From BACON to XML

Why Patent Information Publishers are still Converting Image Data to Searchable Text







- Challenge What is it / how is it being met?
- Brief Introduction Richard Garner
- Lexpat The emergence of full text
- <u>Back File Conversion Project</u> What was it / what did it lead to?
- From the lean years to the growth years 20 years of waiting
- An abundance of full text What is available today?
- Spinning Straw into Gold Adding Value to IP Data
- Questions Thank You



Challenge – What is it / how is it being met?

- 185 WIPO Member States
 - 131 "active" patent issuing authorities
 - Patent applications by residents and non-residents (1999 2009)
- Less than 20% publish <u>regular</u> updates of their full text data
 - Fewer still have published their complete back files
- What, why and how:
 - Some 25 years after the EPO announced the launch of its BACON project
 - Patent information publishing industry is still dependent upon image data to create searchable full text data
- Outline:
 - Sources currently available to industry
 - Systems and processes used to convert images to text
 - The scope for adding value to first-level patent data
 - Discuss some of the new content that is fast emerging
 - Issues faced by users of full text databases
 - Offer insights on overcoming those challenges



Brief Introduction – Richard Garner

- Current position:
 - Product Director IP Research Solutions LexisNexis
- Previous positions:
 - Senior Vice President Business Development Univentio Information BV
 - Director Omnipatent Ltd
 - Director Derwent First Level Data Division
 - General Manager Derwent Business Solutions
 - General Manager Rapid Patent, Crystal City VA
 - General Manager Derwent International Patent Document Delivery Service
- Began career in patent information at the British Library in mid 1980s
 - Established PatentExpress
 - Leading provider of patent document delivery services



LexPat – The emergence of full text

- Prior to 1983, subject-based text was usually limited to the title
- There were no full text databases:
 - Did not exist in machine-readable form or
 - The cost of computer storage made it prohibitively expensive
- Patent abstracts first appeared on SDC (Orbit)
 - Derwent World Patents Index file from 1976
- Lexis (Mead Data Central), launched Lexpat in March 1983
 - First example of full text patent database
 - Granted US patents for the period from 1976 onwards
- Each year's specifications were loaded into a separate files
 - Indexes generated enormous volumes of data
 - The slow search speeds available made analysing documents on screen relatively easy, as the text literally built up as you watched
- By the late 1980s, systems had improved sufficiently to be able to add clipped images from the drawing pages of patent specifications
 - Still with very limited means for viewing them online



Back File Conversion Project – What was it?

- In June 1986 the Administrative Council of the European Patent Organisation approved:
 - Award of a contract for the capture of the complete European PCT minimum documentation, contained in the EPO search collection, to a consortium of three firms.
- The BACON or <u>BAckfile CON</u>version project covered the digitisation and capture of the complete EPO back file:
 - Joint initiative between the USPTO, JPO and the EPO
 - Scan the full text and drawings of the first publication of patent documents by the major industrial property offices, back to 1920 or earlier
 - Image data (in facsimile form) on magnetic tape so as to permit the subsequent creation of an image retrieval system
 - First phase 1920 1987 comprised 125 million pages
 - 12 million documents stored on 65,000 magnetic tapes
 - Began early in 1987 and estimated to take about three years to complete
 - Resulting in some 13 Terabytes of data
 - The back file data available to the examiners, other national offices for their internal use and eventually disseminated for public information



EPOQUE/BNS – Merging text and images

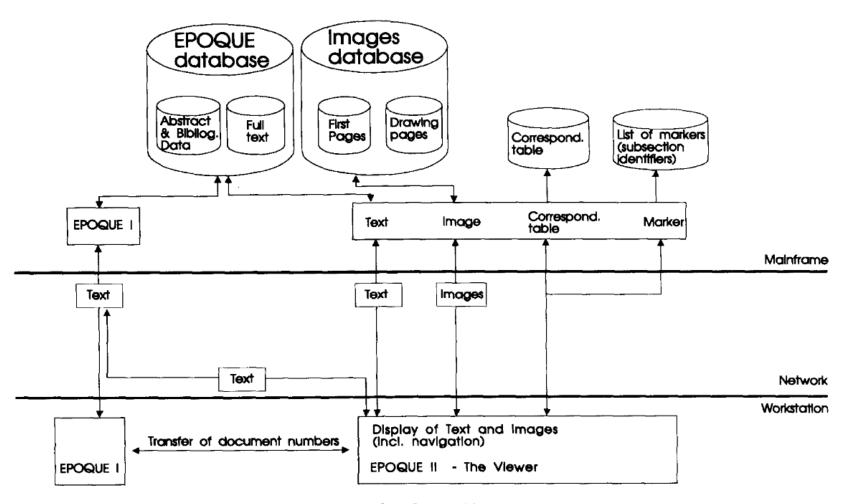
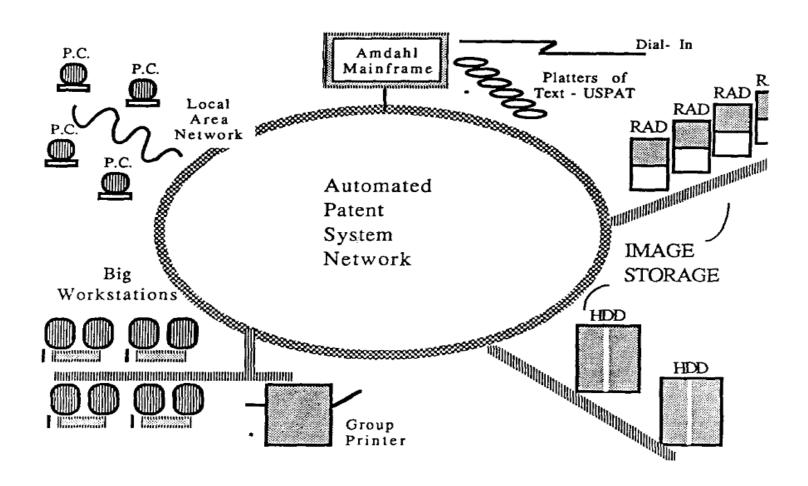


Fig. 2. EPOQUE II architecture.



USPTO APS – Merging text and images

Automation of USPTO





The Wilderness Years – 1983 – 2003

- US applications and granted patents full text
 - 1976/2001 present (USPTO)
 - 1836 present (commercial providers)
- EP applications and granted patents full text
 - 1986/1991 present (EPO)
 - 1978/1980 present (commercial providers)
- WO applications full text
 - 1978 present (commercial providers)
- JP applications, granted patents and utility models full text (in Japanese)
 - 1993 present (JPO)



The Present – US full text and images

	Patent Grant Multi- Page Images	Patent Grant Single- Page Images	Patent Grant Full Text with Embedded Images	Patent Grant Full Text	Patent Grant Bibliographi c Data	Patent Grant OCR Text
Beginning	1790	Current year	2001	1976	1976	1920
End	Present	Present	Present	Present	Present	1979
Issue Frequency						
Weekly	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	
Front-Page Info	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	$\sqrt{}$
Specification	\checkmark	\checkmark	\checkmark	\checkmark		$\sqrt{}$
Claims	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
Embedded Images			\checkmark			
Drawings	\checkmark	\checkmark	\checkmark			
ASCII				\checkmark	\checkmark	\checkmark
SGML			\checkmark	\checkmark	\checkmark	
XML			\checkmark	\checkmark	\checkmark	
TIFF	\checkmark	\checkmark	\checkmark			
PDF						



The Present – ROW full text databases





Full Text – The case for the prosecution

- Full-text databases for chemical information: do searchers need them?
 - Online '84 Conference Proceedings
 - With so many thoroughly-indexed databases available, do chemical searchers need full text?
- Lack of discrimination:
 - Every word (however common or unusual) has equal retrieval value, which may result in poor retrieval unless other strategies are put in place to include an implicit term weighting
- Information overload:
 - Full text has a tendency to err on the side of retrieving too large sets, with a consequent hazard that relevant answers may be missed in the analysis.
- Poor precision:
 - In many searches, common terminology occurs with such high frequency that it ceases to be a usable retrieval mechanism.
- Pressure at the point of search:
 - The lack of a controlled vocabulary means that good retrieval relies upon good 'post-coordinate indexing', i.e. it is up to the searcher to be able to select current and appropriate terminology and all possible variations/synonyms.



Full Text – The case for the defense

- The advantages may be summarised as:
 - Simplicity: any user who has experimented with a simple search engine is able to make sense of full text searching approaches.
 - Cost-efficiency: the overheads involved in creating new abstracts and indexing look less viable options for the production and management of a database.
 - Reduction in indexing errors: little or no manual indexing, the scope for human indexing errors is reduced.
 - Comprehensive retrieval: full text retrieval draws from the entire disclosure and ensures a much better recall of unusual terms which may only occur rarely in a document
 - Fewer skills to maintain: with full text searching, there is no need for the searcher to be continually updating their knowledge of subject heading lists or similar controlled vocabulary



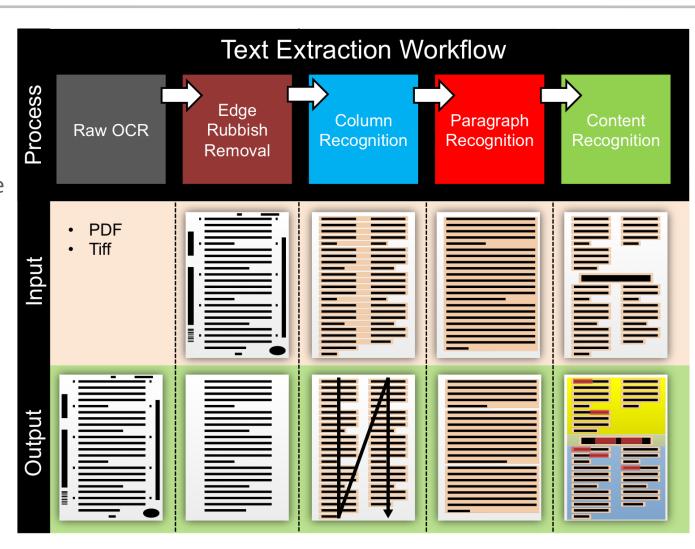
Full Text - In mitigation m'Lud

- Searchable segments
 - Full text
 - Title and/or Abstract
 - Descriptions, Claims
 - First Claim
 - Independent Claims
 - Patent Object, Advantages and Drawbacks
 - Examples
- Boolean, Proximity Operators and Wildcards:
 - AND, OR, AND NOT
 - W/n, W/s, W/p, Not W/n, Pre/n
 - ALLCAPS, NOCAPS
 - ATLEAST
 - PLURAL and SINGULAR
 - Phrase searching, stemming etc.
 - Wildcards and (SLART) truncation to represent one or more characters





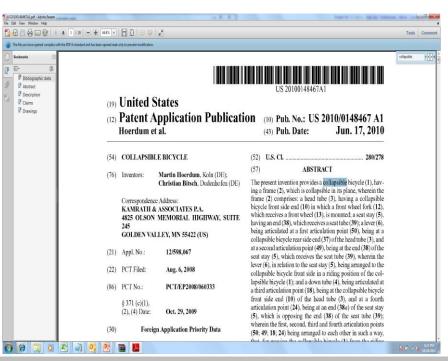
- Vendors overcome lack of full text data by OCR.
- Quality of PDF and different layouts make this task more difficult.
- Automatic layout recognition allows vendors to overcome format changes
- Maintaining paragraphs and basic character attributes increases readability





Value Add - PDF/XML

 Compressed, English bookmarks, searchable PDFs



q id='p-0003-en' num='0002'>The present invention relates to a collapsible bicycle. on id="n-0004-en" num="0003">2. Description of the Related Art apid='p-0005-en' num='0004'>Modern collapsible bicycles have to fulfill a variety of demands. These are in particular a beneficial volume of the collapsed collapsible bike, a high frame stiffness as well as a collapsible mechanism, which enable easy collapsible of the collapsible bicycle. op id="p-0006-en" num="0005">Collapsible bicycles known from prior art cope with these demands only in an insufficient manner. ¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.
¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.

¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.

¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.

¬ id="p-0007-en" num='0006'>Hence, the present invention is to provide an improved collapsible bicycle.

¬ id="p-0007-en" num='0006'>Hence, the present invention inve do id="p-0008-en" num='0007"> According to the present invention this problem is solved by a collapsible bicycle with the characteristics of claim cheading id="h-0002-en" level="1">SUMMARY OF THE INVENTION 👽 id="p-0009-en" num="0008">Accordingly a collapsible bicycle, having a frame, which is collapsible in its plane is provided, wherein the frame comprises; a head tube having a collapsible bicycle front side end in which a front wheel fork, which receives a front wheel, is mounted; a seat stay having an end, which receives a seat thube; a lever being articulated at a first articulation point, being at a collapsible bicycle rear side end of the head tube, and at a second articulation point, being at the end of the seat stay, which receives the seat tube, wherein the lever in relation to the seat stay being arranged to the collapsible bicycle front side in a riding position of the collapsible bicycle; and a down tube being articulated at a third articulation point, being at the collapsible bicycle front side end of the head tube, and at a fourth articulation point, being at an end of the seat stay, which is opposing the end of the seat tube; wherein the first, second, third and fourth articulation points being arranged to each other in such a way, that for moving the collapsible bicycle from the riding position into a sliding position or from the sliding position into the riding position, respectively, the end of the seat stay, which receives the seat tube, being shiftable in direction towards the head tube or in direction away from the head tube, respectively. q id="p-0010-en' num="0009"). The idea of the present invention is based on the fact that a very convenient volume of the collapsible bicycle is reachable when the seat stay with the seat tube is shifted approximately in direction towards the o id='p-0011-en' num='0010'>Advantageous embodiments and improvements of the present invention are to be found in the dependent claims. op id="p-0012-en" num="0011">The term "end" is at present also to be comprehended as an end section. opid='p-0013-en' num='0012'>The term "stay" is at present to be comprehended as an elongated connecting element. Preferably at least one of the stays is provided as a tube. o id='p-0014-en' num='0013'>Accessory or alternatively the stays may be bifurcated in appropriate areas. φ id="p-015-en" num='0014">The head tube comprises at present a possible front tube for supporting the front tube for wheel <|p> φ id="p-0016-en" num='0015">According to a preferred embodiment, the collapsible bicycle further comprises a rear wheel for h, which receives a rear wheel being articulated at a fifth articulation point being at a collapsible bicycle rear side end of the down tube, and a rear wheel stay being articulated at a sixth articulation point being at the collapsible bicycle rear side end of the head tube, and at a seventh articulation point being at the rear wheel fork, wherein the first, second, third, fourth, fifth, sixth and seventh articulation points being arranged to each other in such a way, that, the rear wheel shifts by means of shifting the end of the seat stay, which receives the seat tube, in direction towards the head tube or in direction away from it, respectively, in direction towards the front wheel or in direction away from the front wheel, respectively.
q id=p-0017-en' num='0016' >Advantageously, it is possible by means of the preceding described embodiment that just by shifting the upper end of the seat stay, which receives the seat tube, in direction towards the head tube, the rear wheel is shifted in direction towards the front wheel at the same time. The collapsible bicycle is then in a sliding position, this means, that the front and the rear wheel are furthermore in ground contact and are at the same time rotatable while it is possible for the user to slide and to steer the collapsible bicycle by means of the handlebar. Hence, the collapsible bicycle, according to the present invention, is already provided in the sliding position with a very compact shap that enables the user to entrain the collapsible bicycle in crowded spaces like the underground railway or the departure platform, wherein the user can slide the collapsible bicycle and has not to carry it, like it is required with the collapsible 🕫 id="p-0018-en" num="0017" In another preferred embodiment of the collapsible bicycle according to the present invention, the down tube comprises an end section being bent towards the head tube, at which the seat stay is articulated. The bent end section is preferably bifurcated. This embodiment enables a shifting of the seat stay in direction towards the head tube in such a way, that, the seat stay is essentially aligned over its whole length parallel and adjacent to the head qi d='p-0019-en' num='0018'>In another preferred improvement of the collapsible bicycle according to the present invention, a lug, at which the rear wheel fork is articulated, protrudes from the bent end section. This advantageor determines that the rear wheel fork is able to be widely shifted in direction towards the collapsible bicycle front side in such a way, that, the rear wheel is positioned as near as possible to the front wheel. The lug thereby enables, in section an almost parallel alignment of the rear wheel fork relatively to the down tube. φ id="p-0020-en" num="0019">According to another preferred embodiment of the collapsible bicycle according to the present invention the seat stay comprises a fork-shaped section, through which the head tube extends. id="p-0021-en' num='0020'>In the riding position of the collapsible bicycle the seat stay extends essentially from down to head (in the present application "down" and "head" refer to the ground). Thereby, the seat stay intersects the head tube, which extends essentially from the collapsible bicycle front side to the collapsible bicycle rear side. For providing a high frame stiffness, the seat stay is therefore provided with a bifurcated section and encompasses the head tube. When shifting the head end of the seat stay, which receives the seat tube, in direction towards the head tube, the seat stay is able to move unhindered relative to the head tube, (n) op id="p-0022-en" num="0021". In another preferred improvement of the collapsible bicycle according to the present invention, the head tube comprises a recess, in which the lever is concealed in the riding position of the collapsible bicycle φ id="p-0023-en' num="0022"/In another preferred embodiment of the collapsible bicycle according to the present invention, the seat stay comprises a support element and the head tube comprises a receiving element which provides a form closure in one plane, which extends essentially perpendicular to the seat stay. Consequently, in the riding position high transverse loads, resulting from riding on bumpy ground can be beard from a seat, which is attached to the seat 💠 id="p-0024-en" num="0023" : Initially, the support element and the receiving element are mainly intended to avoid a shifting of the seat stay by means of the lever relatively to the head tube towards down, thus, when the seat is loaded by the ascension of an user. Accessory, the stabilization according to the preceding embodiment is supposed to be provided in a transverse direction. op id="p-0025-en" num="0024" >Preferably the support element has a hollow conical shape and the receiving element has a conical shape, wherein they match together in the riding position of the collapsible bicycle. By means of this embodiment of the support element and the receiving element, a load bearing in a transverse direction and also the bearing of the gravitational force resulting from a user is enabled. φ id='p-0026-en' num='0025'>According to another preferred embodiment of the collapsible bicycle according to the present invention, the collapsible bicycle comprises a carrier and at least one carrier support stay, wherein the carrier is articulated at an eight articulation point, being at the seat stay, and the carrier support stay being articulated at a ninth articulation point, being at the end of the head tube, and at a tenth articulation point, being at the carrier, wherein the eight, ninth and tenth articulation points being arranged to each other in such a way that the carrier is provided in the riding position and in the sliding position with an essentially horizontal position. φ id='p-0027-en' num='0026'>Hence, it is guaranteed that luggage, being clamped by the carrier, stays on the carrier and does not glide off it, while the collapsible bicycle is moved from the riding position to the sliding position. This is especially beneficial for bags, which are fixedly mounted to the carrier. a id=5-0038-en' num=0027'-According to another preferred improvement of the collapsible bicycle according to the present invention, the seat stay is compoundable to the head tube for the riding position of the collapsible bicycle by means of locking a connecting element, being arranged at the seat stay, and the seat stay is releasable for a shifting of the seat stay in a first shifting direction relatively towards the head tube in a plane of the frame for the sliding position of the collapsible bicycle by means of loosening the connecting element. The connecting element ensures in the locked position that a shifting of the seat stay relatively to the head tube is avoided. Such a connecting element enables in an easy way to turn the frame, which is stiff in the riding position, to be collapsible in itself for the sliding position. q id='p-0029-en' num='0028'>According to another preferred improvement of the collapsible bicycle according to the present invention, the connecting element comprises a release handle, being articulated at the seat stay, and a hook, being

> XML standardized to WIPO ST.36 format



Value Add – Forward Citations

- Vendors have the option to add value across documents such as Forward Citations.
- Standardization of publication numbers is required to be able to link to additional information like titles.
- Some vendors merge non IP data with content such as Legal, News, Business etc. information



Retrieve Selected

	Publication Number	Publication Date	Title
	GB2467819A	2010-08-18	Plural sliding housings each with a fan and air channel for an array of data storage elements
	US6924439B1	2005-08-02	Signal conducting applique and method for use with printed circuit board
	US7145770B1	2006-12-05	Method and apparatus of packaging disk drives in a data storage system
	US7061715B2	2006-06-13	Disk array apparatus
	US6795309B2	2004-09-21	Tool-less field replaceable peripheral mounting system
	US7414837B2	2008-08-19	ATCA board compatible hard disk mounting structure
	US7403390B2	2008-07-22	Ultra dense multipurpose server
	US7361081B2	2008-04-22	Small form factor air jet cooling system
	US7362565B2	2008-04-22	Disk drive support system
_	US7304841B2	2007-12-04	Connection arrangement of blade



Value Add – Machine Translation

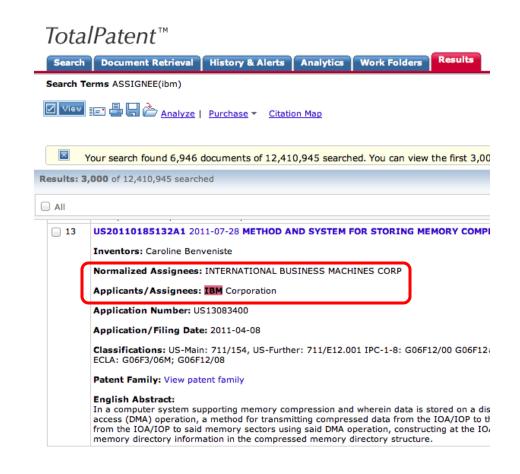
- Customers want help from vendors in finding a more comprehensive result by translating "foreign" full-text to English.
- Getting the concepts right by training translation engines with IP and other scientific material to make documents easier to find.
- Getting the language right by using IP to create the language-model will improve readability.
- Engines need to be fine-tuned by "manually" adding new words.
- Also a continuous cycle of measuring, training, improving software and retranslating the database

Source	нт	MT
多チャンネル化して も光パワーモニター は小型化が可能であ る。	If multiple channels the optical power monitor can be reduced in size.	The multi-channel optical power monitor can be reduced in size.
解決手段:エネルギ 吸収部材10は周壁 12を有する筒状の 本体11を備える。	SOLUTION: The energy absorbing member 10 includes a cylindrical main body 11 having a peripheral wall 12.	SOLUTION: the energy absorbing member 10 has a cylindrical body 11 having a peripheral wall 12.
課題:設置スペース を小さくして据え付 け安定性を良好にで きるコンパクトで冷 却性能の高い冷却ユ ニットを提供する。	PROBLEM TO BE SOLVED: To provide a compact cooling unit with high cooling performance, with improved installation stability by reducing the installation space.	PROBLEM TO BE SOLVED: reducing the installation space to improve the stability of the installation can provide a compact cooling unit with high cooling performance.



Value Add – Smart Indexing

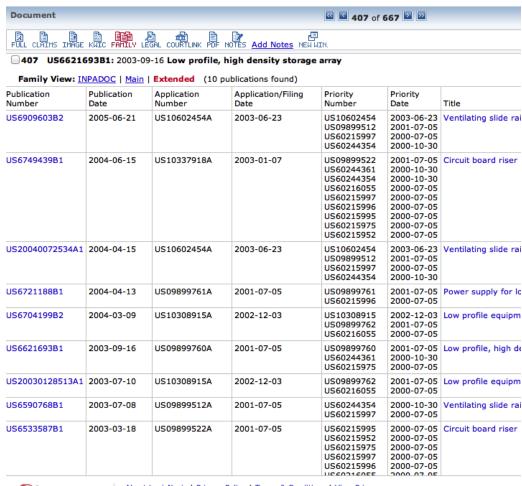
- By normalizing assignee names, compensate for name variations (abbreviations, translations / transliterations, typos etc.)
- Reassignment data (US applications assigned to inventors)
- Determine probable patent assignee using machine learning algorithms and family information





Value Add – Patent Families

- Vendors organize records into families: domestic, strict, extended etc.
- Collapse results by family
- Some vendors use
 (domestic) family to
 enhance publications, by
 "adding" legal status across
 every stage of a publication.
- Standardization of application/priority numbers



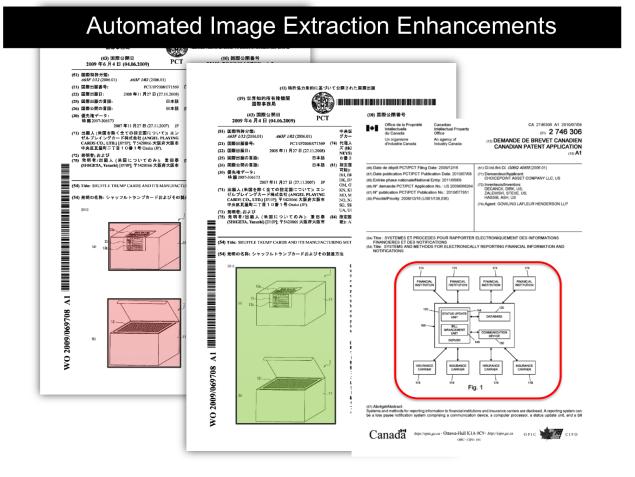


About LexisNexis | Privacy Policy | Terms & Conditions | View Price Copyright© 2011 LexisNexis, a division of Reed Elsevier Inc. All rights reserved



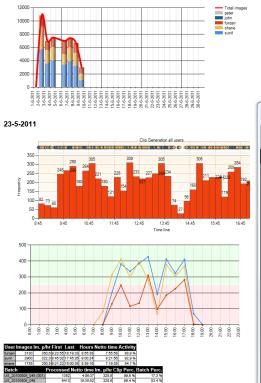


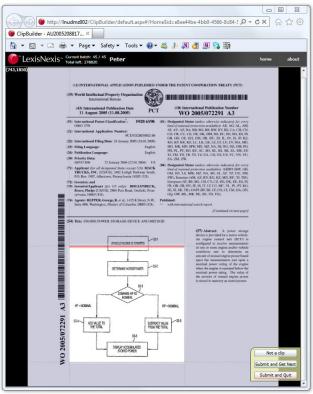
- Extracting clipped images for older and/or more "exotic" authorities.
- Create algorithms to combine "separated images".
- Optimizing extraction programs to handle special types of clips, like flow-charts.

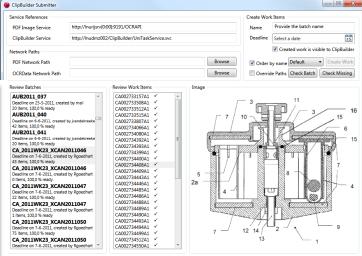




LexisNexis Univentio – Clip Builder









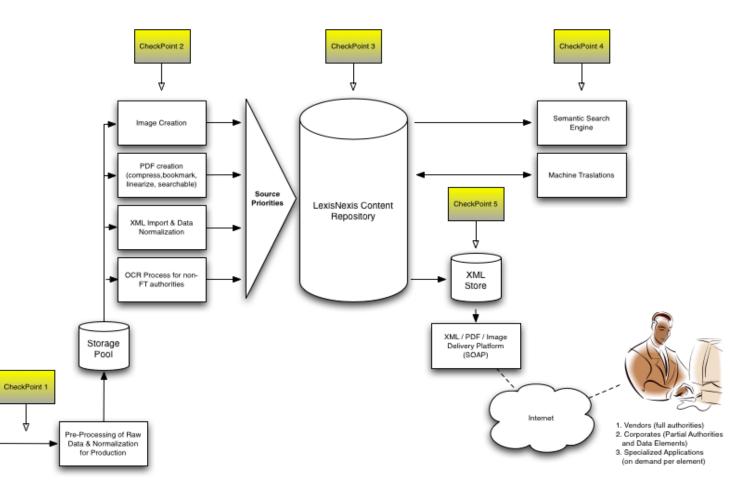
LNU / Systems Overview

LNU Facts & Figures

- Employees: 25
- Located Rijnsburg NL

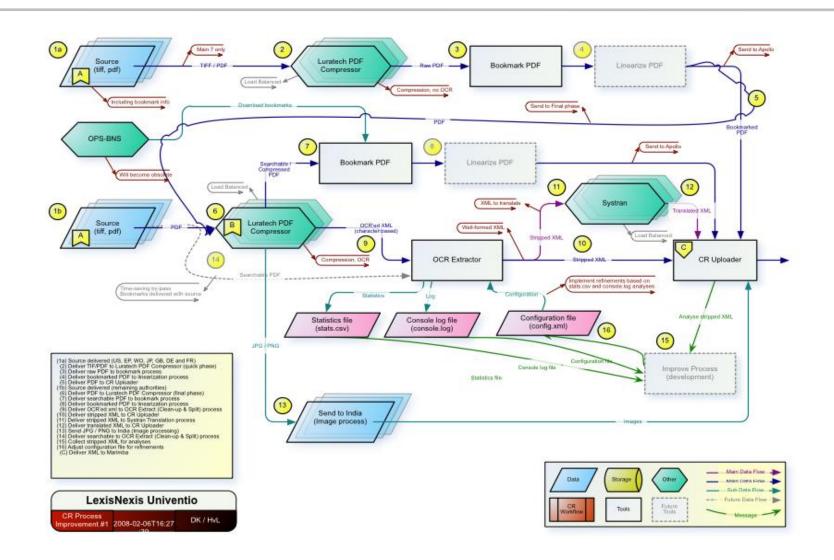
Data Processed / Stored:

- > 80 Mio Publications
- > 55 Mio PDFs
- 200+ Mio Images
- 100 Authorities
- 30 full-text
- Processing:
- 450,000 per day
- 150,000 OCR per day
- 100,000 MAT per day
- High update frequency



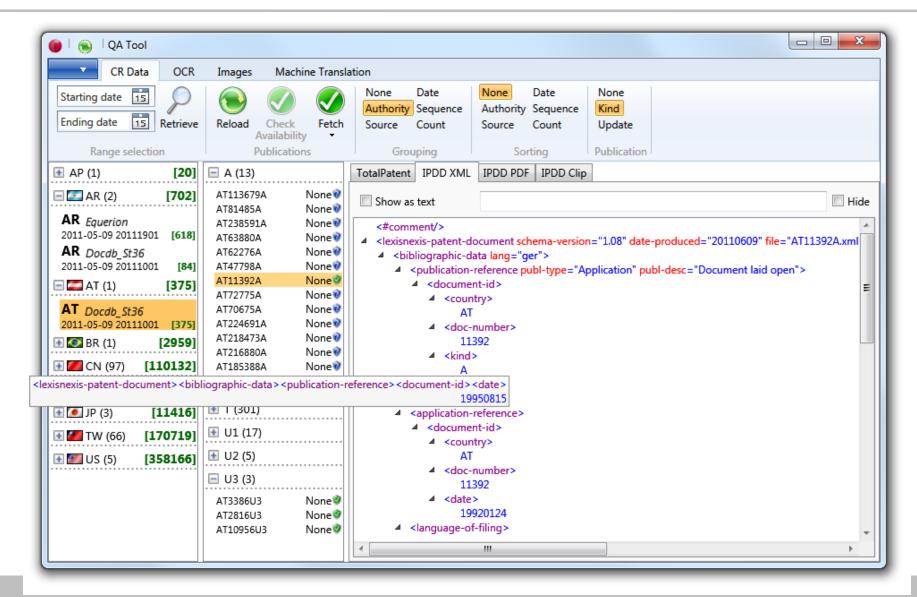


Data Fabrication - PDF/OCR



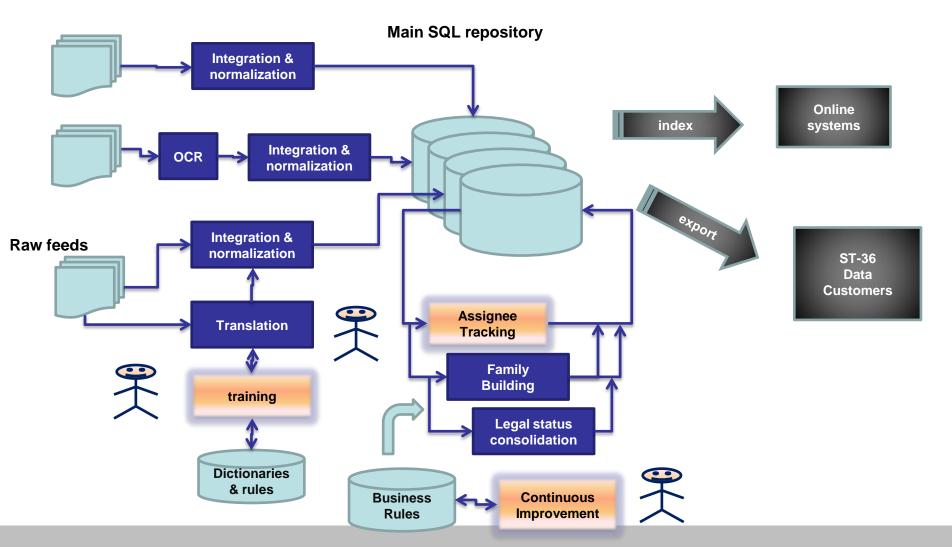


LexisNexis Univentio – QA Tools





Questel – Patent Data Refinery







- Past 25 years has seen a revolution in the IP market.
- After a slow start rapid growth in and availability of full text databases
- Similarly, increase in the number of patent information vendors
- Some patent issuing authorities have not covered themselves in glory
- More authorities now publishing front files very few publish back files and almost no comprehensive "official" full text collections exist
- Industry has often led the way utilizing software technology like OCR, MT and XML.
- It's a dynamic field. Offices change formats, create new classification systems etc. Industry has to respond in order to remain competitive and look for increasingly innovative ways to add value.
- The next 5 year promises to be every bit as exciting.



Acknowledgments

- Eric van Stegeren: LexisNexis
 - Maximizing Value, Minimizing Cost How technology is making value added data affordable for both consumers and vendors of patent information EPOIC Oct 2011
- Laurent Hill: Questel
 - From patent data to IP knowledge How Commercial Patent Information Providers refine crude data, yielding high value patent insights EPOIC Oct 2011
- Stephen Adams: Magister Ltd
 - Information Sources in Patents K.G. Saur
- Michael Blackman: World Patent Information
 - Various articles Elsevier B.V.





Thank You!