wot I done in AI

Alan Dix
note ...

• using ‘intelligent’ to include
  - traditional artificial intelligence
  - statistical methods
  - information retrieval techniques
  - neural nets
  - genetic algorithms
  - simple heuristics
‘intelligent’ things …

- submarine design
  - use genetic algorithms to solve constraints
- database interface: Query-by-Browsing
  - use decision trees to infer queries
- intelligent web interfaces
  - simple pattern matching + blackboard
  + appropriate interaction
- typing error analysis (for dyslexia)
  - iterative deepening search (variant of depth first) to find most likely errors
Query by Browsing

- user chooses records of interest
  - ✓ tick for those wanted
  - ✗ cross for those not wanted

- system infers query
  - web version uses rule induction
  - variant of Quinlan’s ID3

www.meandeviation.com/qbb
Query by Browsing
what it looks like

user asks
system to
make a query

system infers
SQL query

query results
highlighted
Query by Browsing

dual representation

query (intensional) for precision
listing (extensional) for understanding
Query by Browsing

how it works

examples

(machine learning)

SQL query

(decision tree)

cond

cond

SELECT * FROM qbb_ex1 WHERE Wage >= 15000
Query by Browsing keywords

• knowledge representation
  - uniform attribute-value data
  - simple values (numbers, strings)

• rule representation
  - decision tree

• algorithm
  - variant of ID3 - inductive learning
intelligent user interfaces

• have had a bad history!
• why?
  - often focused on clever techniques
  - forget wider interaction
e.g. intelligent menus:
  - monitor command use
  - reorder so most frequent on top 😊
  - but order keeps changing 😞
appropriate intelligence

- often simple heuristics
- combined with the right interaction
rules of standard AI interfaces

1. it should be right as often as possible

2. when it is right it should be good

good for demos
look how clever it is!
rules of appropriate intelligence

1. it should be right as often as possible

2. when it is right it should be good

3. when it isn’t right …
   it shouldn’t mess you up

} what makes a system really work!
Hit or a Miss?

✗ paper clip
- can be good when it works
- but interrupts you if it is wrong

✓ Excel ‘∑’ button
- guesses range to add up
- very simple rules
  (contiguous numbers above/to left)
- if it is wrong ...
  simply select what you would have anyway
... for menus

- small number of most popular at top quick when it gets it right

- alphabetic below still easy to scan when it isn’t

<table>
<thead>
<tr>
<th>Boxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiona</td>
</tr>
<tr>
<td>Miriam</td>
</tr>
<tr>
<td>Esther</td>
</tr>
<tr>
<td>Adrian</td>
</tr>
<tr>
<td>Andy</td>
</tr>
<tr>
<td>Brian</td>
</tr>
<tr>
<td>Charlotte</td>
</tr>
<tr>
<td>Colin</td>
</tr>
<tr>
<td>David</td>
</tr>
<tr>
<td>Devina</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>
designing appropriate intelligence

onCue

- intelligent toolbar
- sits at side of the screen
- watches clipboard for cut/copy
- suggests useful things to do with copied date
onCue in action

user selects text

and copies it to clipboard

slowly icons fade in

the dancing histograms are very useful at

looking out some of the textile sites you

x’s page at http://www.hiraeth.com/
onCue appropriate?

1. it should be right as often as possible
   - uses simple heuristics:
     e.g. words with capitals = name/title

2. when it is right it should be good
   - suggests useful web/desktop resources

3. when it isn’t right it shouldn’t mess you up
   - slow fade-in means doesn’t interrupt
### Kinds of Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>short text</td>
<td>search engines</td>
</tr>
<tr>
<td>single word</td>
<td>thesaurus, spell check</td>
</tr>
<tr>
<td>names</td>
<td>directory services</td>
</tr>
<tr>
<td>post codes</td>
<td>maps, local info</td>
</tr>
<tr>
<td>numbers</td>
<td>SumIt! (add them up)</td>
</tr>
<tr>
<td>custom</td>
<td>order #, cust ref ...</td>
</tr>
<tr>
<td>tables</td>
<td>...</td>
</tr>
</tbody>
</table>
how it works

sensor watching system clipboard

recognisers that determine type of data

words recogniser

single word recogniser

table recogniser

number list recogniser

sort of ‘blackboard’ (but not quite)

services that do things (mainly web-based)

services

histogram

thesaurus, dictionary

Excel Qbit

CD Rom Encyclopedia

onCue framework

various searches

browser Qbit

onCue toolbar (UI)

onCue framework

aQtiveSpace

lower level component infrastructure
how it works

1. User copies text

2. Clipboard watcher puts text on 'blackboard'

3. Recognisers that deal with raw text fire

 recognise

words recogniser

single word recogniser

table recogniser

number list recogniser

Text: "histograms"

onCue framework

aQtiveSpace
how it works

4 one recogniser sees text is words and adds to ‘blackboard’

5 further recognisers fire

6 services based on some ‘words’ become active
how it works

7 recogniser sees it is ‘single ‘word’ and adds to ‘blackboard’

8 more services activate for ‘single word’

9 active services appear in onCue toolbar
onCue

keywords

• knowledge representation
  - heterogeneous attribute-value / frame

• rule representation
  - regular expressions
    sort of production rule
    hand-crafted heuristics

• algorithm
  - multi-agent blackboard, emergent behaviour
    pattern matching
    text mining
Snip!†
web version of onCue
Snip!t

1. Users select in web page and presses “Snip!t” bookmarklet.

2. Snip!t pops up page with suggested things to do with the snip (and saves it for later, like bookmark).
SnipIt

recognises various things e.g. dates

ask for demo later

www.snipit.org
Kinds of recogniser:

syntactic - regular expression / patterns / structure
  e.g. post code, email address

lexicon - large look up tables
  e.g. countries, internet TLD

hybrid - popular first names => full name

+context - telephone number ... what country??
architecture

- server-side ‘intelligence’
- recognisers + services again
- different kinds of recogniser chaining:
  - from semantics to wider representation  
    e.g. postcode suggests look for address
  - from semantic to semantic  
    e.g. domain name in URL
  - from semantic to inner representation  
    e.g. from Amazon author URL to author name
provenance

when you have a recognised term:

• where did it come from
  - text char pos 53-67
  - transformed from Amazon book URL

• how confident are you
  - 99% certain Abraham Lincoln is a person

• how important
  - mother-in-law’s birthday