Or how free software could help us
from the legal gray goo..

Alexandre Dulaunoy & Michael G. Noll

GooDiff

"GooDiff is a service for automated tracking of semantic changes in legal documents."

- http://www.goodiff.org/
- running since March 2006
- 1st free software release and datasets at VJ12

(philosophical)

Background

The pessimistic perspective to technology (en)

"112. People anxious to rescue freedom without sacrificing the supposed benefits of technology will suggest naive schemes for some new form of society that would reconcile freedom with technology. **Apart from the fact that people who make suggestions seldom propose any practical means** by which the new form of society could be set up in the first place, it follows from the fourth principle that even if the new form of society could be once established, it either would collapse or would give results very different from those expected."

Industrial Society and Its Future (1995), Theodore Kaczynski

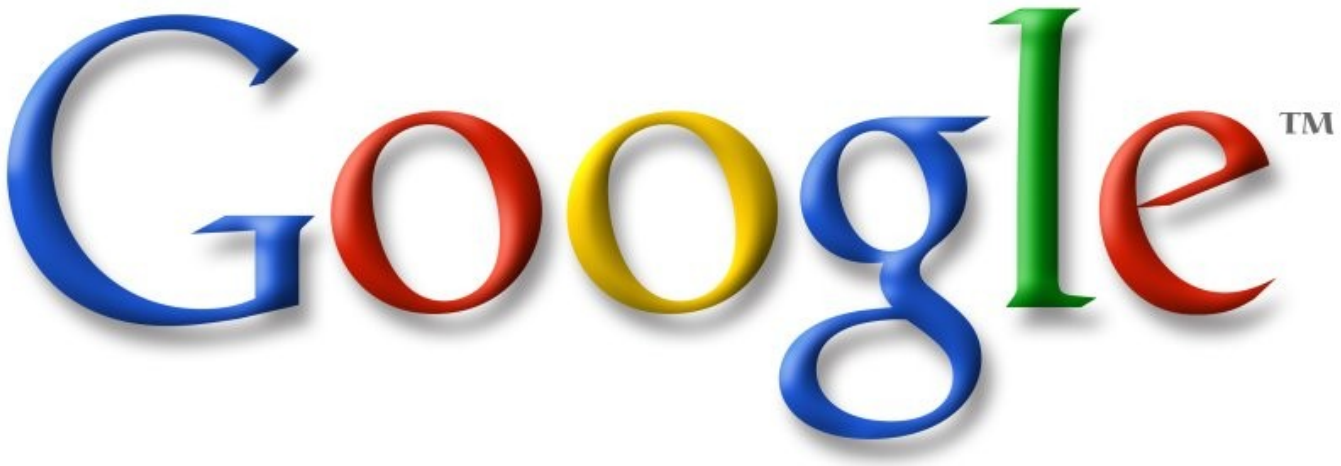
The pessimistic perspective to technology (fr)

"112. Des individus désireux de sauver la liberté sans sacrifier les soi-disant bienfaits de la technologie, ne manqueront pas de proposer des modèles naifs de société où la liberté irait de pair avec la technologie. **Hormis le fait que de telles gens offrent rarement des solutions concrètes permettant de mettre en pratique les idées,** il s'ensuivrait, en vertu du quatrième principe, que même dans le cas où leur nouvelle société pourrait voir le jour, soit elle s'effrondrerait, soit elle aboutirait à des résultats bien différents de ceux attendus."

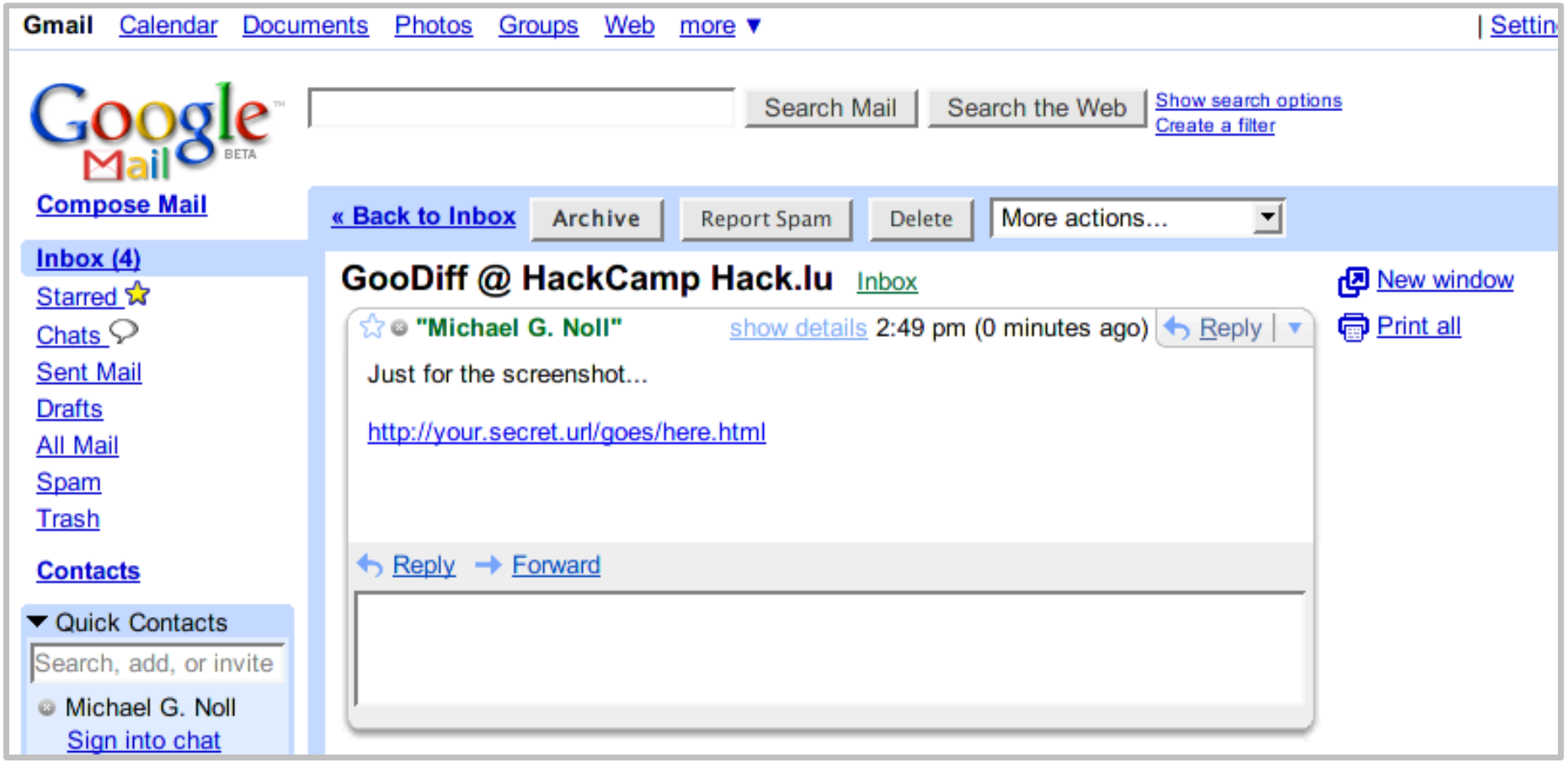
Industrial Society and Its Future (1995), Theodore Kaczynski

(practical)

Background


Google™

Background


Gmail Calendar Documents Photos Groups Web more ▼ | Settin
Google Mail BETA
Search Mail
Search the Web
Show search options
Create a filter
Compose Mail
« Back to Inbox
Archive
Report Spam
Delete
More actions...
Inbox (4)
Starred
Chats
Sent Mail
Drafts
All Mail
Spam
Trash
Contacts
Quick Contacts
Search, add, or invite
Michael G. Noll
Sign into chat
GooDiff @ HackCamp Hack.lu
Inbox
New window
Print all
"Michael G. Noll"
show details 2:49 pm (0 minutes ago)
Reply
Just for the screenshot...
http://your.secret.url/goes/here.html
Reply
Forward

Background (a honeytoken)

- private hyperlink in GMail
- visited by GoogleBot some days later
- no answer from Google

Scientific questions:

- What happened?
- Are they allowed to do that? Are the users informed?

Background

Google Groups, April 2006

http://groups.google.lu/group/Gmail-ABCs/browse_thread/thread/51faf9a3586d2314/2bf9d67b12f64506

Background

- Step #1:
  analyze Google privacy policy etc.
- Step #2:
  changes in the meantime?

= no changelogs

= no archives*

= “you must inform yourself”

Looking around...

- most legal documents (privacy, ToS, etc) have no timestamp, no version, ...
- end user forced to inform himself about changes
- no tool available to help out


GooDiff
+++good
---evil

GooDiff to the rescue

- “Goo” + “Diff” = GooDiff

Components

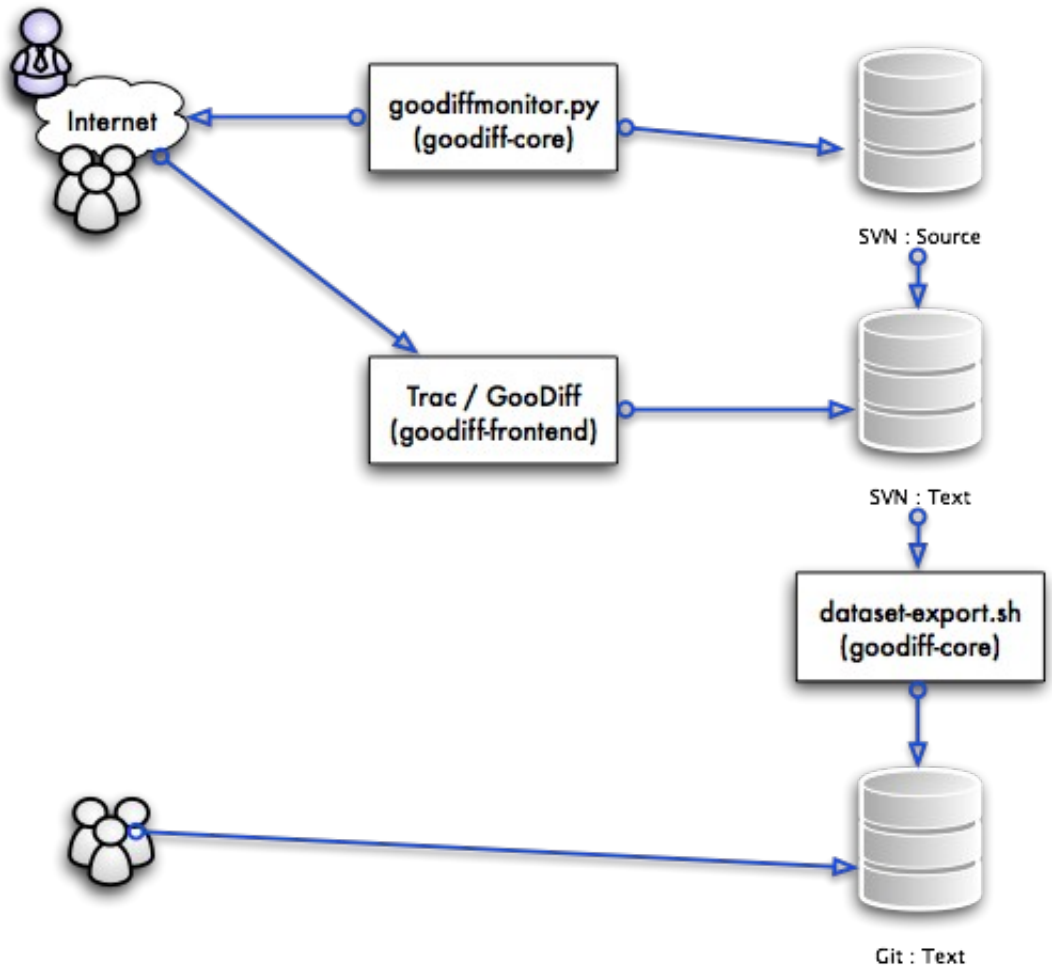
- monitor: fetch documents daily
- revision control: track changes
- UI: browsing, notifications

GooDiff implementation

- Core code rewritten 2 times (and it will be again...) (goodiff-core)

- Python + Subversion + Trac
  - Python
  - SVN 2x, for storing documents
  - Trac GUI front-end (goodiff-frontend)

GooDiff overview



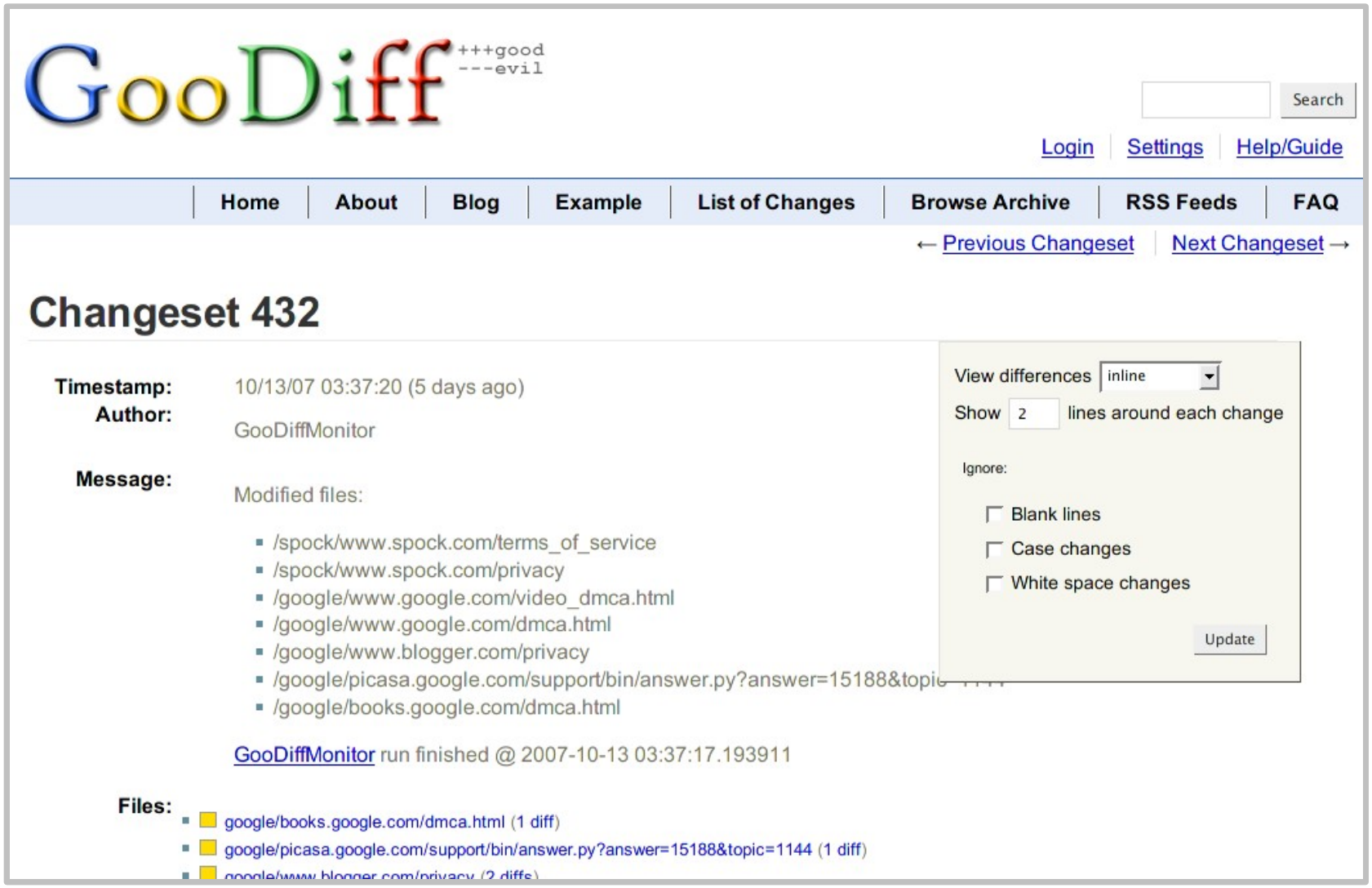
Configuration (sample)

```
<provider name="facebook">
        <service name="facebook">
            <document url="http://www.facebook.com/terms.php">
            <replace pattern='\?pwstdfy=[0-9a-fA-F]{32}' with="?pwstdfy=|REMOVED|" />
          </document>
        </service>
</provider>
```

What we're tracking

+ others...

Examples


GooDiff
+++good
---evil
Search
Login
Settings
Help/Guide
Home
About
Blog
Example
List of Changes
Browse Archive
RSS Feeds
FAQ
← Previous Changeset
Next Changeset →
Changeset 432
Timestamp: 10/13/07 03:37:20 (5 days ago)
Author: GooDiffMonitor
Message: Modified files:
/spock/www.spock.com/terms_of_service
/spock/www.spock.com/privacy
/google/www.google.com/video_dmca.html
/google/www.google.com/dmca.html
/google/www.blogger.com/privacy
/google/picasa.google.com/support/bin/answer.py?answer=15188&topic
/google/books.google.com/dmca.html
GooDiffMonitor run finished @ 2007-10-13 03:37:17.193911
View differences inline
Show 2 lines around each change
Ignore:
Blank lines
Case changes
White space changes
Update
Files:
google/books.google.com/dmca.html (1 diff)
google/picasa.google.com/support/bin/answer.py?answer=15188&topic=1144 (1 diff)

Examples

google/www.google.com/searchhistory/privacy.html

Revision 522		Revision 563	
1	[ ](/)	1	Web History Privacy Notice - Google Web History
2		2	
3	#	3	# [ ](/) Web History
4	** Search History Privacy Notice **		
5		4	
6	[Search History](/searchhistory)	5	* [Web History](/history)
7		6	* [Help](/support/accounts/bin/topic.py?topic=14148)
8	[Help](http://www.google.com/support/accounts/bin/topic.py?topic=10470)	7	* Privacy Policy
9		8	* [Privacy FAQ](privacyfaq.html)
10	** Privacy Policy **		
11			
12	[Privacy FAQ](privacyfaq.html)		
13		9	
14	** Search History Privacy Notice **	10	## Web History Privacy Notice
15		11	
16	November 20, 2006	12	February 22, 2007
17		13	
18	The [Google Privacy Policy](http://www.google.com/privacypolicy.html) describes how we treat personal information when you use Google's products and services, including information provided when you use Search History. In addition, the following describes our privacy practices that are specific to Search History.	14	The [Google Privacy Policy](/privacypolicy.html) describes how we treat personal information when you use Google's products and services, including information provided when you use Web History. In addition, the following describes our privacy practices that are specific to Web History.
19		15	

Examples

- Google Picasa (changeset 28, 30-Mar-06)
- If you send a request to Google's servers, we record standard log information, including Internet Protocol addresses and information related to your request. We will ask before collecting personally identifying information from you and give you an opportunity to opt out.

Examples

- Google Picasa (changeset 28, 30-Mar-06)
- If you send a request to Google's servers, we record standard log information, including Internet Protocol addresses and information related to your request. We will ask before collecting personally identifying information from you and give you an opportunity to opt out.
- ...and information related to your request. **We also log information about the installation process when you download Picasa. We also log information about the installation process and your system and settings when you download Picasa.** We will ask before ...

Examples

- Blogger.com (changeset 432, 13-Oct-07)
- Google-Server zeichnen automatisch Daten zu Ihrer Verwendung des Service auf, z. B. dazu, wann Sie Blogger verwenden und die Haufigkeit sowie die Grosse von Datentransfers. Informationen, die auf der Blogger-Oberflache angezeigt werden oder auf die geklickt wird (z. B. UI-Elemente, Einstellungen und andere Informationen), werden ebenfalls aufgezeichnet.

Examples

- Blogger.com (changeset 432, 13-Oct-07)
- Google-Server zeichnen automatisch Daten zu Ihrer Verwendung des Service auf, z. B. dazu, wann Sie Blogger verwenden und die Haufigkeit sowie die Grosse von Datentransfers. Informationen, die auf der Blogger-Oberflache angezeigt werden oder auf die geklickt wird (z. B. UI-Elemente, Einstellungen und andere Informationen), werden ebenfalls aufgezeichnet.
- ...werden ebenfalls aufgezeichnet. **If you are logged in we may associate that information with your account.**

Examples

- Apple.com / privacy (changeset 568, Sept. 2009)
- Pixel tags also enable us to send email messages in a format customers can read. And they tell us whether emails have been opened to ensure that we're sending only messages that are of interest to our customers. We store all of this information in a secure database located in Cupertino, California, in the United States.
- … to our customers. **We may use this information to reduce or eliminate messages sent to a customer.** We store all..

Examples

- Google Talk (changeset 563, July, 14 2009)
- ...for more information.)
- ...for more information.) *** _ SMS. _ When you send and receive SMS messages to or from Google Talk, we collect and maintain information associated with those messages, such as the phone number, the wireless carrier associated with the phone number, the content of the message, and the date and time of the transaction.**

Examples of interesting updates

- Apple
  - Addition of iPhone to the iTunes portfolio
- Google Finance
  - Real-time stock prices instead of X mins delay
- Google “Search History”- changeset 563
  - Is now called “Web History”

Statistics & comments

- Tracking stuff since ~ 36 months
- #changesets: 570 (raw), ~360 (real)
  - changeset: >=1 changed documents
- Documents change more often than we would have expected
- Timestamps IN documents (if present) are often not matching the time when we track a change

Theory vs. Practice

- How to handle 404, 301, ...?
  - “YouTube will be undergoing scheduled maintenance, starting around 7:00 pm PDT.“
- Dynamic content:
  ads, web bugs, session IDs, ...
- How to rebuild DB from scratch while preserving change time etc.?
- Service providers don't like us

GooDiff is not a robot/crawler

“A robot is a program that automatically traverses the Web's hypertext structure by retrieving a document, and recursively retrieving all documents that are referenced. “

- GooDiff Monitor is only fetching a defined set of URLs
- The set of URLs is defined by human users
- The main implication for us : we don't have to follow robots.txt files

If you want to help...

- http://www.gitorious.org/goodiff/
- More analysis of the datasets
- Human feedback is required for finding semantic changes
- Fork it (and make it more useful)
- For feedback, ideas :
  - info@goodiff.org

Future activity (based on hackerspace.be discussion)

- Lawyer2text – can this be automated?
- Timeline representation of documents changes
  - Along with company/service provider events
- Words/terms detection to avoid displaying non-semantic changes
- Collaborative evaluation of change set

(philosophical)

Closing

“The crisis can be solved only if we learn to invert the present deep structure of tools; **if we give people tools that guarantee their right to work with high, independent efficiency, thus simultaneously eliminating the need for either slaves or masters and enhancing each person’s range of freedom.** People need new tools to work with rather than tools that “work” for them. They need technology to make the most of the energy and imagination each has, rather than more well-programmed energy slaves.”

Tools for Conviviality (1973) - Ivan Illich

“**Une société conviviale est une société qui donne à l'homme la possibilité d'exercer l'action la plus autonome et la plus créative, à l'aide d'outils moins contrôlables par autrui**.”

La reconstruction conviviale (1973) - Ivan Illich

“**Une société conviviale est une société qui donne à l'homme la possibilité d'exercer l'action la plus autonome et la plus créative, à l'aide d'outils moins contrôlables par autrui**.”

La reconstruction conviviale (1973) - Ivan Illich

Bibliography