PatCit

Building a comprehensive patent citations dataset

Latest Version

G. Cristelli

EPFL

G. de Rassenfosse

EPFL

F. Gerotto

SciencesPo

K. Higham

Hitotsubashi University

C. Verluise

Collège de France

L. Violon

HEC Paris

You!

Thanks

P. Lopez
T. Simcoe

Sections

US00RE41175E

(19) **United States**

(12) **Reissued Patent**
Vashisth et al.

(10) **Patent Number: US RE41,175 E**
(45) **Date of Reissued Patent: Mar. 30, 2010**

(54) **GPS-ENHANCED SYSTEM AND METHOD FOR AUTOMATICALLY CAPTURING AND CO-REGISTERING VIRTUAL MODELS OF A SITE**

(75) Inventors: **Robert M. Vashisth**, Salt Lake City, UT (US); **James U. Jensen**, Salt Lake City, UT (US); **James W. Bunger**, Salt Lake City, UT (US)

(73) Assignee: **Intelisum, Inc.**, Salt Lake City, UT (US)

(21) Appl. No.: **11/480,248**
(22) Filed: **Jun. 30, 2006**

Related U.S. Patent Documents

Reissue of:
(64) Patent No.: **6,759,979**
Issued: **Jul. 6, 2004**
Appl. No.: **10/348,275**
Filed: **Jan. 21, 2003**

U.S. Applications:
(60) Provisional application No. 60/350,860, filed on Jan. 22, 2002.

(51) **Int. Cl.**
G01S 5/14 (2006.01)
G06T 15/00 (2006.01)
G09G 5/00 (2006.01)
H04N 5/225 (2006.01)

(52) **U.S. Cl.** **342/357.13**; 345/419; 345/629; 348/218.1

(58) **Field of Classification Search** 342/357.13; 345/419, 629; 348/218.1
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,337,149	A	8/1994	Kozah et al.	
5,988,862	A	11/1999	Kacyra et al.	
6,166,744	A	12/2000	Jaszlics et al.	
6,246,468	B1	6/2001	Dimsdale	
6,249,600	B1	6/2001	Reed et al.	
6,292,215	B1	9/2001	Vincent	
6,307,556	B1	10/2001	Ellenby et al.	
6,330,523	B1	12/2001	Kacyra et al.	
6,420,698	B1	7/2002	Dimsdale	
6,473,079	B1	10/2002	Kacyra et al.	
6,526,352	B1	2/2003	Breed et al.	
6,664,529	B2 *	12/2003	Pack et al.	250/208.1
6,759,979	B2	7/2004	Vashisth et al.	
2001/0010546	A1	8/2001	Chen	
2002/0060784	A1	5/2002	Pack et al.	
2003/0090415	A1	5/2003	Miyasaka et al.	
2004/0105573	A1	6/2004	Neumann et al.	
2005/0057745	A1	3/2005	Bontje	

FOREIGN PATENT DOCUMENTS

WO	WO97/40342	A2 *	10/1997
WO	WO 01/04576		1/2001
WO	WO 01/88565		11/2001
WO	WO 01/88566		11/2001
WO	WO 01/88741		11/2001
WO	WO 01/88849		11/2001
WO	WO 02/16865		2/2002
WO	WO 97/40342		10/2006

OTHER PUBLICATIONS

Allen, P. et al, Avenue: automated site modeling in urban environments—3–D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on, 2001, pp. 357–364.*

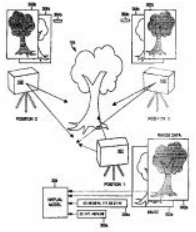
(Continued)

Primary Examiner—Gregory C Issing
(74) *Attorney, Agent, or Firm*—Austin Rapp & Hardman

(57) **ABSTRACT**

A system for capturing a virtual model of a site includes a range scanner for scanning the site to generate range data indicating distances from the range scanner to real-world objects. The system also includes a global positioning system (GPS) receiver coupled to the range scanner for acquiring GPS data for the range scanner at a scanning location. In addition, the system includes a communication interface for outputting a virtual model comprising the range data and the GPS data.

43 Claims, 11 Drawing Sheets



Pedro F. Felzenswalb, et al., "Pictorial Structures for Object Recognition," Artificial Intelligence Lab, Massachusetts Institute of Technology, Computer Science Department, Cornell University, 2003.

Helmut Pottmann, et al., "Registration without ICP," Geometric Modeling and Industrial Geometry Group, Vienna University of Technology, Mar. 5, 2004.

Matthew P. Tait, "Point Cloud registration: Current State of the Science," Schulich School of Engineering, University of Calgary, Mar. 27, 2006.

Steven Alexander Sablerolle, "Automatic Registration of Laser Scanning Data and Colour Images," TUDelft, Oct. 2006.

Ayman F. Habib, et al., "Automatic Surface Matching for the Registration of LIDAR Data and MR Imagery," ETRI Journal, vol. 28, No. 2, Apr. 2006.

Peter K. Allen, et al., "3D Modeling of Historic Sites Using Range and Image Data," Dept. of Computer Science, Columbia University, Dept. of Computer Science, Hunger College, May 16, 2008.

Steven Alexander Sablerolle, "Automatic Registration of Laser Scanning Data and Colour Images," TU Delft, Faculty of Civil Engineering, Final Presentation Msc. Geomatics, Oct. 31, 2006.

Steven Sablerolle, "Graduation Research Report," TU Delft, Sep. 2006.

Mehdi Bouroumand, et al., "The Fusion of Laser Scanning and Close Range Photogrammetry in Bam Laser–Photogrammetric Mapping of Bam Citadel (Arg–E–Bam)/Iran," KNT University in Tehran, 2004.

Isi–Gi, Dec. 21, 2006.

"HLODs: Hierarchical Levels of Detail, hierarchical Simplification for Faster Display of Massive Geometric environments," Department of Computer Science, University of North Carolina at Chapel Hill, Feb. 2004.

Wikipedia, the free encyclopedia, "Level of Detail," http://en.wikipedia.org/wiki/Level_of_detail_(programming), May 23, 2008.

Tahir Rabbani Shah, "Automatic Reconstruction of Industrial Installations Using Point Clouds and Images," Geodesy 6.2, Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission, Delft, May 2006.

T. Rabbani, et al., "Segmentation of Point clouds Using Smoothness Constraint," ISPRS Commission V Symposium 'Image Engineering and Vision Metrology,' 2006.

Tahir Rabbani, et al., "Efficient Hough Transform for Automatic detection of Cylinders in Point Clouds," Workshop "Laser Scanning 2005," Enschede, the Netherlands, Sep. 12–14, 2005.

B. Vangelder, et al., "Modern Technologies for Design Data Collection. " Civil Engineering, Joint Transportation Research Program, Purdue Libraries, 2005.

Dinesh Manadhar, "Extraction of linear features from vehicle–borne laser data," Remote Sensing, Singapore, vol. 2, p. 113–1118, Nov. 5–9, 2001.

Hendrik P. A. Lensch, et al., "Automated Texture Registration and Stitching for Real World Models," Max–Planck–Institute for Computer Science, Baarbrucken, Germany, 2000.

"A Point–and–Shoot Color 3D Camera," Askold V. Strat, Manuel M. Oliveira, Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–8.

"Adaptive Enhancement of 3D Scenes using Hierarchical Registration of Texture–Mapped 3D models," Srikumar Ramalingam and Suresh K. Lodha, Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–8.

"Automatic Registration of Range Images Based on Correspondence of Complete Plane Patches," Wenfeng He, Wei Ma, Hongbin Zha, Proceedings of the Fifth International Conference on 3–D Digital Imaging and Modeling, 2005, pp. 1–6.

"A Multi–Resolution ICP with Heuristic Closest Point Search for Fast and Robust 3D Registration of Range Images," Timothée Jost and Heinz Hügli, Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–7.

"Effective Nearest Neighbor Search for Aligning and Merging Range Images," Ryusuke Sagawa et al., Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–8.

"Enhanced, Robust Genetic Algorithms for Multiview Range Image Registration," Luciano Silva et al., Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–8.

"Non–parametric 3D Surface Completion," Toby P. Breckon, Robert B. Fisher, Proceedings of the Fifth International Conference on 3–D Digital Imaging and Modeling, 2005, pp. 1–8.

"Projective Surface Matching of Colored 3D Scans," Kari Pulli et al., Proceedings of the Fifth International Conference on 3–D Digital Imaging and Modeling, 2005, pp. 1–8.

"Combining texture and shape for automatic crude patch registration," Joris Vanden Wyngaerd et al., Proceedings of the Fourth International Conference on 3–D Digital Imaging and Modeling, 2003, pp. 1–8.

"Automated reconstruction of 3D models from real environments," V. Sequeira, K. Ng, E. Wolfart, J.G.M. Gonçalves, D. Hogg, ISPRS Journal of Photogrammetry & Remote Sensing 54, 1999, pp. 1–22.

"3D Reality Modelling: Photo–Realistic 3D Models of Real World Scenes," Vitor Sequeira, João G.M. Gonçalves, Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission, 2002, pp. 1–8.

"An Integrated Multi–Sensory System for Photo–Realistic 3D Scene Reconstruction," Kia Ng et al., School of Computer Studies, University of Leeds, European Commission—Joint Research Centre, pp. 1–9.

"Lecture Notes In Computer Science," http://portal.acm.org/toc.cfm?id=645595&type=proceeding&coll=GUI..., Jul. 6, 2006, pp. 1–6.

"2–D and 3–D Image Registration: A Tutorial," http://www.cs.wright.edu/~agoshtas/CVPR04_Registration_Tutorial.html, Jul. 7, 2006, pp. 1–3.

"Medical Image Processing," http://www.sce.carleton.ca/faculty/adler/elg7173/elg7173.html, Jul. 7, 2006, pp. 1–5.

"Gaussian Random Fields on Sub–Manifolds for Characterizing Brain Surfaces," http://portal.acm.org/citation.cfm?id=645595.660540&coll=GUIDE&d..., Jul. 6, 2006, pp. 1–3.

"Image Image–Guided Interventions Workshop Guided Interventions Workshop," John Haller, National Institute of Biomedical Imaging and Bioengineering, May 13, 2004–May 14, 2004, pp. 2–11.

"Workshop on symmetries, inverse problems and image processing," http://www.indmath.uni–linz.ac.at/people/bila/Workshop.html, Jan. 12, 2005, pp. 1–4.

Mission

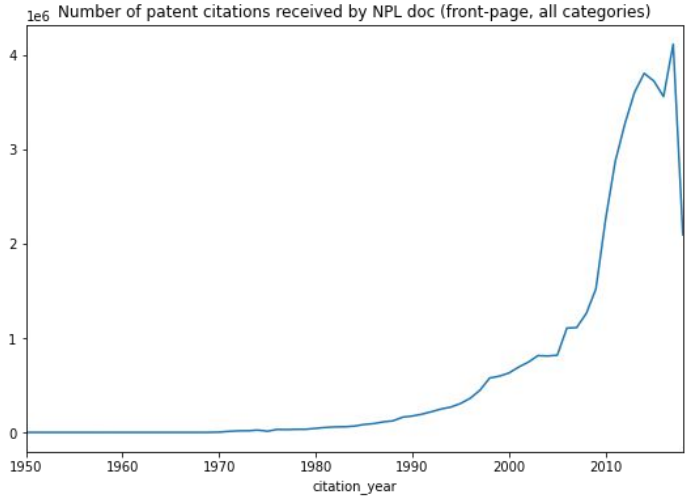
Build a community driven comprehensive patent citations dataset

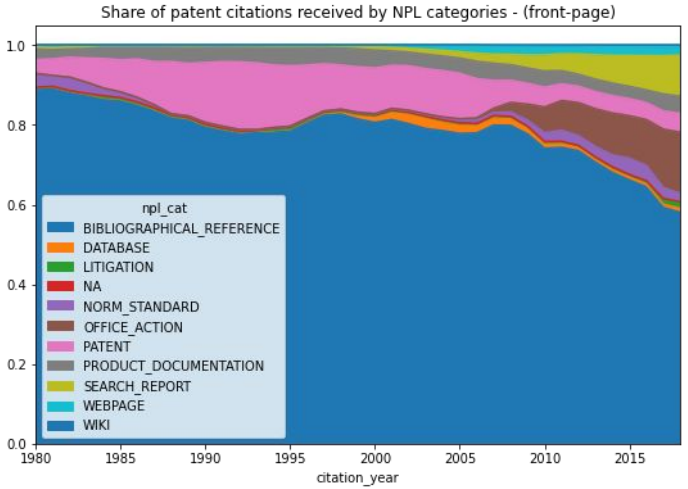
- Worldwide
- Front page & In-text
- All categories of NPL

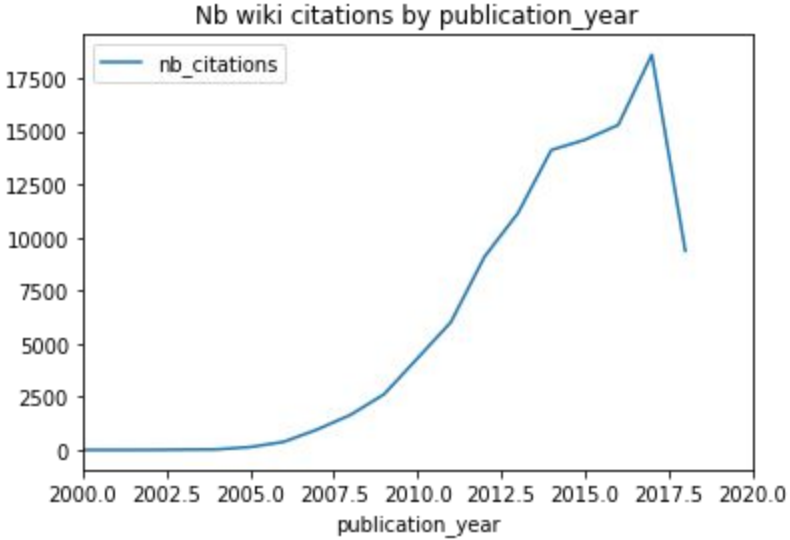
→ *One-stop-shop by and for the community + we take care of data engineering*

Motivations

- Embed patents in the innovation system: science, *open knowledge*, IP institutions & competition
- Improve knowledge flows measurement


Number of patent citations received by NPL doc (front-page, all categories)
1e6
4
3
2
1
0
1950
1960
1970
1980
1990
2000
2010
citation_year


Share of patent citations received by NPL categories - (front-page)
1.0
0.8
0.6
0.4
0.2
0.0
1980
1985
1990
1995
2000
2005
2010
2015
citation_year
npl_cat
BIBLIOGRAPHICAL_REFERENCE
DATABASE
LITIGATION
NA
NORM_STANDARD
OFFICE_ACTION
PATENT
PRODUCT_DOCUMENTATION
SEARCH_REPORT
WEBPAGE
WIKI



Graph

- Nb wiki nodes: 70,052
- Nb patent nodes: 64,224
- Nb edges: 108,475

Previous work

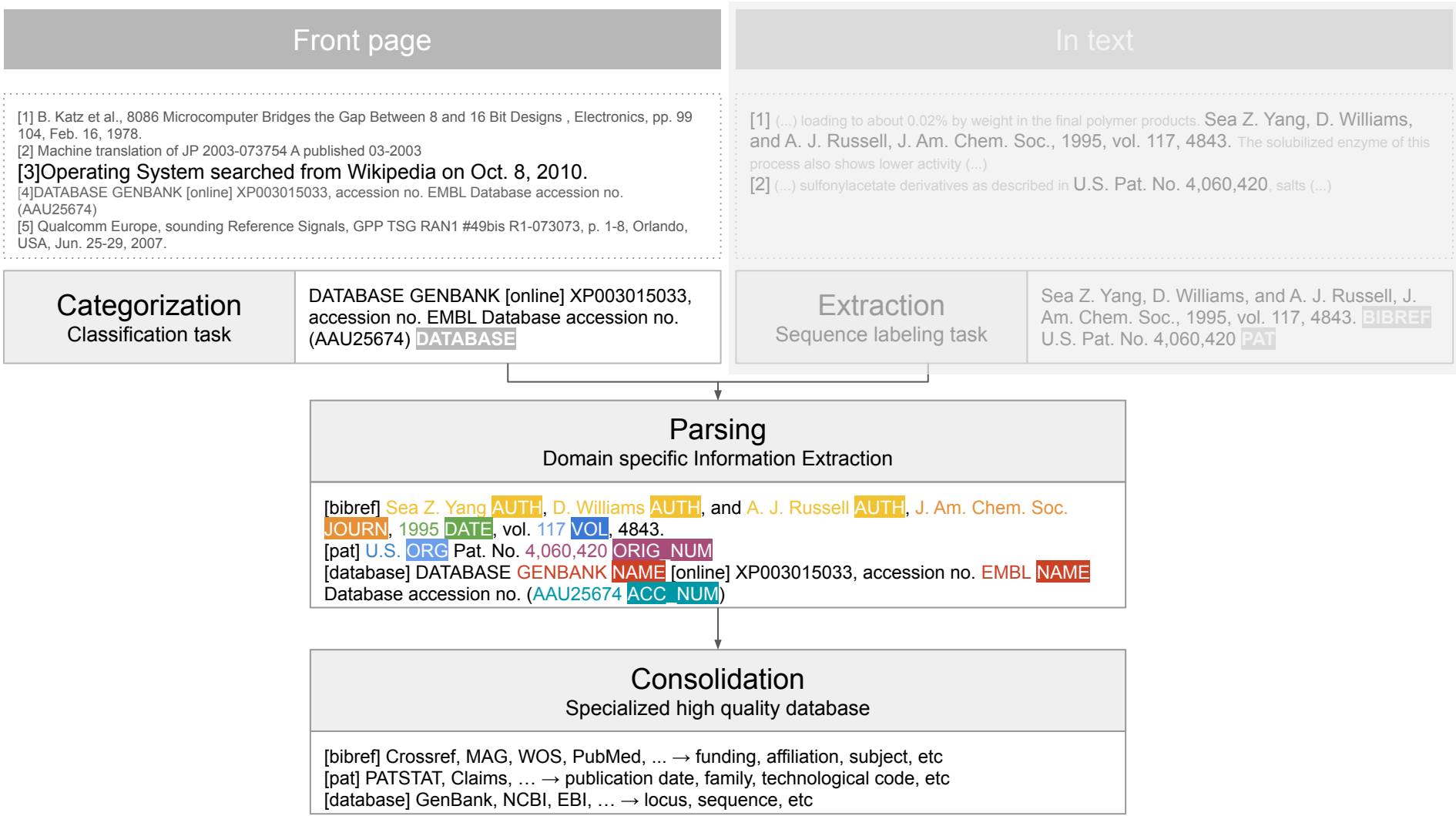
- Patents and scientific citations
  - Jefferson et al (2018)
  - Marx and Fuegi (2020a,b)
  - Bryan et al (2020)
- Natural Language Processing & citations extraction/parsing
  - Galibert et al (2010)
  - Verberne et al (2018)
  - Grobid (2008-2020)

→ WIKI not addressed

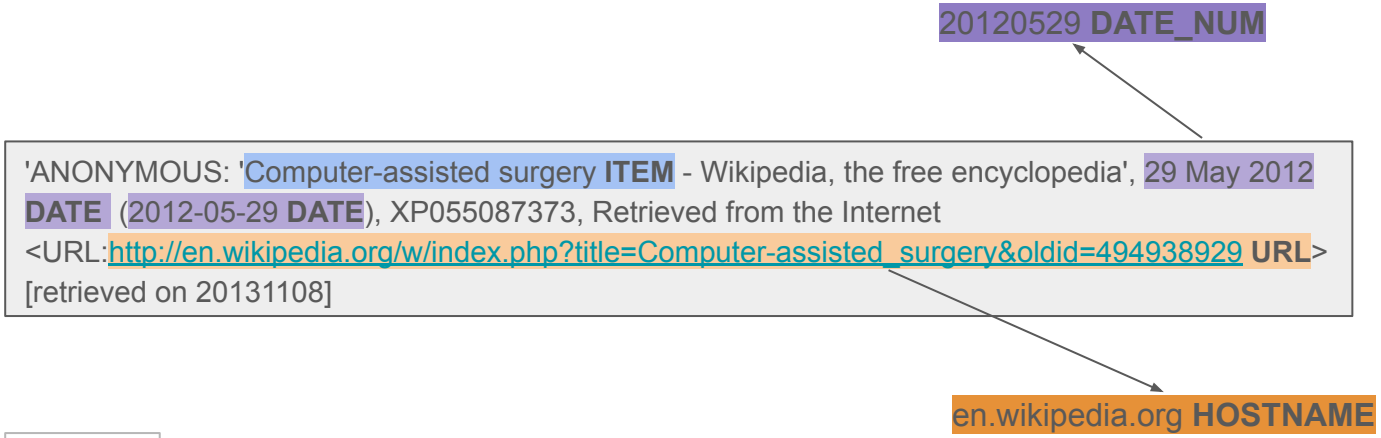
Methodology

Bridging the gap

- Syndicate field expertise
- Domain specific information extraction models
- Data enrichment


Front page
In text
[1] B. Katz et al., 8086 Microcomputer Bridges the Gap Between 8 and 16 Bit Designs , Electronics, pp. 99 104, Feb. 16, 1978.
[2] Machine translation of JP 2003-073754 A published 03-2003
[3]Operating System searched from Wikipedia on Oct. 8, 2010.
[4]DATABASE GENBANK [online] XP003015033, accession no. EMBL Database accession no. (AAU25674)
[5] Qualcomm Europe, sounding Reference Signals, GPP TSG RAN1 #49bis R1-073073, p. 1-8, Orlando, USA, Jun. 25-29, 2007.
[1] (...) loading to about 0.02% by weight in the final polymer products. Sea Z. Yang, D. Williams, and A. J. Russell, J. Am. Chem. Soc., 1995, vol. 117, 4843. The solubilized enzyme of this process also shows lower activity (...)
[2] (...) sulfonylacetate derivatives as described in U.S. Pat. No. 4,060,420, salts (...)
Categorization
Classification task
DATABASE GENBANK [online] XP003015033, accession no. EMBL Database accession no. (AAU25674) DATABASE
Extraction
Sequence labeling task
Sea Z. Yang, D. Williams, and A. J. Russell, J. Am. Chem. Soc., 1995, vol. 117, 4843. BIBREF
U.S. Pat. No. 4,060,420 PAT
Parsing
Domain specific Information Extraction
[bibref] Sea Z. Yang AUTH, D. Williams AUTH, and A. J. Russell AUTH, J. Am. Chem. Soc. JOURN, 1995 DATE, vol. 117 VOL, 4843.
[pat] U.S. ORG Pat. No. 4,060,420 ORIG_NUM
[database] DATABASE GENBANK NAME [online] XP003015033, accession no. EMBL NAME Database accession no. (AAU25674 ACC_NUM)
Consolidation
Specialized high quality database
[bibref] Crossref, MAG, WOS, PubMed, ... → funding, affiliation, subject, etc
[pat] PATSTAT, Claims, … → publication date, family, technological code, etc
[database] GenBank, NCBI, EBI, … → locus, sequence, etc

Domain specific information extraction - WIKI



Regex URL

NER ITEM, DATE (F-score resp. 85% and 94%)

Results

- In-text
  - Coverage: worldwide
  - Categorize: 10 categories, f1:.86
  - Parse: already 4 category specific information extraction models
  - Consolidate: 40% bibliographical references enriched with Crossref
- Front-page
  - Coverage: US all time
  - Extract: bibliographical references & patents, f1: .69 & .89
  - Parse: both categories
  - Consolidate: bibliographical references enriched with Crossref and patents enriched with PATSTAT/Claims

Front page

	Categorization	Parsing	Consolidation
Bibliographical reference	p: .84 | r: .92 | f: .87	a: [.77-.92][| no DOI]	Crossref: 40%
Office action	p: .89 | r: .93 | f: .90		
Patent	p: .78 | r: .89 | f: .83		
Search report	p: .96 | r: .87 | f: .91		
Norm Standard	p: .87 | r: .71 | f: .78	p:[.94-.99]| r:[.94-.100]| f:[.95-.99]	Dev - help welcome
Product documentation	p: .66 | r: .62 | f: .64		
Webpage	p: .57 | r: .50 | f: .53		
Database	p: .94 | r: .88 | f: .90	p:[.91-.95]| r:[.90-.92]| f:[.91-.93]	Dev - help welcome
Litigation	p: .93 | r: .84 | f: .88		
Wiki	p: - | r: - | f: -	p:[.92-.96]| r:[.92-.98]| f:[.92-.97]	Dev - help welcome

a=accuracy | p = precision | r = recall | f = f1-score

In-text

	Extraction	Parsing	Consolidation
Bibliographical reference	p: .80 | r: .60 | f: .69 [ex-post]	a: [.77-.92][| no DOI]	Crossref
Patent	p: .97| r: .82 | f: .89	a: [.96-98]	PATSTAT/Claims

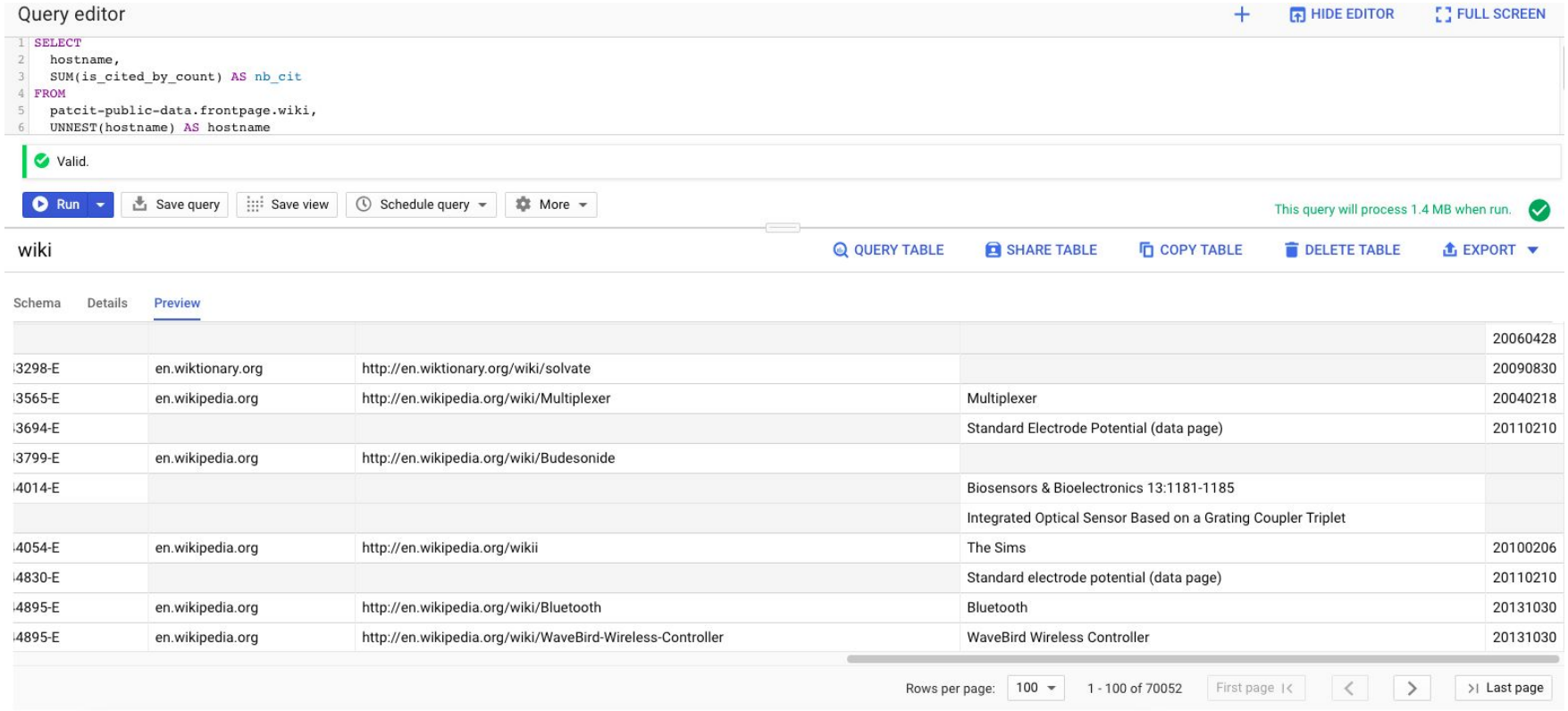
p = precision | r = recall | f = f1-score
Ex-post: we apply a classifier on top of Grobid extraction model to improve the precision. Ex ante p: .41, r: .64, f: .51

Data

FAIR

- Find
  - Zenodo [dump + data versioning]
  - BigQuery [interactive + latest]
- Access
  - Documentation website
  - Issues/requests/etc on GitHub
- Interoperability
  - [bibref] DOI, PMID, PMCID
  - [patent] publication number
  - [database] accession number
  - [wiki] url
- Reproduce
  - GitHub
  - DVC (model + training data)

Time to play!


Query editor
HIDE EDITOR
FULL SCREEN
SELECT
hostname,
SUM(is_cited_by_count) AS nb_cit
FROM
patcit-public-data.frontpage.wiki,
UNNEST(hostname) AS hostname
Valid.
Run
Save query
Save view
Schedule query
More
This query will process 1.4 MB when run.
wiki
QUERY TABLE
SHARE TABLE
COPY TABLE
DELETE TABLE
EXPORT
Schema
Details
Preview
20060428
3298-E
en.wiktionary.org
http://en.wiktionary.org/wiki/solvate
20090830
3565-E
en.wikipedia.org
http://en.wikipedia.org/wiki/Multiplexer
Multiplexer
20040218
3694-E
Standard Electrode Potential (data page)
20110210
3799-E
en.wikipedia.org
http://en.wikipedia.org/wiki/Budesonide
4014-E
Biosensors & Bioelectronics 13:1181-1185
Integrated Optical Sensor Based on a Grating Coupler Triplet
4054-E
en.wikipedia.org
http://en.wikipedia.org/wikii
The Sims
20100206
4830-E
Standard electrode potential (data page)
20110210
4895-E
en.wikipedia.org
http://en.wikipedia.org/wiki/Bluetooth
Bluetooth
20131030
4895-E
en.wikipedia.org
http://en.wikipedia.org/wiki/WaveBird-Wireless-Controller
WaveBird Wireless Controller
20131030
Rows per page:
100
1 - 100 of 70052
First page
Last page

Updates

Stay informed & Get involved

PatCit in your mails

`v0.3` **Coming soon!**

PatCit on GitHub

Thank You!

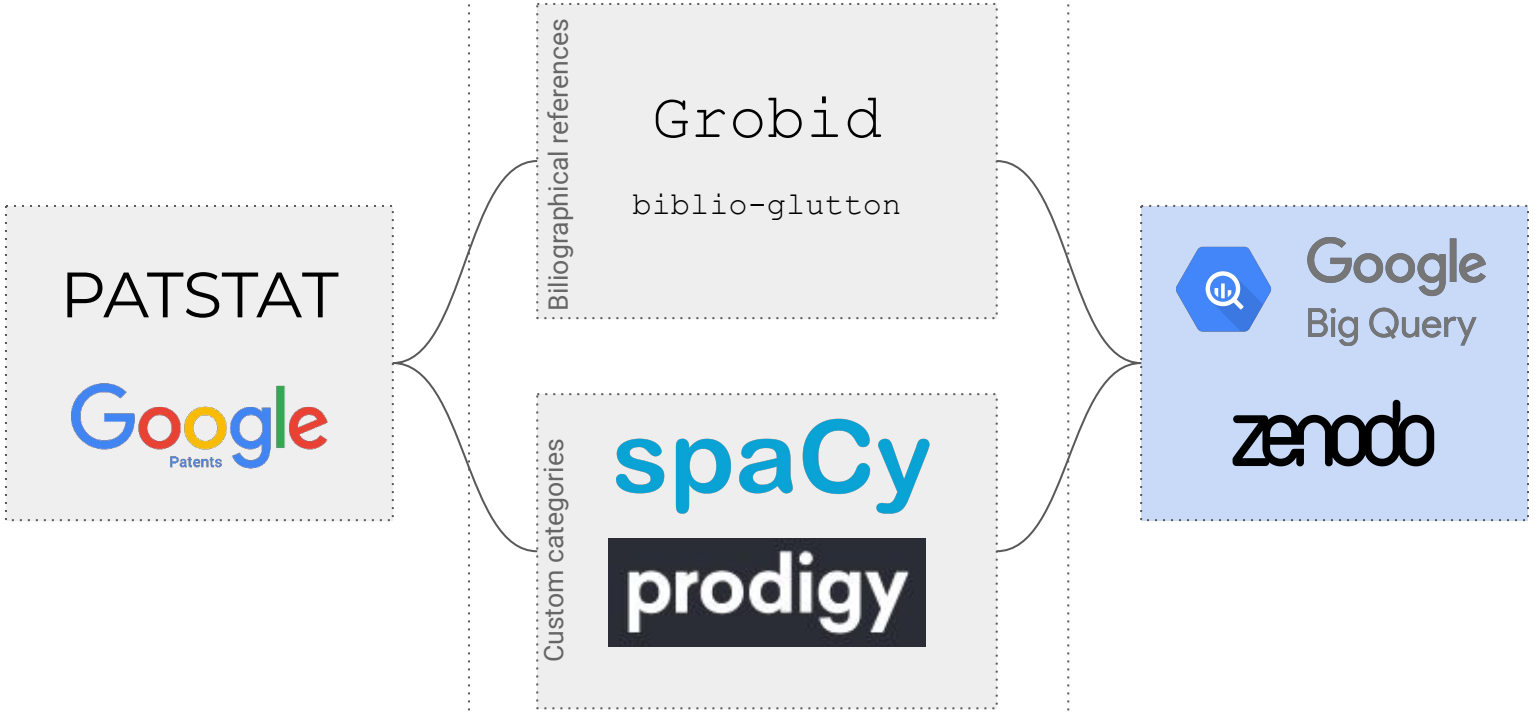
NPL categories (1/2)

Category	Group	Example
Bibliographical reference	Science	B. Katz et al., 8086 Microcomputer Bridges the Gap Between 8 and 16 Bit Designs , Electronics, pp. 99 104, Feb. 16, 1978.
Database	Science	DATABASE GENBANK [online] XP003015033, accession no. EMBL Database accession no. (AAU25674)
Norm Standard	IO/Competion	Qualcomm Europe, sounding Reference Signals, GPP TSG RAN1 #49bis R1-073073, p. 1-8, Orlando, USA, Jun. 25-29, 2007.
Product documentation	IO/Competion	Brochure, Roadtec SP 100 Gravity Feed Paver (undated).
Litigation	IO/Competion	Suggestion for Interference filed in connection with U.S. Appl. No. 10/689,866 on May 11, 2010.

NPL categories (2/2)

Category	Group	Example
Wiki	Open knowledge	Operating System searched from Wikipedia on Oct. 8, 2010.
Webpage	Open knowledge	'P2P Internet', halfbakery, http://www.halfbakery.com/idea/P2P-20Internet#10952502800, 3 pages.
Search report	IP institutions	International Search report and Written Opinion for PCT Application PCT/US2010/043935 Search report dated Nov. 16, 2010.
Office action	IP institutions	First Office Action issued by the State Intellectual Property Office of People's Republic of China in corresponding CN Application No. 201080037669.8, dated Nov. 18, 2013.
Patent		Machine translation of JP 2003-073754 A published 03-2003

Technical stack



Named Entity Recognition - Review

TODO

Vocabulary

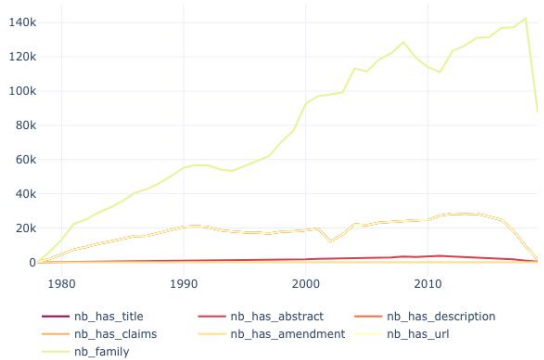
Accuracy - Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition: *Accuracy = # correct predictions/ # of predictions*.

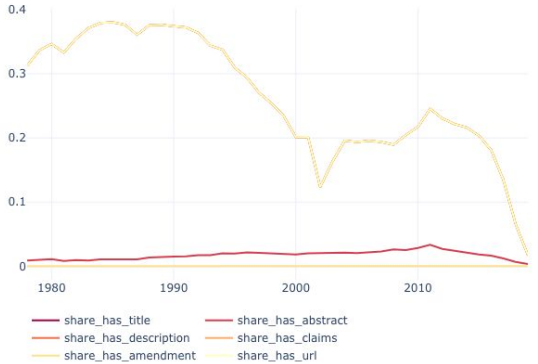
Precision - Informally, precision attempts to answer the following question: "What proportion of positive identifications was actually correct?". Formally, precision has the following definition: *Precision = # True Positives/ (# True Positives + # False Positives)*.

Recall - Informally, recall attempts to answer the following question: “What proportion of actual positives was identified correctly?”. Formally, recall has the following definition: *Recall = # True Positives / (# True Positives + # False Negatives)*

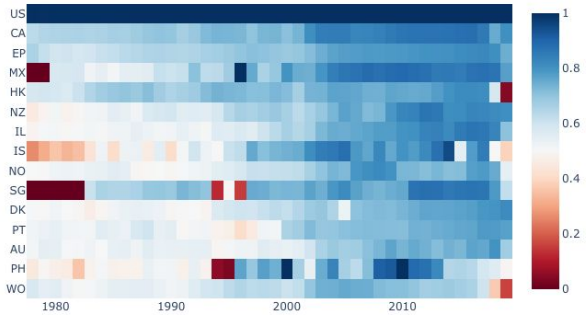
EPO full-text

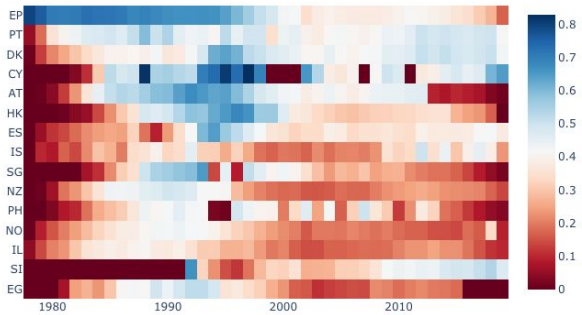
Number and share of EPO patent families with full-text data coverage *in english*





USPTO and EPO full-text patent description data coverage





→ Paradoxically, the **USPTO** full-text data implicitly cover a larger share of **EP** families than the EP bulk dataset.