Beyond text searching in patents: Review of PatMedia and its potential for improving retrieval for mechanical and device searchers.

Jane List, Technical Product Manager, Dialog.
IRF Symposium, Vienna, June 4th
Overview

Introductions - Dialog, ITI PatMedia, Patents, text and image searching

An example of figure and table extraction from industry

Patmedia
image search, filtering, figure extraction & categorisation

some test searches, results and recommendations

Conclusions
invention representation for the future
Towards content-based patent image retrieval: A framework perspective.


http://mklab-services.iti.gr/PatMedia

130,000 images
6,200 optical recording EP patents
Searchable elements in a patent specification

Front Page
--Bibliographic information
--Abstract

Full Specification
--Background to the Invention
--Summary of the Invention
  --Brief Description of the Drawings
--Detailed Description of the Invention
--Claims

‘Representative’ (clipped) image

Figures
Why do we need to analyse all the drawings?

- drawing pages may be half content of patent specification, typically 5-10 Figures
- same front page image may be used for many patents
- prior art may even be used for ‘representative’ images
- design patents, some utility models need graphical representation

Mechanical patent Searching a moving target, Demarco, D, PIUG 2010
Fig. 3. Time trend of nutrient enrichment into Laholm Bay in southwestern Sweden from 1950 until 1988. Phosphorus loads decreased after 1970 in response to management for eutrophication, while nitrogen loads, which were unregulated, continued to increase. Early evidence of eutrophication in the Bay appeared in the early 1970s with filamentous green algae and plankton blooms.

Sentence referring to the figure or table

A fourth whole-ecosystem study explored long-term changes in Laholm Bay, an estuary on the southwestern coast of Sweden (Fig. 3). Early signs of eutrophication appeared...
Example of Extraction of Key Terms from Figures and/or Tables (scanning OCR)

raw terms from the figure:
- N kg yr$^{-1}$
- nitrogen
- Phosphorus
- Early indications of eutrophication
- Filamentous green algae
- Exceptional plankton blooms
- P kg yr$^{-1}$
- Year

raw terms from the table:
- Basin
- Choptank
- Patuxent
- Land area (km$^2$)
- % Land Cover
- Agri.
- Urban
- Inter. Wet.
- Pop. Den. (no. km$^{-2}$)
- Sewage (L d$^{-1}$)
- Estuary area
- Land/water

<table>
<thead>
<tr>
<th>Basin</th>
<th>Land area (km$^2$)</th>
<th>% Land Cover</th>
<th>Inter. wet.</th>
<th>Pop. den. (No. km$^{-2}$)</th>
<th>Sewage (10$^6$ L d$^{-1}$)</th>
<th>Estuary area (km$^2$)</th>
<th>Land/water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choptank</td>
<td>1,756</td>
<td>30.9</td>
<td>61.5</td>
<td>5.4</td>
<td>22.7</td>
<td>301</td>
<td>5.83</td>
</tr>
<tr>
<td>Patuxent</td>
<td>2,260</td>
<td>63.5</td>
<td>20.3</td>
<td>15.7</td>
<td>235</td>
<td>137</td>
<td>16.5</td>
</tr>
</tbody>
</table>
Overview of PatMedia

• Search Images:
  • upload an example to find similar drawings

Refine results
• Filter on Keywords
  • from the description in specification
• Filter on 10 Categories
Three test image searches to find similar figures

- upload a figure known to be in the database
  - result 36 images
  - 4 close visually, no exact match

- upload a simple shape
  - Results no similar

- upload a time chart
  - result 3 images
  - All visually similar no exact match
Sample result screen

**Hits**
- **Category**
- **Identifier**
PatMedia Review - figure extraction and categorisation

Figure, Diagram, WaveForm, Block Diagram, FlowChart, Graph, Circuit, Technical Drawing, TimeChart,
• Unknown

- Improve categorisation of figures
  - too many Unknown,
  - Figure vs. Diagram?
  - Use text from both ‘Description of Drawings’ section and Specification

- Improve figure separation programme
  - EP224935 has 5 Figures
  - PatMedia shows 11. Figure 5 waveform progression; but two waveforms missing.
PatMedia Review - image search

• Enable categorisation of figures by subject pre or post search
  ➢ Optical (lens), electrical drawings
  ➢ prior art?

➢ Consider orientation, scale, dimensions
➢ improve exact match
Back to the future

patent models – required in US from 1790 to 1870

Blender.org
CAD
solidworks

History on a small scale, Toylike models, required for patents a century ago, reveal a vanished past
By Corydon Ireland, Harvard News Office, Thursday, July 2, 2009
Thank you

Jane.list@dialog.com

Stefanos@iti.gr