using the web of data
the user

Alan Dix
InfoLab21, Lancaster University, UK
www.hcibook.com/alan/papers/WOD-PD-2008/
www.alandix.com/blog

in the end
it’s not about data
it’s about use
user and the web of data
two sides

• generating the web of data

• accessing the web of data

and maybe linked through user interaction

the ecology of the web

on the web

on the desktop

standards: ⭐ structured

openness: ⛔ closed

sharedness: 😊 private

ontology based ⭐ SemWeb

local data

desktop apps

web data

web apps

browser

web services
generating the web of data

top down
bottom up

top down

• from legacy sources
taking existing online data
  – from structured data, e.g. SQL -> RDF
  – web scraping
e.g. Google Social Graph API

• from scratch
  – building new RDF stores for new data
  – user contributed
top down – user creation

creating public data
  – Freebase
    • ontology based
  – Swivel
    • tables & graphs

• maybe sharing more private data
  • Google spreadsheets
  • linking desktop ontologies

bottom up

• implicit
  – folksonomies => structured data *

• explicit
  – personal ontologies

• automatic
  – semantic desktop – PIM mining

personal ontologies

- use ‘general’ categories:
  - post code, name, place

- linking to personal ontology
  - users own entities and categories
    - egocentric & ideocentric

- how to build?
  - by hand (during useful interactions)
  - automatically (mining files, emails, etc.)
    - e.g. Gnowsis, Neomuk and other semantic desktop projects

user data collection

- private
  - personal ontologies
  - semantic desktop

- public
  - freebase
  - folksonomies critical mass
accessing the web of data

- exposing the WoD
- users own WoD
- hiding the WoD
- linking to the WoD
### Visualising Web of Data

<table>
<thead>
<tr>
<th>Link to WoD</th>
<th>Implemented using WoD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Traditional apps</th>
<th>e.g. Facebook</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td></td>
</tr>
<tr>
<td></td>
<td>YES</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User sees a WoD</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hiding WoD</td>
<td></td>
</tr>
<tr>
<td>Exposing WoD</td>
<td></td>
</tr>
</tbody>
</table>

### Hiding vs. Exposing

Similar issues for RDMS

- **Exposing**
  - MySQLAdmin
  - mainly for administrators/developers

- **Hiding**
  - most DB apps
  - for end users
exposing the web of data

user sees a WoD

<table>
<thead>
<tr>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

implemented using WoD

YES

full WoD

exposing WoD

exposing the web of data tabulator

• outliner style + query list views + editing

exposing the web of data

**tabulator**

- ouliner style + **query list views + editing**

exposing the web of data

tabulator

- **exposing the web of data graph ...**
- visualising graph ... example

exposing the web of data
foaf explorer
• drill-down / link style
• part bespoke

http://xml.mfd-consult.dk/foaf/explorer/

exposing the web of data
mspace
• faceted browsing

http://www.iam.ecs.soton.ac.uk/projects/292.html
http://mspace.fm/
hiding the web of data

users don’t want data
they want to do things
hiding the web of data

Cenote

- dedicated view for books
- draws together several linked data sources
- semantic web structure not apparent to user

http://cenote.talis.com/

user’s own web of data

<table>
<thead>
<tr>
<th>implemented using WoD</th>
<th>user sees a WoD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>e.g. FaceBook</td>
</tr>
<tr>
<td></td>
<td>personal WoD</td>
</tr>
</tbody>
</table>

user’s own WoD
user’s own web of data

- FaceBook
  - apps add relationships
  - but insulated from one another
  - NOT RDF
- Developer API
  - add own relationships

linking to the web of data

<table>
<thead>
<tr>
<th>link to WoD</th>
<th>implemented using WoD</th>
<th>user sees a WoD</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

traditional apps
linking to the web of data

some rich semantic web services ...
... but most web pages and apps are not!

can we make them loci of entry to WoD?
  • at point of creation:
    – microformats, zLinks
  • at point of use:
    – data detectors and intelligent stuff

linking at point of creation

offer alternative RDF or XML formats
or use markup in web page
  – dedicated markup
    e.g. zLinks
  – microformats
    human readable text ...
    ... but also machine readable
    for search engines or plug-ins

<p class="vcard">Hi, my name is <span class="fn">Jamie Jones</span> and I dig microformats!</p>
linking at point of use
Snip!t 🌸

1. University
2. 1/2

linking at point of use
Snip!t 🌸
class of systems ‘data detectors’

- late 1990s
  - Intel selection recognition agent
  - Apple Data Detectors (Bonnie Nardi)
  - CyberDesk (Andy Wood led to onCue)
- recently
  - Microsoft SmartTags
  - Google extensions
  - Citrine – clipboard converter
  - CREO system (Faaberg, 2006)
- way back
  - Microcosm (Hypertext external linkage)

syntactic / regexp

'Semantic' / lookup

SnipIt uses combination

do it yourself!

cf. Calais yesterday

- tellmeabout web service
  - XML version of Snip!t detection and actions
- Tag the Net
  - used in WordPress

http://tagthe.net/api/?text=Alan+Dix+is+at+Lancaster+not+London-on-29-Oct-2008

<?xml version="1.0" encoding="UTF-8"?>
<memes>
  <dim type="topic">London</dim>
  <dim type="person">Alan Dix</dim>
  <dim type="location">Lancaster</dim>
  <dim type="language">english</dim>
</memes>
linking in the hidden web

hidden web / deep web
- searchable data bases etc.
- 90% of web accessible data not searchable

add meta-information
- internal (page markup) or external (e.g. Snipit)

reweaving the hidden web through user interaction

web of data static and dynamic
meta-data describing input
hidden web application
web pages with microformats or data detectors
from use to data
Amazon don’t ask people what they like
... people just buy what they like

action is data

from use to data

using interaction to generate semantics
• selection:
  – user selects data and uses it in semantic field
• confirmation
  – if user uses inferred data assume correct
• web forms
  – type annotation from use
context in forms

but what is the relationship?
maybe semantic markup on form
   – good SemWeb style ... but rare
   ... or more inference ...

context in forms - inference

match terms in form to ontology
look for ‘least cost paths’
   • number of relationships traversed, fan-out

context in forms - inference

match terms in form to ontology
look for ‘least cost paths’
  • number of relationships traversed, fan-out
later suggest based on rules