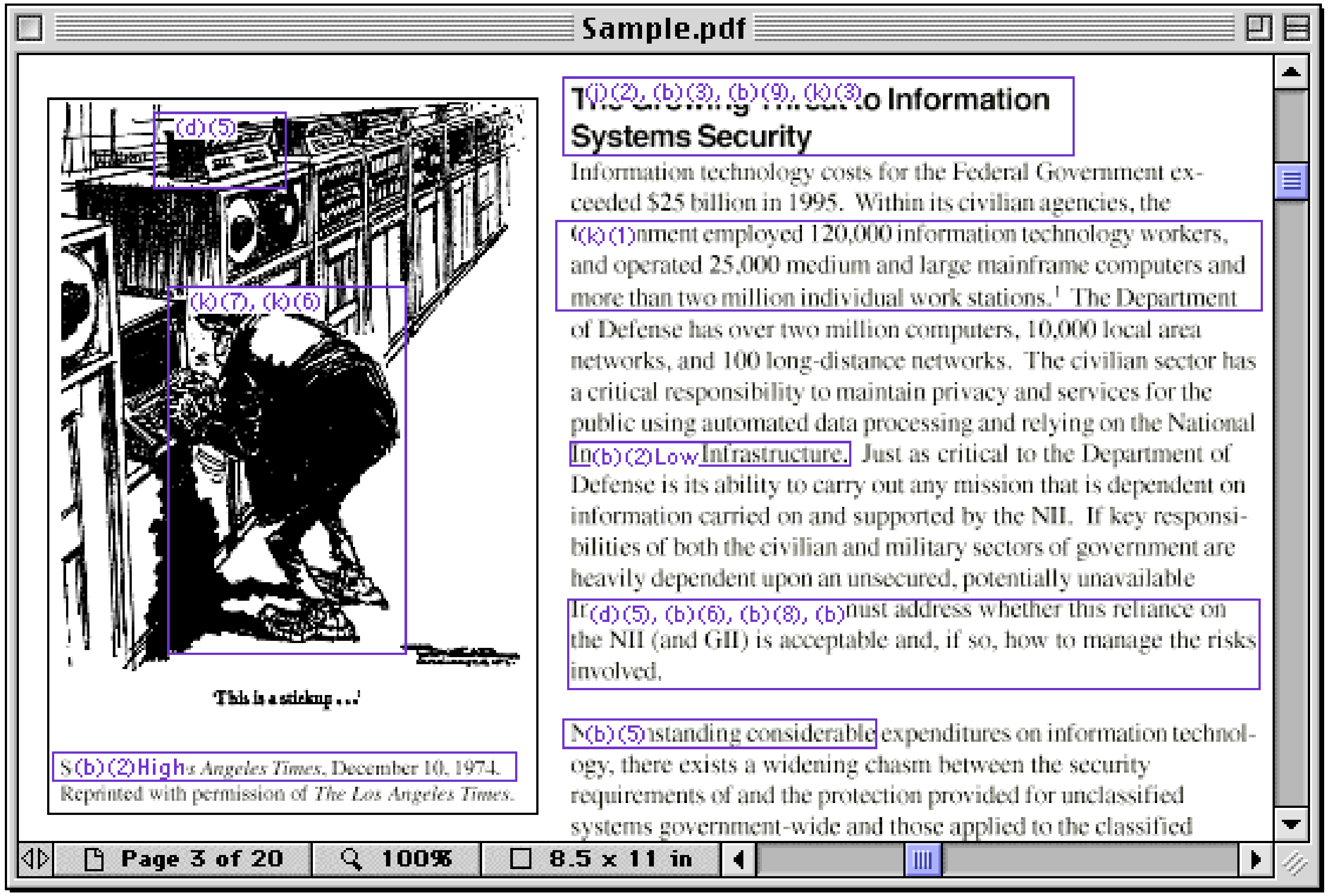
Source: *The Los Angeles Times*, December 10, 1974. Reprinted with permission of *The Los Angeles Times*.

The Growing Threat to Information Systems Security

Information technology costs for the Federal Government exceeded $25 billion in 1995. Within its civilian agencies, the Government employed 120,000 information technology workers, and operated 25,000 medium and large mainframe computers and more than two million individual work stations.[1] The Department of Defense has over two million computers, 10,000 local area networks, and 100 long-distance networks. The civilian sector has a critical responsibility to maintain privacy and services for the public using automated data processing and relying on the National Information Infrastructure. Just as critical to the Department of Defense is its ability to carry out any mission that is dependent on information carried on and supported by the NII. If key responsibilities of both the civilian and military sectors of government are heavily dependent upon an unsecured, potentially unavailable Internet, the Government must address whether this reliance on the NII (and GII) is acceptable and, if so, how to manage the risks involved.

Notwithstanding considerable expenditures on information technology, there exists a widening chasm between the security requirements of and the protection provided for unclassified systems government-wide and those applied to the classified

PDF Scrubbing



'This is a stickup . . .'

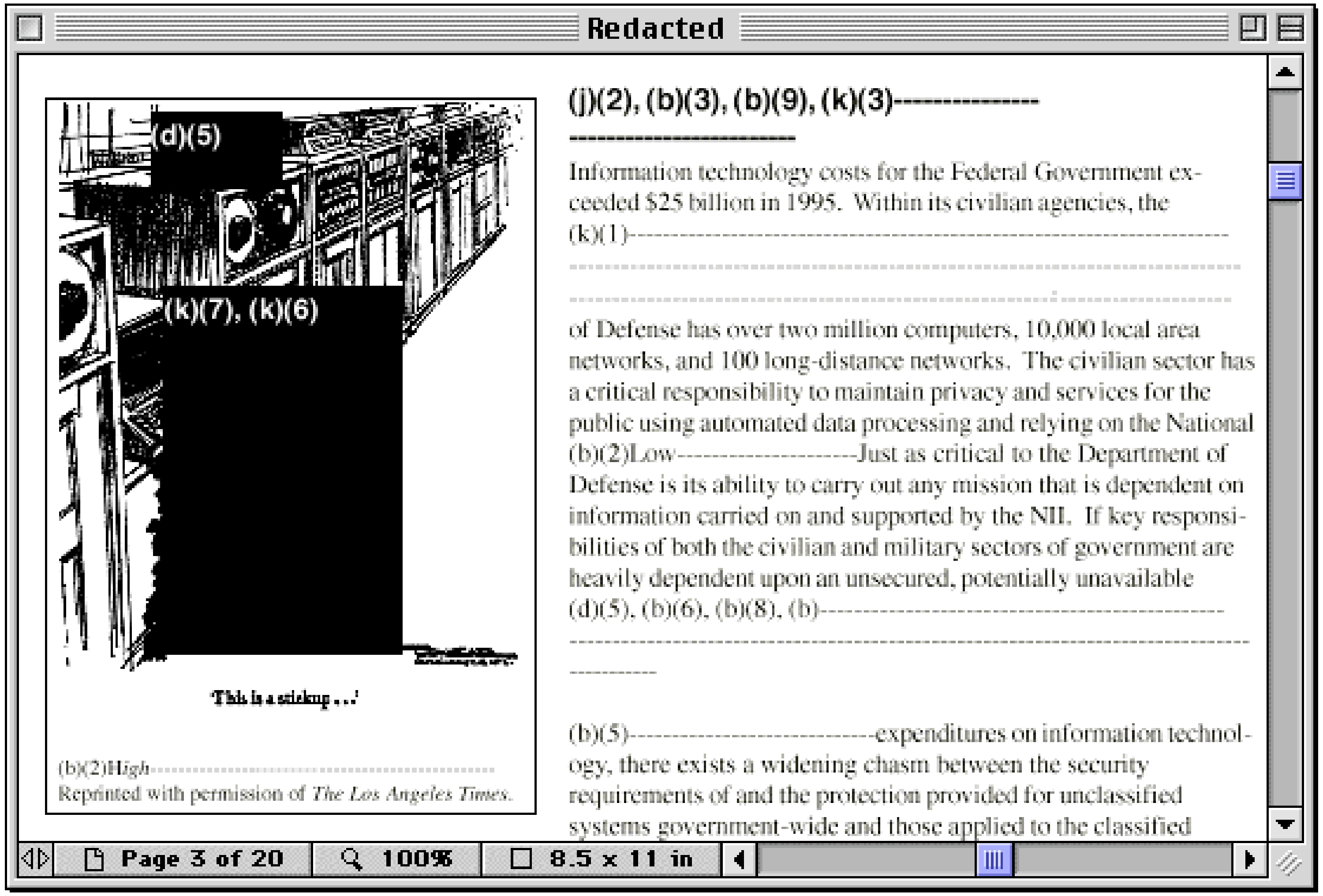
S(b)(2)Highs *Angeles Times*, December 10, 1974.
Reprinted with permission of *The Los Angeles Times*.

T(j)(2), (b)(3), (b)(9), (k)(3)o Information Systems Security

Information technology costs for the Federal Government exceeded $25 billion in 1995. Within its civilian agencies, the ((k)(1)nment employed 120,000 information technology workers, and operated 25,000 medium and large mainframe computers and more than two million individual work stations.[1] The Department of Defense has over two million computers, 10,000 local area networks, and 100 long-distance networks. The civilian sector has a critical responsibility to maintain privacy and services for the public using automated data processing and relying on the National In(b)(2)LowInfrastructure. Just as critical to the Department of Defense is its ability to carry out any mission that is dependent on information carried on and supported by the NII. If key responsibilities of both the civilian and military sectors of government are heavily dependent upon an unsecured, potentially unavailable Ir(d)(5), (b)(6), (b)(8), (b)must address whether this reliance on the NII (and GII) is acceptable and, if so, how to manage the risks involved.

N(b)(5)nstanding considerable expenditures on information technology, there exists a widening chasm between the security requirements of and the protection provided for unclassified systems government-wide and those applied to the classified

PDF Scrubbing



[redacted]

[redacted]

'This is a stickup . . .'

(b)(2)High---

Reprinted with permission of *The Los Angeles Times.*

(j)(2), (b)(3), (b)(9), (k)(3)--

Information technology costs for the Federal Government exceeded $25 billion in 1995. Within its civilian agencies, the (k)(1)--- of Defense has over two million computers, 10,000 local area networks, and 100 long-distance networks. The civilian sector has a critical responsibility to maintain privacy and services for the public using automated data processing and relying on the National (b)(2)Low--------------------Just as critical to the Department of Defense is its ability to carry out any mission that is dependent on information carried on and supported by the NII. If key responsibilities of both the civilian and military sectors of government are heavily dependent upon an unsecured, potentially unavailable (d)(5), (b)(6), (b)(8), (b)---

(b)(5)---------------------------expenditures on information technology, there exists a widening chasm between the security requirements of and the protection provided for unclassified systems government-wide and those applied to the classified

Dave Aitel's umask

- Tries to compare characteristics of text to
  - guess if authors match
  - authors are of the same region etc.
- Minimal documentation