

VISUAL STEP-BY-STEP CASE STUDY: PATENT TROLLS

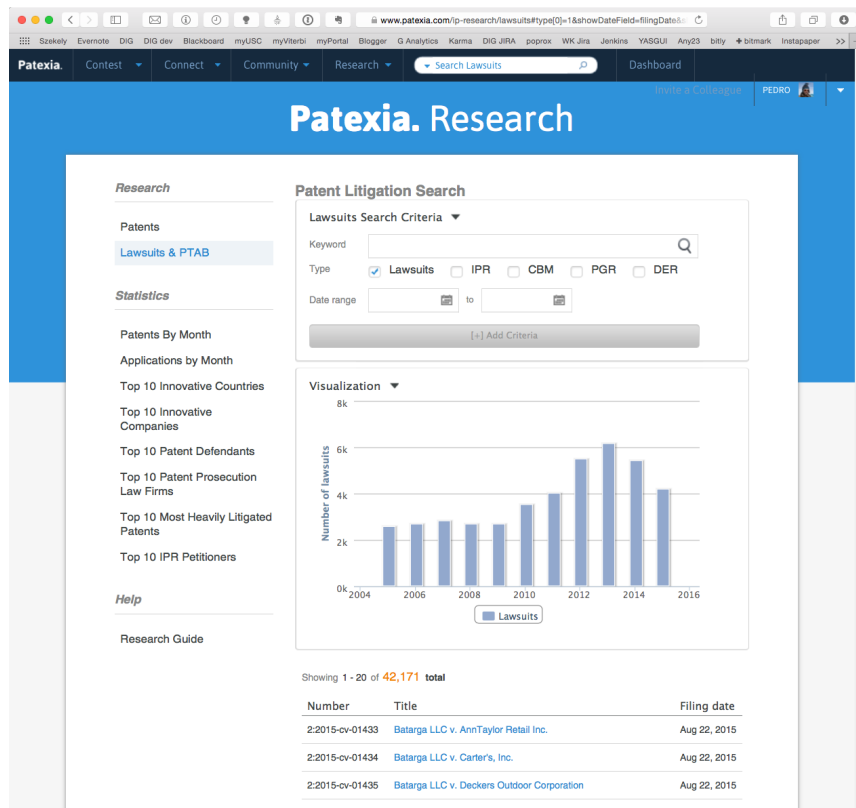
Let's apply the course concepts we have studied in an actual information extraction application

Opening prompt / problem statement

- Companies use patents to protect their intellectual property. Patent Trolls are companies that purchase patents and then threaten to sue many companies for patent infringement, looking to settle rather than go to court
 - Who are the patent trolls?
 - Which patents are being trolled?
 - Who are their victims?

Context and opportunity

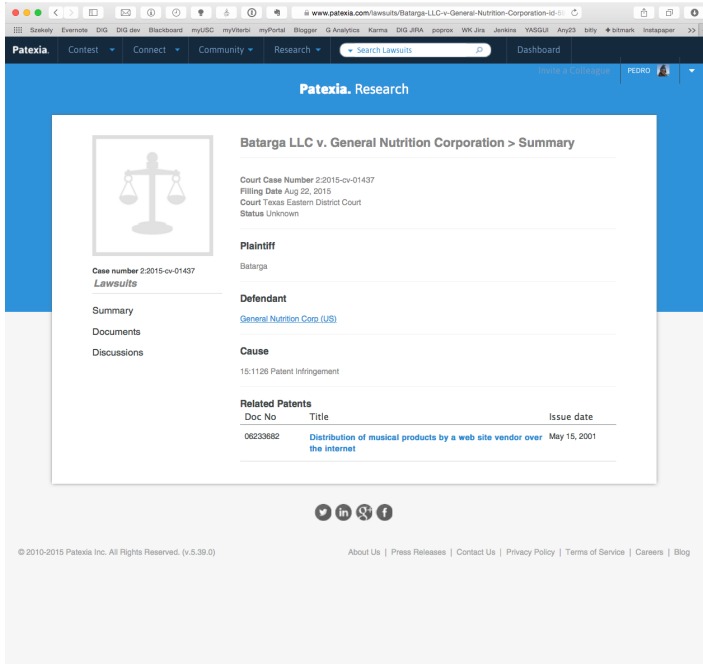
Patent research is a major application area



[https://www.patexia.com/ip-research/lawsuits#type\[0\]=1&showDateField=filingDate&showCharts=1](https://www.patexia.com/ip-research/lawsuits#type[0]=1&showDateField=filingDate&showCharts=1)

VISUAL STEP-BY-STEP CASE STUDY: PATENT TROLLS

Example lawsuit

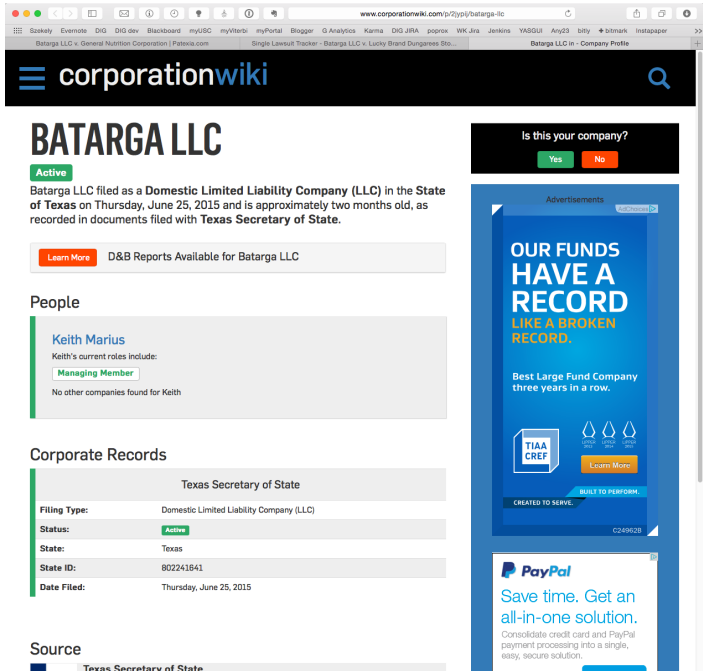


The screenshot shows the Patexia Research page for the lawsuit **Batarga LLC v. General Nutrition Corporation**. The page includes a sidebar with navigation links: **Lawsuits**, **Summary**, **Documents**, and **Discussions**. The main content area displays the following information:

- Case number:** 2:2015-cv-01437
- Summary:** Court Case Number 2:2015-cv-01437, Filing Date Aug 22, 2015, Court Texas Eastern District Court, Status Unknown.
- Plaintiff:** Batarga
- Defendant:** General Nutrition Corp (GNC)
- Cause:** 15:1126 Patent Infringement
- Related Patents:** A table with columns for Doc No, Title, and Issue date.

Doc No	Title	Issue date
06233682	Distribution of musical products by a web site vendor over the internet	May 15, 2001

At the bottom of the page, there are social media icons for Twitter, LinkedIn, Facebook, and YouTube, along with a copyright notice: © 2010-2015 Patexia Inc. All Rights Reserved. (v.5.39.0).



The screenshot shows the CorporationWiki page for **BATARGA LLC**. The page includes a sidebar with navigation links: **People**, **Corporate Records**, and **Source**. The main content area displays the following information:

- Active:** Batarga LLC filed as a Domestic Limited Liability Company (LLC) in the State of Texas on Thursday, June 25, 2015 and is approximately two months old, as recorded in documents filed with Texas Secretary of State.
- Learn More:** D&B Reports Available for Batarga LLC
- People:** Keith Marius, Keith's current roles include: Managing Member. No other companies found for Keith.
- Corporate Records:** Texas Secretary of State. Filing Type: Domestic Limited Liability Company (LLC). Status: Active. State: Texas. State ID: 802241641. Date Filed: Thursday, June 25, 2015.
- Source:** Texas Secretary of State

On the right side of the page, there are two advertisements. The top one is for TIAA CREF, titled "OUR FUNDS HAVE A RECORD LIKE A BROKEN RECORD." The bottom one is for PayPal, titled "Save time. Get an all-in-one solution."

<http://www.corporationwiki.com/p/2jypij/batarga-llc>

VISUAL STEP-BY-STEP CASE STUDY: PATENT TROLLS

Downloading US Patent Office patents in bulk

The screenshot shows a web browser window displaying the Google Books page for "USPTO Bulk Downloads: Patent Grant Full Text". The page title is "USPTO Bulk Downloads: Patent Grant Full Text". Below the title, it states: "Full text of USPTO patent grants: 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1999, 1998, 1997, 1996, 1995, 1994, 1993, 1992, 1991, 1990, 1989, 1988, 1987, 1986, 1985, 1984, 1983, 1982, 1981, 1980, 1979, 1978, 1977, 1976." It then says "Patent Grant Full Text (2001 to present):" and describes the content: "Contains the full text including tables, sequence data and 'in-line' mathematical expressions of each patent grant issued weekly (Tuesdays) from January. The file is a concatenation of the Standard Generalized Markup Language (SGML) in accordance with the U.S. Patent Grant Version 2.4 Document Type Definition (DTD) and eXtensible Markup Language (XML) in accordance with the U.S. Patent Grant Version 2.5; 4.0 International Common Element (ICE); 4.1 ICE; 4.2 ICE Document Type Definitions (DTDs). Sequence data XML text in accordance with the ICE SEQLST V1.2 DTD (us-sequence-listing-2004-03-09.dtd) is concatenated next to the containing grant SGML or XML text. References to the following external files are present but the external files are not present:" followed by a bulleted list: "Mega Sequence Listing data files", "Mathematica Notebook (NB) files", "CS ChemDraw (CDX) and MDL Information Systems (MOL) files", and "Drawings, mathematical expressions, and chemical structures image (TIFF) files". Below this, it says "Refer to the following USPTO web site for additional patent data information (2001 through present) to include the Document Type Definitions (DTDs):" and provides a URL: "http://www.uspto.gov/web/offices/ac/ido/oeip/sgml/st32/redbook/rb2004/rb2004.html". It then says "Patent Grant Full Text (1976 to 2001):" and describes the content: "Contains the full text of each patent grant issued weekly (Tuesdays) from January 1976 to December 2001. The file format is ASCII text (a.k.a. Patent Grant Green Book). Included are tables and 'in-line' mathematical equations, where appropriate, appearing as text data. Chemical structures are not present, but their location is indicated by a structure call-out. Includes patent number, series code and application number, type of patent, filing date, title, issue date, inventor information, assignee name at time of issue, foreign priority information, related US patent documents, classification information, US and foreign references, attorney, agent or firm/legal representative, Patent Cooperation Treaty (PCT) information, abstract, specification, and claims. Approximately 4,000 patent grants per week." It then says "Refer to the following link for additional Patent Grant Data/APS documentation:" and provides a URL: "PatentFullTextAPSDoc_GreenBook.pdf". Below this, it lists the years 2015, 2014, 2013, and 2012, each followed by a list of patent grant files (e.g., ipg150106.zip, ipg150113.zip, etc.).

Google **USPTO Bulk Downloads: Patent Grant Full Text**

Full text of USPTO patent grants: 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000, 1999, 1998, 1997, 1996, 1995, 1994, 1993, 1992, 1991, 1990, 1989, 1988, 1987, 1986, 1985, 1984, 1983, 1982, 1981, 1980, 1979, 1978, 1977, 1976.

Patent Grant Full Text (2001 to present):

Contains the full text including tables, sequence data and "in-line" mathematical expressions of each patent grant issued weekly (Tuesdays) from January. The file is a concatenation of the Standard Generalized Markup Language (SGML) in accordance with the U.S. Patent Grant Version 2.4 Document Type Definition (DTD) and eXtensible Markup Language (XML) in accordance with the U.S. Patent Grant Version 2.5; 4.0 International Common Element (ICE); 4.1 ICE; 4.2 ICE Document Type Definitions (DTDs). Sequence data XML text in accordance with the ICE SEQLST V1.2 DTD (us-sequence-listing-2004-03-09.dtd) is concatenated next to the containing grant SGML or XML text. References to the following external files are present but the external files are not present:

- Mega Sequence Listing data files
- Mathematica Notebook (NB) files
- CS ChemDraw (CDX) and MDL Information Systems (MOL) files
- Drawings, mathematical expressions, and chemical structures image (TIFF) files

Refer to the following USPTO web site for additional patent data information (2001 through present) to include the Document Type Definitions (DTDs):
<http://www.uspto.gov/web/offices/ac/ido/oeip/sgml/st32/redbook/rb2004/rb2004.html>

Patent Grant Full Text (1976 to 2001):

Contains the full text of each patent grant issued weekly (Tuesdays) from January 1976 to December 2001. The file format is ASCII text (a.k.a. Patent Grant Green Book). Included are tables and "in-line" mathematical equations, where appropriate, appearing as text data. Chemical structures are not present, but their location is indicated by a structure call-out. Includes patent number, series code and application number, type of patent, filing date, title, issue date, inventor information, assignee name at time of issue, foreign priority information, related US patent documents, classification information, US and foreign references, attorney, agent or firm/legal representative, Patent Cooperation Treaty (PCT) information, abstract, specification, and claims. Approximately 4,000 patent grants per week.

Refer to the following link for additional Patent Grant Data/APS documentation:
PatentFullTextAPSDoc_GreenBook.pdf

2015

[ipg150106.zip](#) [ipg150113.zip](#) [ipg150120.zip](#) [ipg150127.zip](#) [ipg150203.zip](#) [ipg150210.zip](#) [ipg150217.zip](#) [ipg150224.zip](#) [ipg150303.zip](#) [ipg150310.zip](#) [ipg150317.zip](#)

2014

[ipg140107.zip](#) [ipg140114.zip](#) [ipg140121.zip](#) [ipg140128.zip](#) [ipg140204.zip](#) [ipg140211.zip](#) [ipg140218.zip](#) [ipg140225.zip](#) [ipg140304.zip](#) [ipg140311.zip](#) [ipg140318.zip](#) [ipg140325.zip](#) [ipg140401.zip](#) [ipg140408.zip](#) [ipg140415.zip](#) [ipg140422.zip](#) [ipg140429.zip](#) [ipg140506.zip](#) [ipg140513.zip](#) [ipg140520.zip](#) [ipg140527.zip](#) [ipg140603.zip](#) [ipg140610.zip](#) [ipg140617.zip](#) [ipg140624.zip](#) [ipg140701.zip](#) [ipg140708.zip](#) [ipg140715.zip](#) [ipg140722.zip](#) [ipg140729.zip](#) [ipg140805.zip](#) [ipg140812.zip](#) [ipg140819.zip](#) [ipg140826.zip](#) [ipg140902.zip](#) [ipg140909.zip](#) [ipg140916.zip](#) [ipg140923.zip](#) [ipg140930.zip](#) [ipg141007.zip](#) [ipg141014.zip](#) [ipg141021.zip](#) [ipg141028.zip](#) [ipg141104.zip](#) [ipg141111.zip](#) [ipg141118.zip](#) [ipg141125.zip](#) [ipg141202.zip](#) [ipg141209.zip](#) [ipg141216.zip](#) [ipg141223.zip](#) [ipg141230.zip](#)

2013

[ipg130101.zip](#) [ipg130108.zip](#) [ipg130115.zip](#) [ipg130122.zip](#) [ipg130129.zip](#) [ipg130205.zip](#) [ipg130212.zip](#) [ipg130219.zip](#) [ipg130226.zip](#) [ipg130305.zip](#) [ipg130312.zip](#) [ipg130319.zip](#) [ipg130326.zip](#) [ipg130402.zip](#) [ipg130409.zip](#) [ipg130416.zip](#) [ipg130423.zip](#) [ipg130430.zip](#) [ipg130507.zip](#) [ipg130514.zip](#) [ipg130521.zip](#) [ipg130528.zip](#) [ipg130604.zip](#) [ipg130611.zip](#) [ipg130618.zip](#) [ipg130625.zip](#) [ipg130702.zip](#) [ipg130709.zip](#) [ipg130716.zip](#) [ipg130723.zip](#) [ipg130730.zip](#) [ipg130806.zip](#) [ipg130813.zip](#) [ipg130820.zip](#) [ipg130827.zip](#) [ipg130903.zip](#) [ipg130910.zip](#) [ipg130917.zip](#) [ipg130924.zip](#) [ipg131001.zip](#) [ipg131008.zip](#) [ipg131015.zip](#) [ipg131022.zip](#) [ipg131029.zip](#) [ipg131105.zip](#) [ipg131112.zip](#) [ipg131119.zip](#) [ipg131126.zip](#) [ipg131203.zip](#) [ipg131210.zip](#) [ipg131217.zip](#) [ipg131224.zip](#) [ipg131231.zip](#)

2012

[ipg120103.zip](#) [ipg120110.zip](#) [ipg120117.zip](#) [ipg120124.zip](#) [ipg120131.zip](#) [ipg120207.zip](#) [ipg120214.zip](#) [ipg120221.zip](#) [ipg120228.zip](#) [ipg120306.zip](#) [ipg120313.zip](#) [ipg120320.zip](#) [ipg120327.zip](#) [ipg120403.zip](#) [ipg120410.zip](#) [ipg120417.zip](#) [ipg120424.zip](#) [ipg120501.zip](#) [ipg120508.zip](#) [ipg120515.zip](#) [ipg120522.zip](#) [ipg120529.zip](#) [ipg120605.zip](#) [ipg120612.zip](#) [ipg120619.zip](#) [ipg120626.zip](#) [ipg120703.zip](#) [ipg120710.zip](#) [ipg120717.zip](#) [ipg120724.zip](#) [ipg120731.zip](#) [ipg120807.zip](#) [ipg120814.zip](#) [ipg120821.zip](#) [ipg120828.zip](#) [ipg120904.zip](#) [ipg120911.zip](#) [ipg120918.zip](#) [ipg120925.zip](#) [ipg121002.zip](#) [ipg121009.zip](#) [ipg121016.zip](#) [ipg121023.zip](#) [ipg121030.zip](#) [ipg121106.zip](#) [ipg121113.zip](#) [ipg121120.zip](#) [ipg121127.zip](#) [ipg121204.zip](#) [ipg121211.zip](#) [ipg121218.zip](#) [ipg121225.zip](#)

<https://www.google.com/googlebooks/uspto-patents-grants-text.html>

<http://patents.reedtech.com/assignment.php>

VISUAL STEP-BY-STEP CASE STUDY: PATENT TROLLS

The screenshot shows the ReedTech USPTO Data Sets website. The header includes the ReedTech logo, "USPTO Data Sets", and LexisNexis. The navigation bar has links for HOME, PATENTS, TRADEMARKS, PUBLIC PAIR, ABOUT / CONTACT, and FAQ. The main content area is titled "Patent Assignment Text (1980 – present)" and contains a description of the data set, a note about the Hague Agreement, and a table of file names, sizes, and posting dates. A sidebar on the left includes a search bar, contact information, and website links.

Patent Assignment Text (1980 – present)

Contains both daily (front files) and a (backfile) created annually of patent assignments text derived from patent assignment recordings from August 1980 to present made at the USPTO for granted patents. The file format is Extensible Markup Language (XML) in accordance with the Patent Assignment Daily XML (PADX) Version 0.3 Document Type Definition (DTD).

Refer to the following USPTO web site for additional patent data information (2001 through present) to include the Document Type Definitions (DTDs): <http://www.uspto.gov/products/xml-resources.jsp>

***** 02/13/2015 NOTE:** In anticipation of the Hague Agreement Concerning the International Registration of Industrial Designs (Hague Agreement) having effect in the United States in the future, the description for certain Data Type Definition (DTD) element names has been changed to accommodate international design applications. For more information, please see the documentation, "PADX-File-Description-v2_Hague.doc". Specifically, the description for the following DTD element names has been changed: country; doc-number; kind; and date. DTD element names are not changing. In the future, content relating to international design applications may be present in these DTD fields. The date the Hague Agreement will take effect with respect to these files is *** May 13, 2015. ***

2015 | 1980-2014

2015 [Top]

File Name	File Size (bytes)	Posting Date
ad20150824.zip	93,277	08/25/2015
ad20150823.zip	989	08/24/2015
ad20150822.zip	989	08/23/2015
ad20150821.zip	242,837	08/22/2015
ad20150820.zip	3,824,526	08/21/2015
ad20150819.zip	127,913	08/20/2015
ad20150818.zip	20,243,939	08/19/2015
ad20150817.zip	127,494	08/18/2015
ad20150816.zip	989	08/17/2015
ad20150815.zip	989	08/16/2015
ad20150814.zip	115,639	08/15/2015
ad20150813.zip	3,725,640	08/14/2015
ad20150812.zip	76,882	08/13/2015
ad20150811.zip	17,422,614	08/12/2015
ad20150810.zip	246,392	08/11/2015
ad20150809.zip	988	08/10/2015
ad20150808.zip	988	08/09/2015
ad20150807.zip	107,768	08/08/2015
ad20150806.zip	4,278,693	08/07/2015
ad20150805.zip	108,578	08/06/2015
ad20150804.zip	18,794,105	08/05/2015
ad20150803.zip	183,715	08/04/2015

Case Questions

- 1) The last visualization shows that the bulk patent dataset is divided up into many files. How would you go about automatically downloading all of those using a script?
- 2) Patent files contain images, abstracts and text, among other elements. What aspects are expected to be useful for this case?
- 3) Draw an information integration pipeline (the class diagram is a guide only) and specify the technologies you will use at each stage of the pipeline.
- 4) Recall the three original questions we specified in the prompt. Using the material provided in the case study, and the architecture you drew in Q3, how would you go about providing decision support for those questions to an interested user?
- 5) Is a traditional database (i.e. relational database) or knowledge graph infrastructure more appropriate for this problem? Why or why not?