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

Dialog

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

An example of figure and table extraction from industry

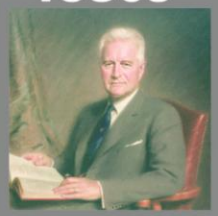
Patmedia
image search , filtering , figure extraction & categorisation

some test searches, results and recommendations

Conclusions
invention representation for the future

DIALOG & PROQUEST
OVER 70 YEARS OF INDUSTRY LEADERSHIP

1930s

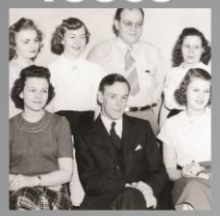


Eugene B. Power founds UMI® in 1938, which later evolves into ProQuest. Early English Books are filmed before WWII threatens the British Library.

1940s

Newspapers in microfilm program begins including *The New York Times*

1950s



UMI becomes the Association of Research Libraries publisher of record for U.S. dissertations in 1951.

Cambridge Scientific Abstracts (CSA) is founded during the 1950s

1960s

UMI is acquired by Xerox in 1961. Roger Summit founds Dialog as Lockheed NASA project in 1968.

1970s

Dialog is launched in 1972

After commercial launch in 1972, Dialog receives "Product of the Year" award from the Information Industry Association (IIA).

1980s

DataStar is launched in 1981.

CSA launches CD-ROM products in 1985

CSA is the first publisher to offer scientific data on CD-ROM from the National Library of Medicine's Medline.

Bell & Howell acquires UMI in 1985

The first full-image periodical database on CD-ROM is produced.

In 1986, Dialog OneSearch is launched and receives IIA "Product of the Year" award in 1987. The first Dialog OnDisc product is launched.

Dialog becomes part of Knight-Ridder in 1988

Knight-Ridder VuText newspaper collection is integrated into Dialog.

1990s

Dialog & DataStar Web interfaces are introduced in 1996.

ProQuest launches ProQuest Direct on the Web in 1996

In 1997, CSA Web Resources Database of scholarly content is introduced. UMI becomes offsite repository of Digital Dissertations for the Library of Congress in 1998.

2000s

The Washington Post

President Refuses to Turn Over Tapes; Ervin Committee, Cox Issue Subpoenas

In 2001, ProQuest Historical Newspapers program is introduced, and CSA acquires Bowker from Reed Elsevier. In 2002, RefWorks was founded. In 2003, ProQuest launches Electronic Dissertation program.

CIG acquires ProQuest and merges with CSA in 2007

In 2004, ProQuest acquires SerialsSolutions, adds Scholar Universe, and launches Smart Search. In 2006, ProQuest introduces COS Research Support Services and One Click™ Searching.

In 2008, ProQuest acquires RefWorks and Dialog.

Dialog®

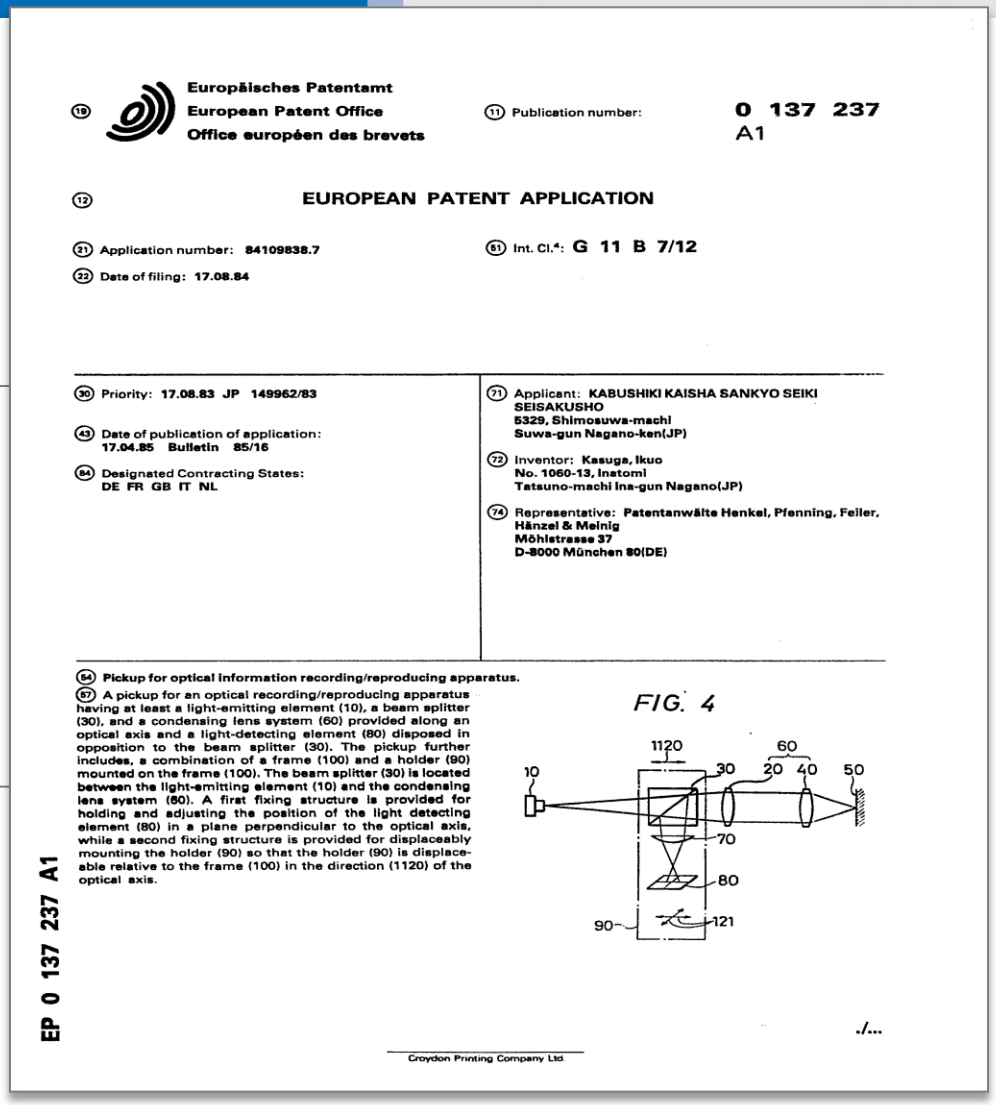
Authoritative answers enriched by ProQuest.

Image Search Project: PatMedia

INFORMATICS & TELEMATICS INSTITUTE
centre for research and technology - hellas

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

130,000 images
6,200 optical recording EP patents



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 137 237 A1

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 84109838.7

(22) Date of filing: 17.08.84

(51) Int. Cl.⁴: G 11 B 7/12

(30) Priority: 17.08.83 JP 149962/83

(43) Date of publication of application: 17.04.85 Bulletin 85/16

(84) Designated Contracting States: DE FR GB IT NL

(71) Applicant: KABUSHIKI KAISHA SANKYO SEIKI SEISAKUSHO
5329, Shimosuwa-machi
Suwa-gun Nagano-ken(JP)

(72) Inventor: Kasuga, Ikuo
No. 1060-13, Inatomi
Tatsuno-machi Ina-gun Nagano(JP)

(74) Representative: Patentanwälte Henkel, Pfenning, Feiler, Hänzel & Meinig
Möhlstrasse 37
D-8000 München 80(DE)

(54) Pickup for optical information recording/reproducing apparatus.

(57) A pickup for an optical recording/reproducing apparatus having at least a light-emitting element (10), a beam splitter (30), and a condensing lens system (60) provided along an optical axis and a light-detecting element (80) disposed in opposition to the beam splitter (30). The pickup further includes, a combination of a frame (100) and a holder (90) mounted on the frame (100). The beam splitter (30) is located between the light-emitting element (10) and the condensing lens system (60). A first fixing structure is provided for holding and adjusting the position of the light detecting element (80) in a plane perpendicular to the optical axis, while a second fixing structure is provided for displaceably mounting the holder (90) so that the holder (90) is displaceable relative to the frame (100) in the direction (1120) of the optical axis.

FIG. 4

EP 0 137 237 A1

./...

Croydon Printing Company Ltd.

Towards content-based patent image retrieval: A framework perspective. *Vrochidis S, Papadopoulos S, et al.* World Patent Information, 2009.

Searchable elements in a patent specification

1973

Front Page

--Bibliographic information

--Abstract

1994

Full Specification

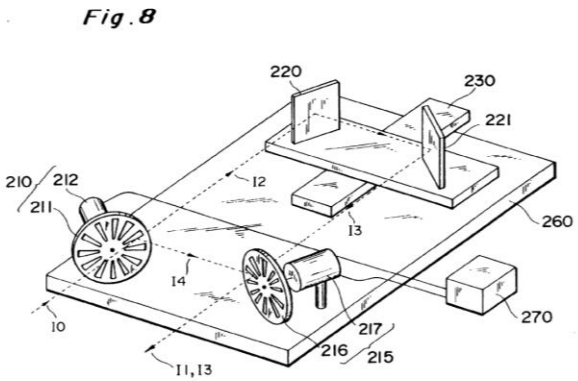
--Background to the Invention

--Summary of the Invention

--*Brief Description of the Drawings*

--Detailed Description of the Invention

--Claims

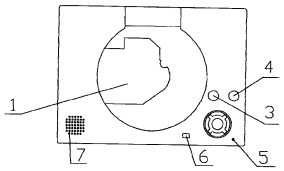


2000

'Representative' (clipped) image

2013?

Figures



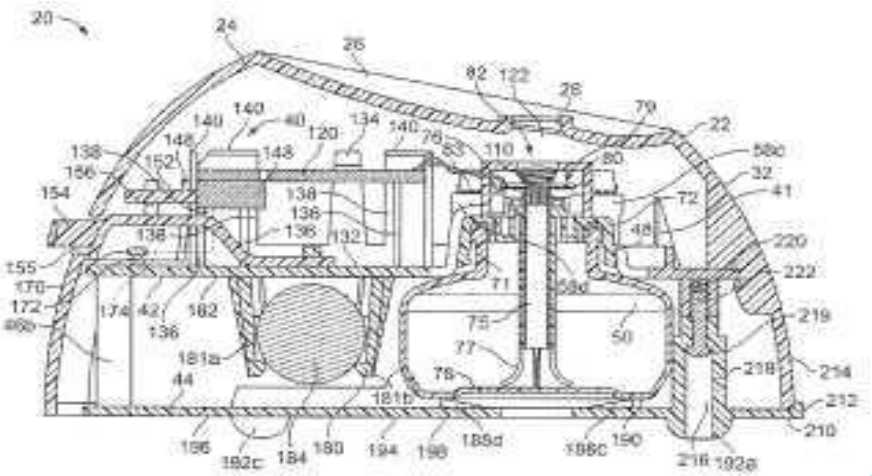
Why do we need to analyse all the drawings?

- 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

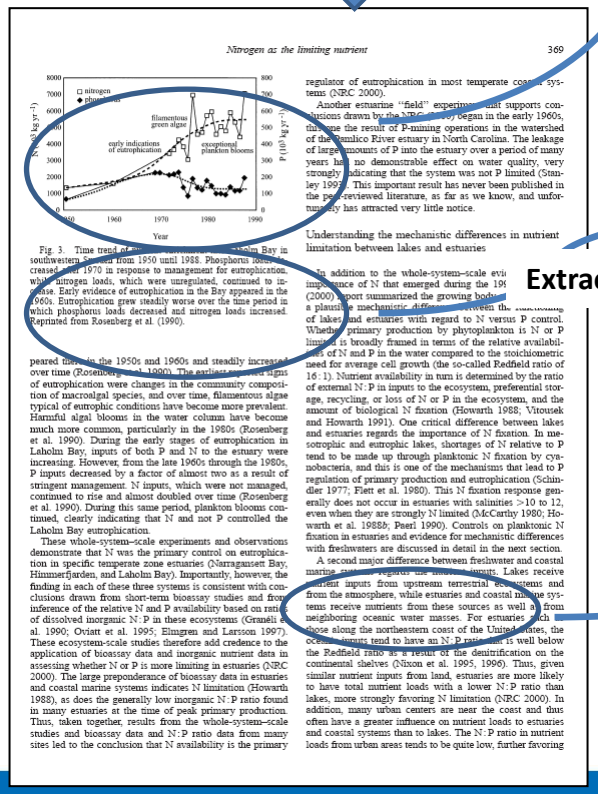
- design patents, some utility models need graphical representation



Mechanical patent Searching a moving target, Demarco, D , PIUG 2010

ProQuest Deep Indexing Process

Acquire Fulltext



Nitrogen as the limiting nutrient 369

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 1960s. Eutrophication grew steadily worse over the time period in which phosphorus loads decreased and nitrogen loads increased. Reprinted from Rosenberg et al. (1990).

Extraction

Extraction

Extraction

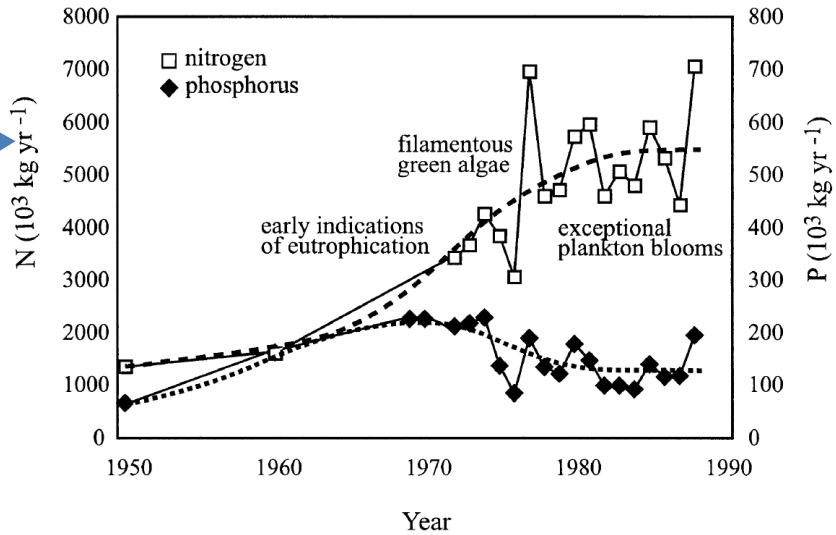


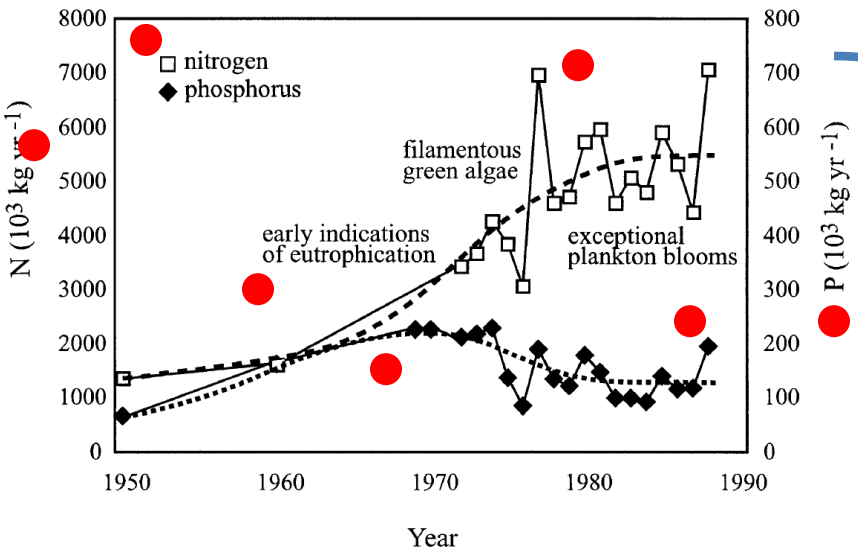
Figure or Table caption

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

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 ap-

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

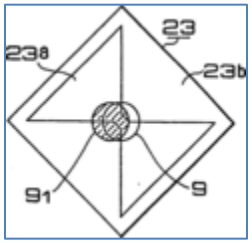
- Basin
- Choptank
- Patuxent
- Land area (km2)
- % Land Cover
- Agri.
- Urban
- Inter. Wet.
- Pop. Den. (no. km-2)
- Sewage (L d-1)
- Estuary area
- Land/water

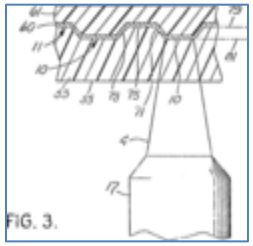
raw terms from the table:

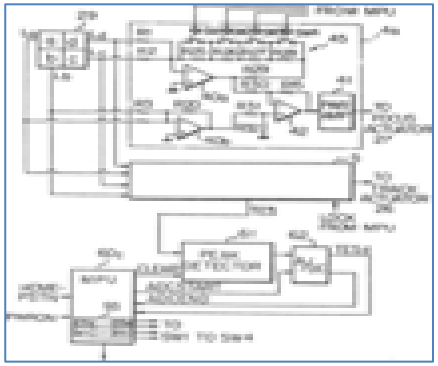
Basin	Land area (km^2)	% land cover				Pop. den. (No. km^{-2})	Sewage (10^6 L d^{-1})	Estuary area (km^2)	Land/water
		Forest	Agri.	Urban	Inter. wet.				
Choptank	1,756	30.9	61.5	5.4	2.2	59	22.7	301	5.83
Patuxent	2,260	63.5	20.3	15.7	0.4	262	235	137	16.5

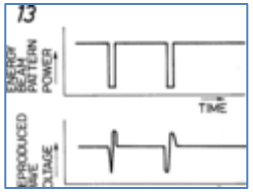
Overview of PatMedia

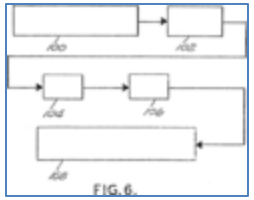
- Search Images:
  - upload an example to find similar drawings

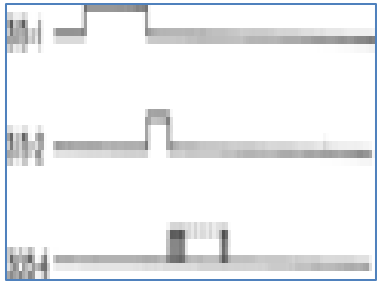










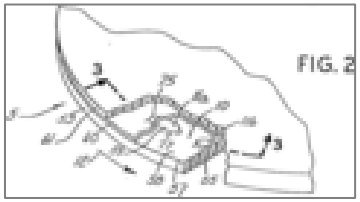


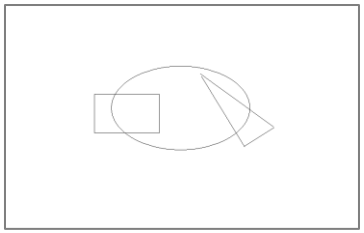
Refine results

- Filter on Keywords
  - from the description in specification
- Filter on 10 Categories

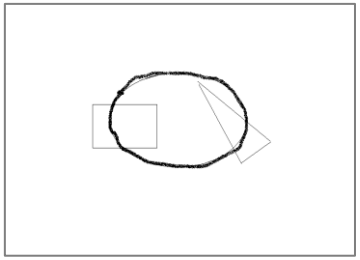
Three test image searches to find similar figures

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

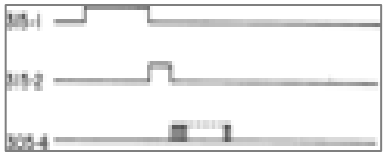




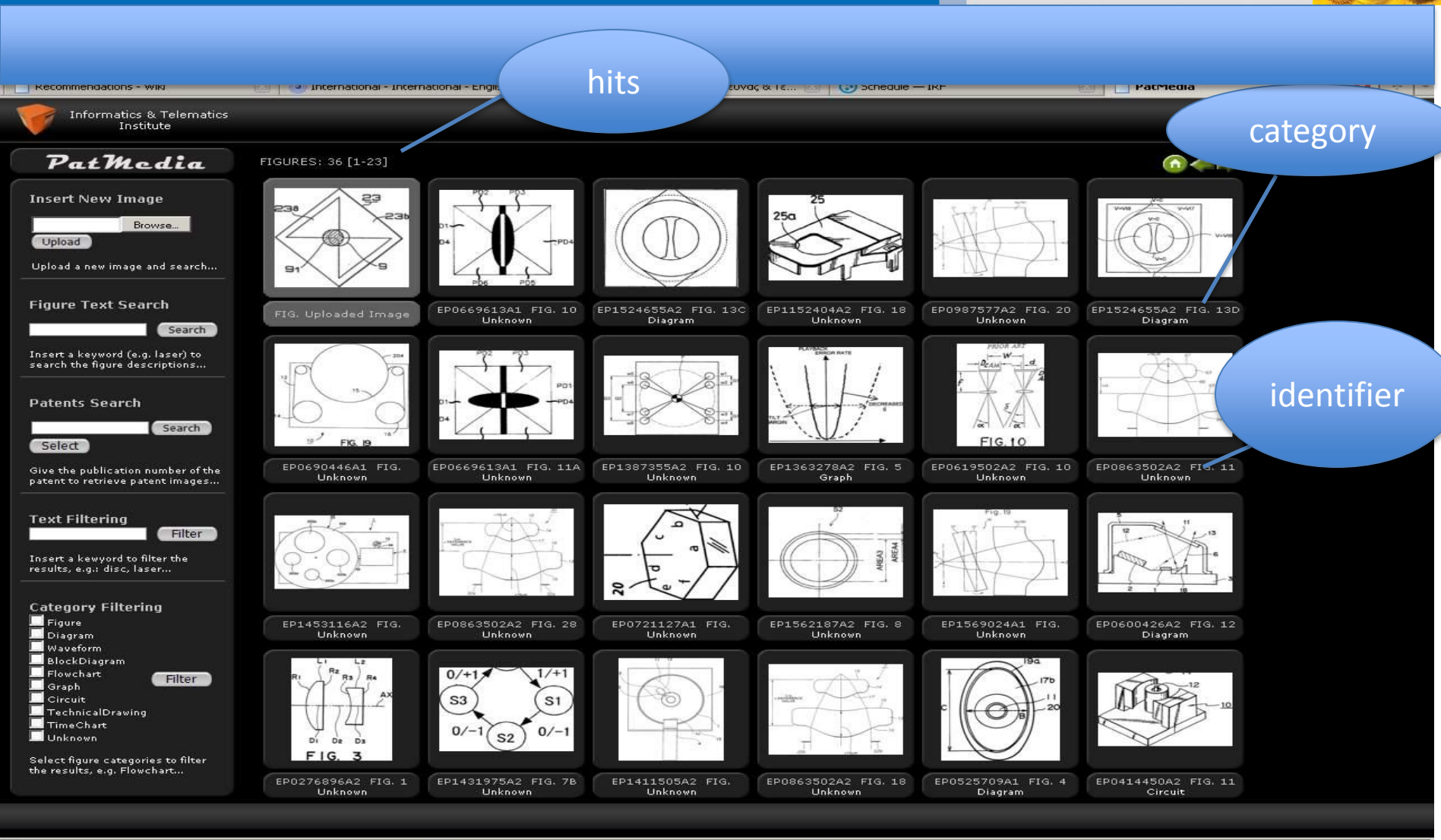
- upload a simple shape
  - Results no similar



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



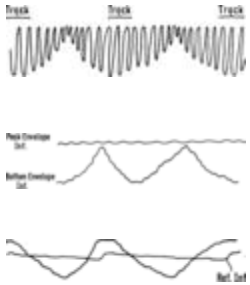
Sample result screen



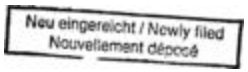
PatMedia Review - figure extraction and categorisation

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



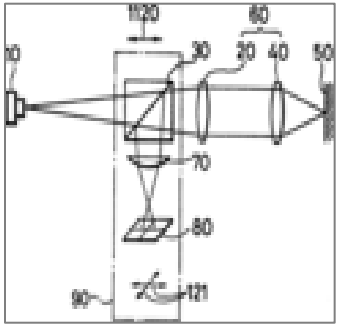
- Improve figure separation programme
  - EP224935 has 5 Figures
  - PatMedia shows 11. Figure 5 waveform progression; but two waveforms missing.

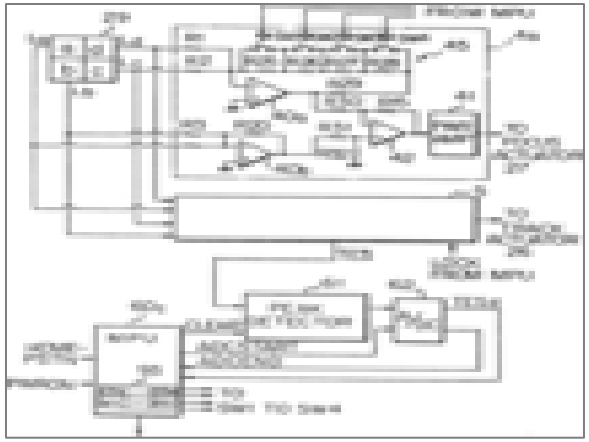


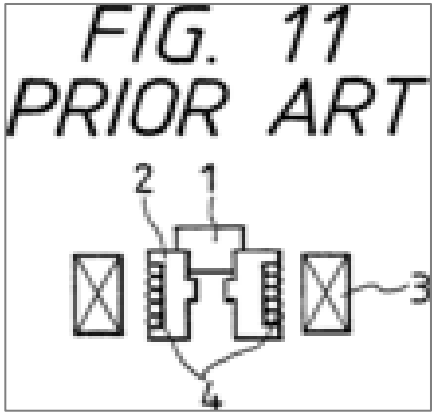
PatMedia Review - image search

- 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