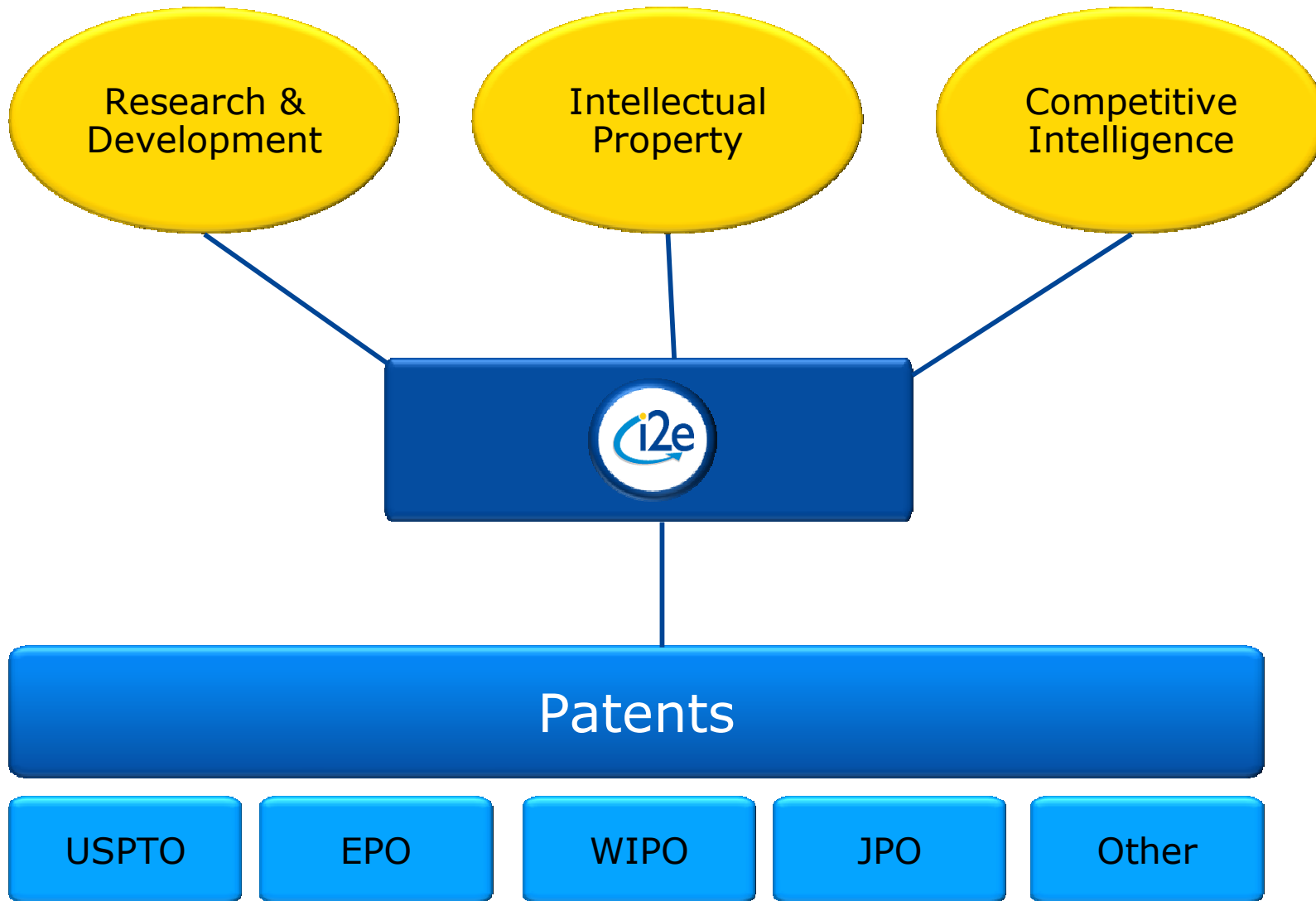# I2E for Patent Mining
ICIC

David Milward

October 2010

Linguamatics

I2E for Patent Mining

Mining Patent Documents

- Patents as a knowledge source

  - R&D

  - Commercial intelligence

- Finding and reviewing relevant patent documents

  - IP

    - Finding the key patent for prior art search

    - Categorising patents to prioritize review

    - Profiling to enable rapid filtering

Linguamatics

Patents as a Knowledge Source

- Find information reported in patents but not the scientific literature

  – Individual ad hoc searches

  – Generation of database content from patents

- Link between entities in patents, or between patents and scientific/internal literature/databases to pick up indirect relationships

- Integration with workflow tools e.g. Pipeline Pilot for:

  – Automated analysis e.g. specialised email/RSS alerting

  – Dashboards e.g. monthly trend analysis

Linguamatics

R&D: What coatings are used for capsules?

- Find list of chemicals used in coatings

- Pick out all elements mentioned in the list of coatings e.g. shellac, sugar or gelatin

Gelatin	▶ 3	7514267	1	... capsules may be coated with gelatin, wax, shellac or ...
nanoparticle	1	7597950	1	... the capsule is coated with nanoparticles using the techniques of the ...
Polymer	1	7588779	1	... core is preferably coated with a polymeric coating to form a membrane around ...
Shellac	▶ 18	7473760	1	... tablets may be coated with shellac, sugar or both.
sugar	▶ 18	7473760	1	... be coated with shellac, sugar or both.

Competitive Intelligence Examples

- What are competitors key areas

- Examine competitors/acquisitions/licencees patent portfolio & activity

- What are the trends in patenting for a particular disease, therapy area etc.

- Find candidate technologies for in-licensing

- Who are the key players in particular areas

Company Patenting Activity

- Company activity in patenting for chemotherapeutics
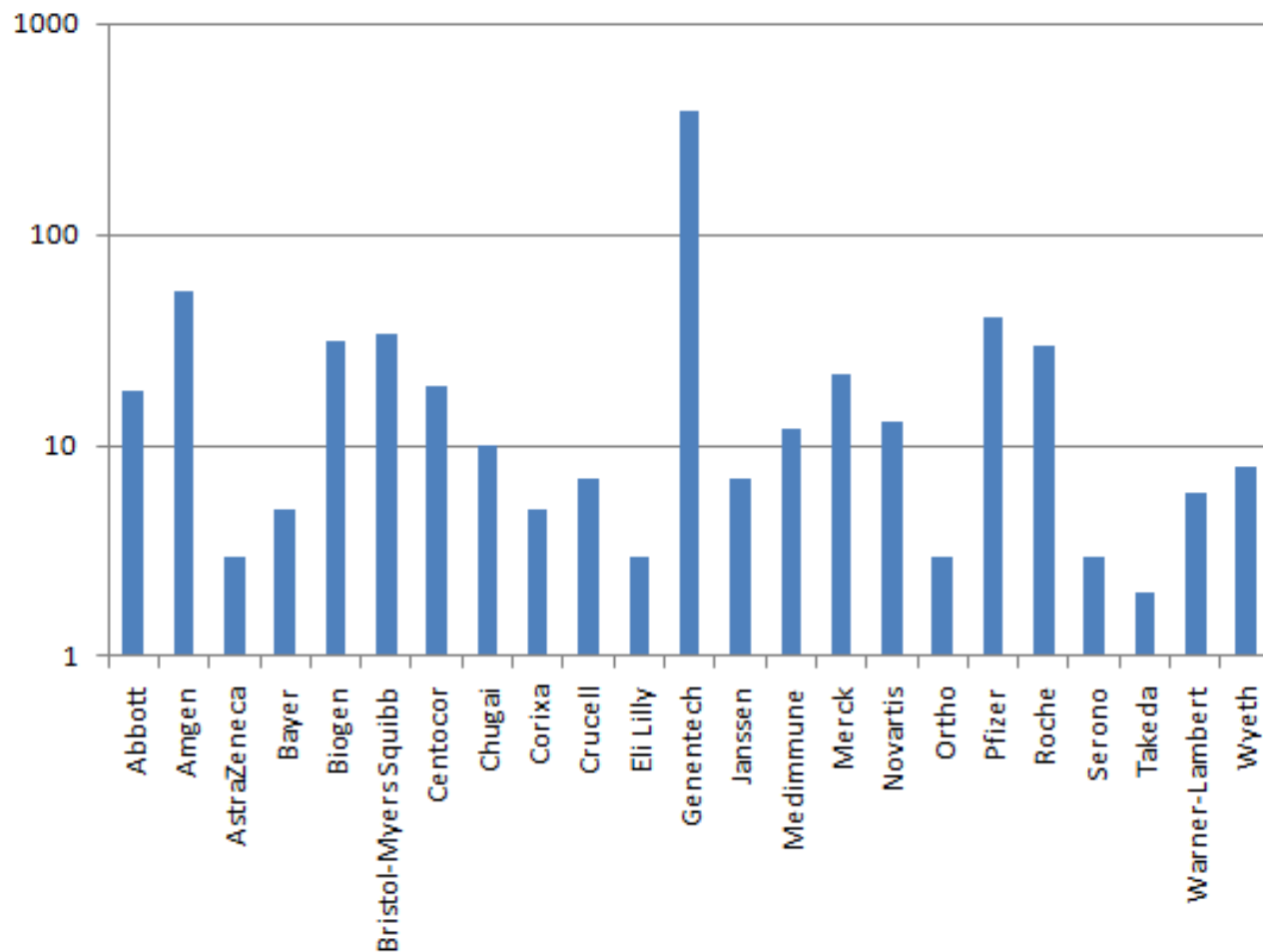
Finding Relevant Patents

- **Terminologies**
  - Can incorporate large-scale built in terminologies
  - Use I2E to **generate** vocabulary from the data

- **Ability to run high throughput searches**
  - Alternative lists of 500 genes, chemicals etc.
  - Any neoplasm etc. using ontologies
  - Chemical substructure and similarity

- **Precision**
  - Linguistics
  - Filtering via local negation
  - Search within regions
  - Search within the same list (limited length)

Linguamatics
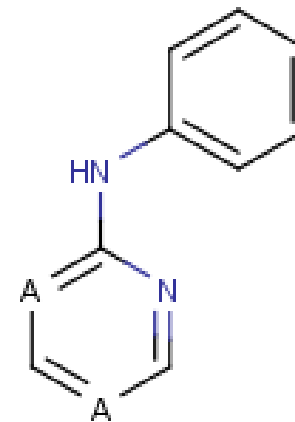
Using Terminologies

- I2E allows users to plug in public domain, commercial or proprietary terminologies e.g.

    - MedDRA, Entrez Genes, MeSH, NCI

- Can search for individual concepts, or whole classes of concepts e.g. any kinase, any skin disorder

    - A single class may be searching for 10s K concepts and all their synonyms

- Users can also create their own run-time hierarchical terminologies, or input lists of hundreds of alternative terms

Filtering via local negation

- I2E allows very precise filtering of unwanted hits

- Typical search engines have document level negation which is not precise enough

- I2E linguistic processing can filter out e.g.

  - "Mole concentration" for nevus

  without ruling out a document containing both "mole" and "concentration". Similarly, it can distinguish:

  - "Family history of cancer"

  - "History of cancer"

  - "No history of cancer"

Linguamatics

Co-occurrence within Claims

- Can go from very precise sentence co-occurrence to looking for co-occurrence somewhere within the Claims, or stages inbetween e.g.

    - Sentence within claims

    - Individual claim

    - Pair of claims in a claim chain

    - Claims as a whole

Linguamatics

Claim Pair

- A claim is often dependent on another claim

- Useful to find co-occurrence within the pair of claims, even if they are not textually adjacent

US7635560	Breast Neoplasms	expression	▶ 20	1. A method for determining the presence or absence of a cancer in a patient, wherein the cancer is at least one of cervical cancer, breast cancer, ovarian cancer and lung cancer, the method comprising the steps of: (a) determining the level of gene expression of hPygo2 as shown in SEQ ID NO: 1 or 2 in a biological sample obtained from a patient, and (b) comparing the level of hPygo2 gene expression in the biological sample to a predetermined cut-off value, wherein the predetermined cut-off value is the level of hPygo2 gene expression in a normal biological sample, to determine whether hPygo2 expression is higher in the biological sample; therefrom determining the presence or absence of cancer in the patient.
US7638122	Bile Duct Neoplasms	expression	1	1. A method for treating cancer in a subject, said method comprising: (a) decreasing Stat3 expression or function in a tumor cell ; (b) culturing the tumor cells obtained in step (a); (c) contacting antigen-presenting cells with supernatant obtained from the culture of step (b), wherein the supernatant is irradiated before contacting the antigen-presenting cells with the supernatant; (d) contacting cytotoxic T cells with the antigen-presenting cells of step (c), thereby activating the cytotoxic T cells, and (e) administering the cytotoxic T cells to the subject, such that the cancer in the subject is treated. ... 6. The method of claim 1 , wherein the cancer is selected from the group consisting of a fibrosarcoma, a myxosarcoma, a liposarcoma, a chondrosarcoma, an osteogenic sarcoma, a chordoma, an angiosarcoma, an endotheliosarcoma, a lymphangiosarcoma, a lymphangioendotheliosarcoma, a synovioma, a mesothelioma, an Ewing's tumor, a leiomyosarcoma, a rhabdomyosarcoma, a colon carcinoma, a pancreatic cancer, a breast cancer, an ovarian cancer, a prostate cancer, a squamous cell carcinoma, a basal cell carcinoma, an adenocarcinoma, a sweat gland carcinoma, a sebaceous gland carcinoma, a papillary carcinoma, a papillary adenocarcinoma, a cystadenocarcinoma, a medullary carcinoma, a bronchogenic carcinoma, a renal cell carcinoma, a hepatoma, a bile duct carcinoma, a choriocarcinoma, a seminoma, an embryonal carcinoma, a Wilms' tumor, a cervical cancer, a testicular tumor, a lung carcinoma, a small cell lung carcinoma, a bladder carcinoma, an epithelial carcinoma, a glioma, an astrocytoma, a medulloblastoma, a craniopharyngioma, an ependymoma, a pinealoma, a hemangioblastoma, an acoustic neuroma, an oligodendroglioma, a meningioma, a melanoma, a neuroblastoma, a retinoblastoma, an acute lymphocytic leukemia, an acute myelocytic leukemia, a chronic leukemia, a polycythemia vera, Hodgkin's disease, a non-Hodgkin's disease, a multiple myeloma, a Waldenstrom's macroglobulinemia, and a heavy chain disease.

Adding Claim Chain Information

- For information in claims, often want to work back along the chain of claims

7582287	14. The method of claim 12 , wherein the chemotherapeutic agent is selected from the group consisting of bleomycin, chlorambucil, epirubicin, 5-fluorouracil, ifosfamide, mitomycin, methotrexate, vincristine, cisplatin and vinblastine.	12. The method of claim 10 , wherein the cervical cancer cells are treated with radiation in conjunction with the polypeptide.	▶ 10. The method of claim 5 , wherein the polypeptide comprises amino acid residues 1-176 of SEQ ID NO:2.	5. A method for inhibiting the growth and or proliferation of cervical cancer cells comprising bringing a polypeptide comprising amino acid residues 33-176 of SEQ ID NO:2 into contact with the cervical cancer cells.	1	7582287	1	... claim 12 , wherein the chemotherapeutic agent is selected from the ...
	7. The method of claim 5 , wherein the cervical cancer cells are treated with one or more chemotherapeutic agents in conjunction with the polypeptide.	5. A method for inhibiting the growth and or proliferation of cervical cancer cells comprising bringing a polypeptide comprising amino acid residues 33-176 of SEQ ID NO:2 into contact with the cervical cancer cells.			1	7582287	1	... treated with one or more chemotherapeutic agents in conjunction with the ...

Linguamatics

Document Review

- Can provide extra information for faster review of the patents found by a search

- For example:

  - Type of applicant e.g.

    - Commercial

      - Competitor

      - Potential partner

    - Academic ...

  - Priority date

  - Topics mentioned

Linguamatics

Adding Structured Information for Review

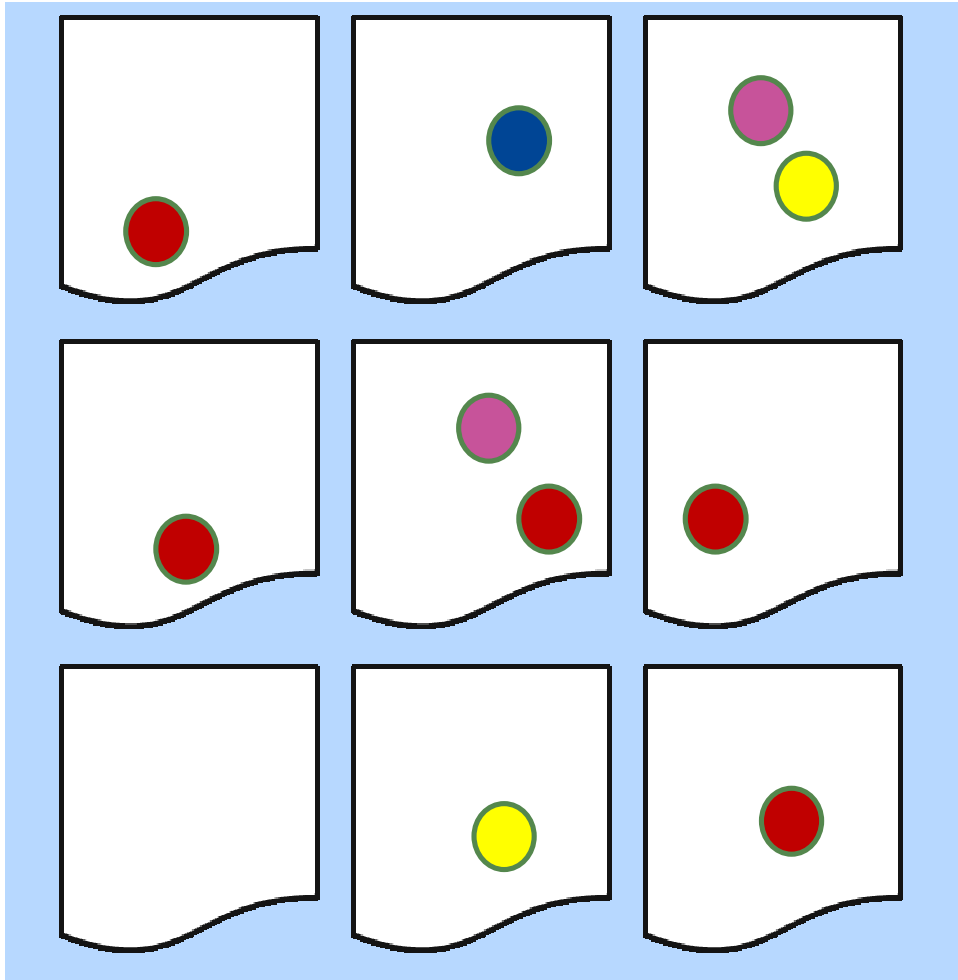|									
▶ Assignee	Corporation		1	7479276	1	... "Optimizing Liposomes for Delivery of Chemotherapeutic Agents to Solid Tumors," Pharmacol... of HER-2 antisense oligonucleotide and chemotherapeutic agents used for the treatment ... e.g., antifungal and conventional chemotherapeutic agents) are already on ... Tumor Cells Causing Sensitization to Chemotherapeutic Agents ... 7479276	1	7479276	1	SynerGene Therapeutics, Inc. 02 ... 7479276
▶ Priority Date	Feb. 22, 1999.		1	7479276	1	... "Optimizing Liposomes for Delivery of Chemotherapeutic Agents to Solid Tumors," Pharmacol... of HER-2 antisense oligonucleotide and chemotherapeutic agents used for the treatment ... e.g., antifungal and conventional chemotherapeutic agents) are already on ... Tumor Cells Causing Sensitization to Chemotherapeutic Agents ... 7479276	1	7479276	1	This application is a 371 filing of PCT/US00/04392, filed Feb. 22, 2000, which claims priority from U.S. provisional application 60/121,133, filed Feb. 22, 1999.... 7479276
Assignee	University		1	7479376	1	... is meant to refer to chemotherapeutics, toxins, radiotherapeutics, ... may be used to deliver chemotherapeutics, toxins, enzymes, ... 7479376	1	7479376	1	Thomas Jefferson University 02... 7479376
▶ Assignee	Corporation	Janssen	1	7479541	1	Many chemotherapeutic agents are designed to disrupt ... efficacy of their use as chemotherapeutic agents. ... use of lower doses of chemotherapeutic agents to effect greater damage ... sensitize cells for treatment with chemotherapeutics, radiation, DNA damaging ... pharmaceutical compositions can further include chemotherapeutic agents for the use in ... and in combination with other chemotherapeutic agents, and in a ... such as a DNA damaging chemotherapeutic agent, and a pharmaceutically ... repair inhibitor can enhance the chemotherapeutic effect of cytotoxic agents used ... 7479541	1	7479541	1	Janssen Pharmaceutica N.V. ... 7479541

Linguamatics

Categorizing Set of Patents

- By categorizing we can prioritize to certain sets of patents, and more easily see any duplicates

- Label each document according to its properties using rule-based categorization

- Often have positive categories: those we will definitely want to explore, and negative categories, those definitely do not need to look at

Linguamatics

Categorizing Documents into Exclusive Sets

|				
US7628988	Diabetes Type 1&2	Methods and compositions for treating type 1 diabetes	1	For the treatment of type I or type II diabetes, insulin-producing islet cells (e.g., islet cells expressing MHC class I that has at least one allele that matches the patient and that presents a peptide) or cells that differentiate into insulin-producing islet cells in vivo (e.g., cells originating from splenocytes, bone marrow origin cells, blood origin cells, or fibroblasts) can be transplanted into the patient.... The data suggest that the defect is widespread in Type 1 diabetes, with the majority of patients showing a detectable abnormality in T-cells (with a relatively larger fraction of T-cells with heightened TNF-α sensitivity).... Alternatively, such antibodies can be screened in vitro using assays in which agonists are identified by their ability to kill activated T-cells obtained, for example, from a patient with lymphoma or newly diagnosed type-2 diabetes. ...While the therapies described herein are likely to be effective in treating pre-diabetics, i.e., patients diagnosed as progressing to type I diabetes, but who are not yet hyperglycemic, we note that the methods of the inventions also may be used to treat a mammal, for example, a human with type I diabetes or any other autoimmune disease.... Accordingly, before beginning to treat a subject with type I diabetes or any other autoimmune condition, we may determine from blood analysis alone whether the subject has either or both of these two phenotypes and, therefore, is amenable to therapy.... In addition, in treating type I diabetes, it may be desirable that the patient maintains as close to normoglycemia as possible.
US7628993	Diabetes	Compositions and methods for vaccinating against HSV-2	1	Med. 4: 109-118 (1998)), AdvantaJet (Linmayer, I., et al., Diabetes Care 9:294-297 (1986)) Medi-jector (Martins, J., and Roedl, E. J. Occup.
US7629006	Diabetes	Process for preparing extract from rhus verniciflua and pharmaceutical composition containing the same	1	The sap of Rhus verniciflua is referred to as lacquer, and dried lacquer has been traditionally known in the Oriental medicine to have the functions of removing extravasated blood, promoting blood circulation and having efficacy against intestinal worms, abdominal pain, hyperacidity, coughs, tuberculosis, amenorrhea, constipation, diabetes mellitus, malaria and the like.... In addition, lacquer is known to have efficacy against gastric diseases, heart diseases, arthritis, hypertension, diabetes, stroke, arthritis, chronic fatigue and the like.
US7629119	Diabetes Type 1	Animal genotyping method	1	There are a number of reports indicating that the presence of β-casein A 1 in the human diet is linked with the incidence of certain diseases, specifically diabetes (Elliott, R. B., Harris, D. P., Hill, J. P., Bibby, N. J., Wasmuth, H. E., Type I (Insulin-Dependent) Diabetes Mellitus and Cow Milk: Casein Variant Consumption, Diabetologia 1999; 42:292-6; Wasmuth, H. E., Rosenbauer, J., Elliot, R. B., McLachlan, C., Erhardt, G., Giani, G., Kolb, H., β-Casein A1 Consumption and Incidence of Type 1 Diabetes in Germany.

Linguamatics

Directly Answering the Question

Which companies are mentioned in patents talking about chemotherapeutics?

Pharma Biotech Compa..	
Pharmacia	▶ 35
Roche	▶ 22
Becton Dickinson	▶ 21
Genentech	▶ 13
Merck	▶ 13
Sigma-Aldrich	▶ 12
Takeda	▶ 9
Amersham Biosciences	▶ 9
Bayer	▶ 8
Bristol-Myers Squibb	▶ 6
Genzyme	▶ 6

Linguamatics

Summarizing a Set of Documents
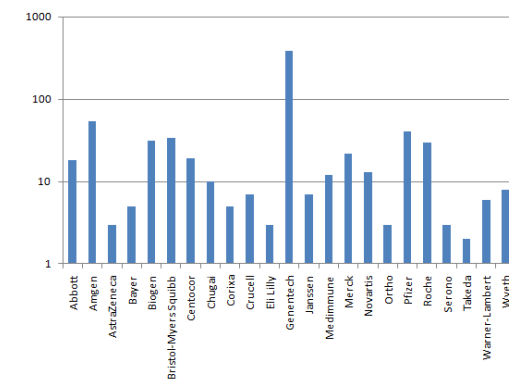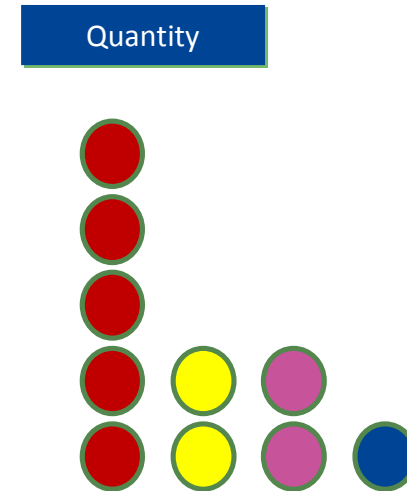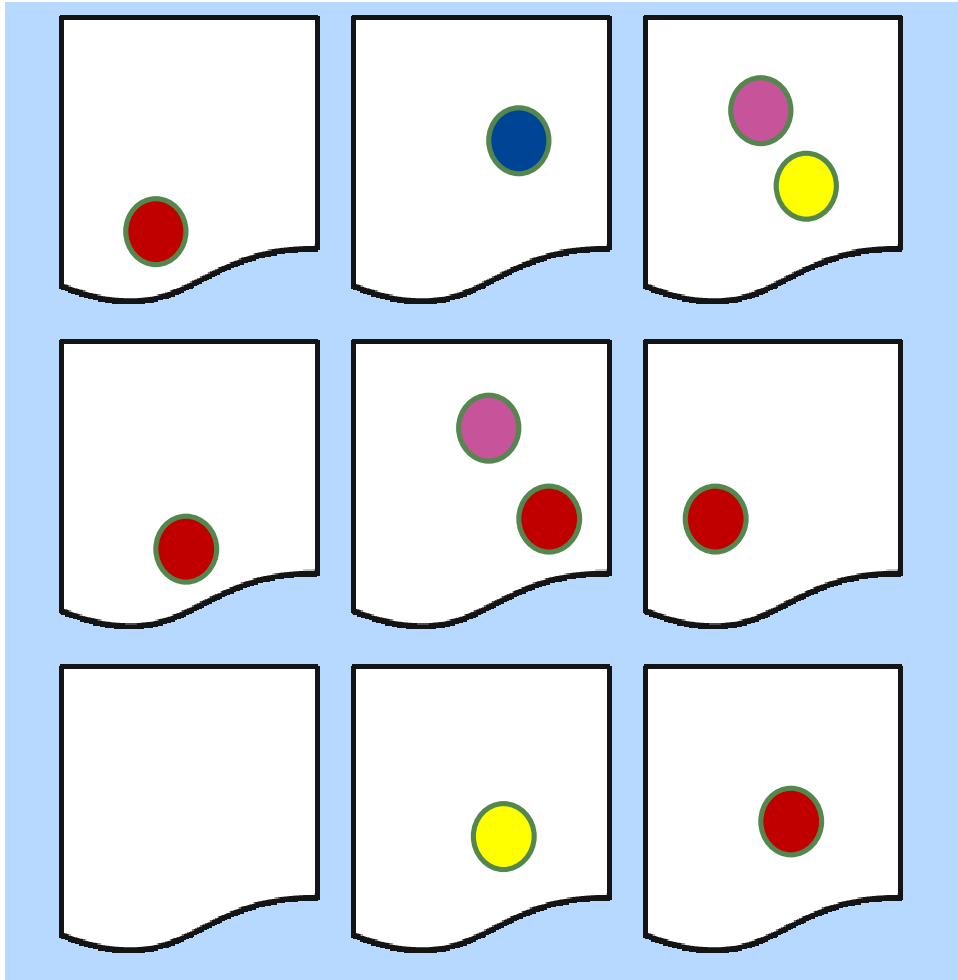
Summarizing Each Document
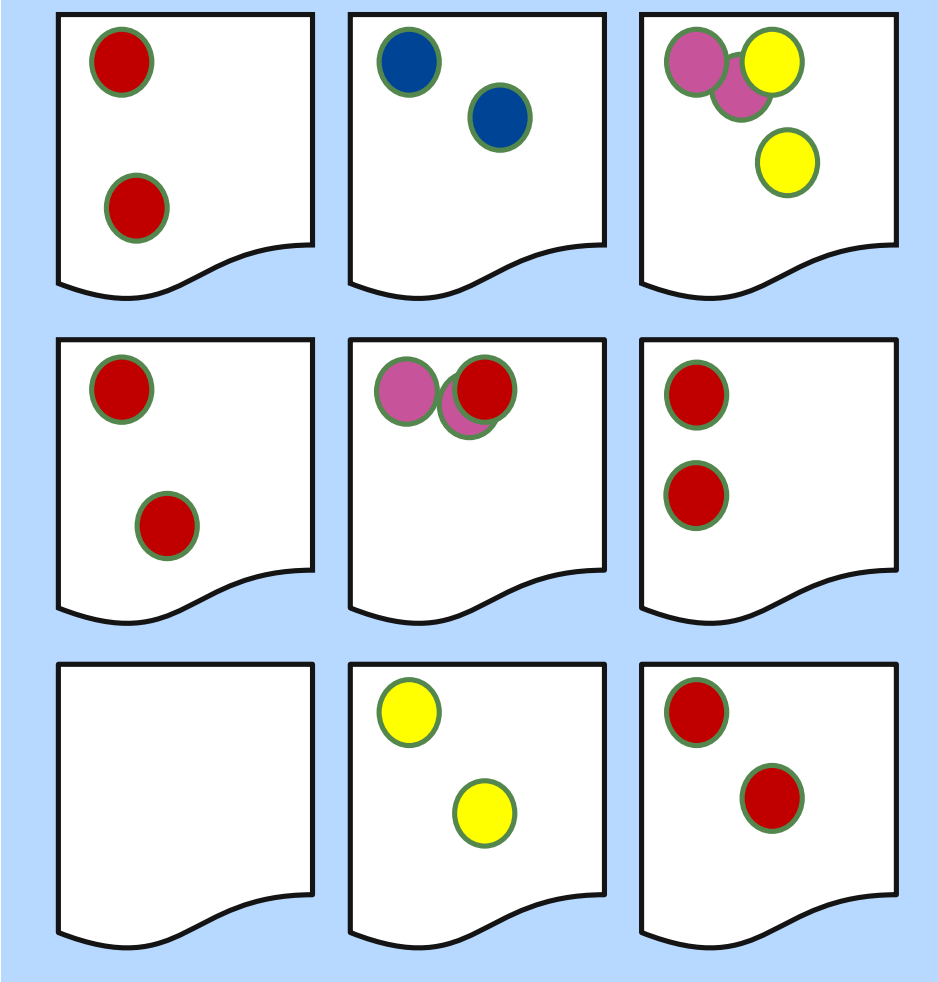
By Topic

US7629006	Diabetes	Process for preparing extract from rhus verniciflua and pharmaceutical composition containing the same	
US7629119	Diabetes Type 1	Animal genotyping method	

Document Profile

▶ Assignee	Corporation
▶ Priority Date	Feb. 22, 1999.

Clustering The Documents
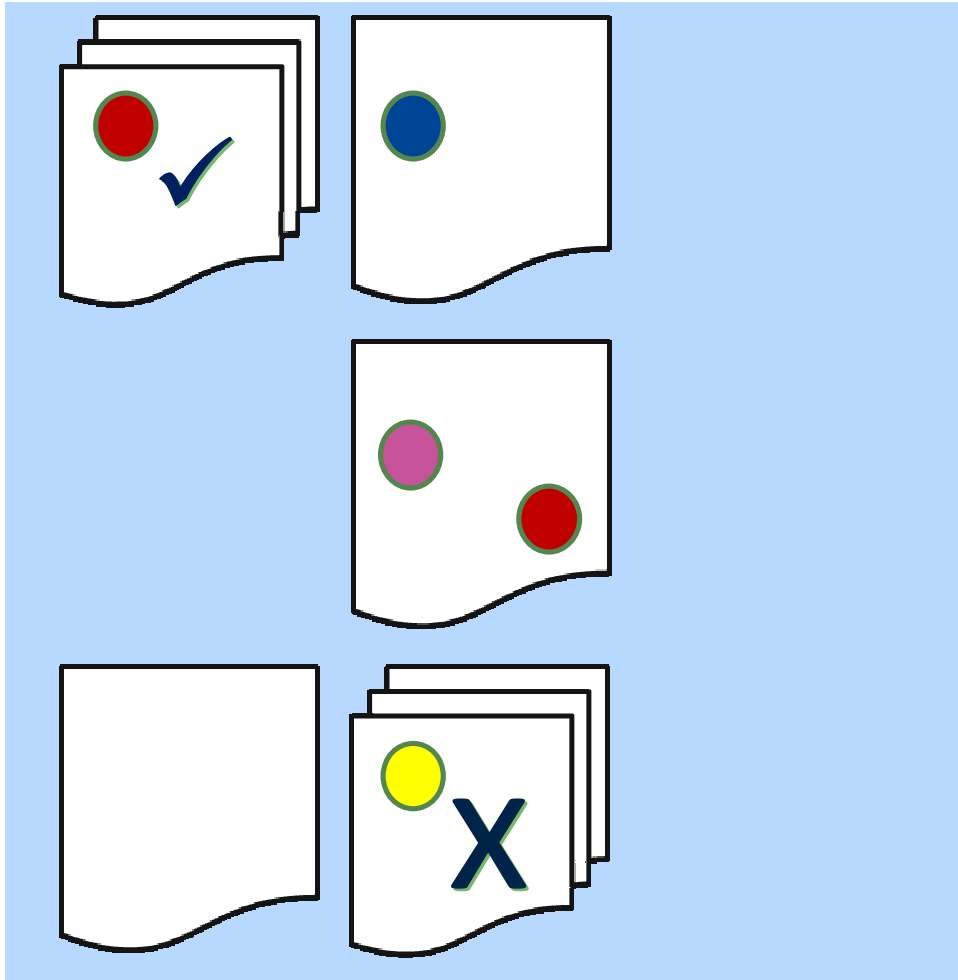
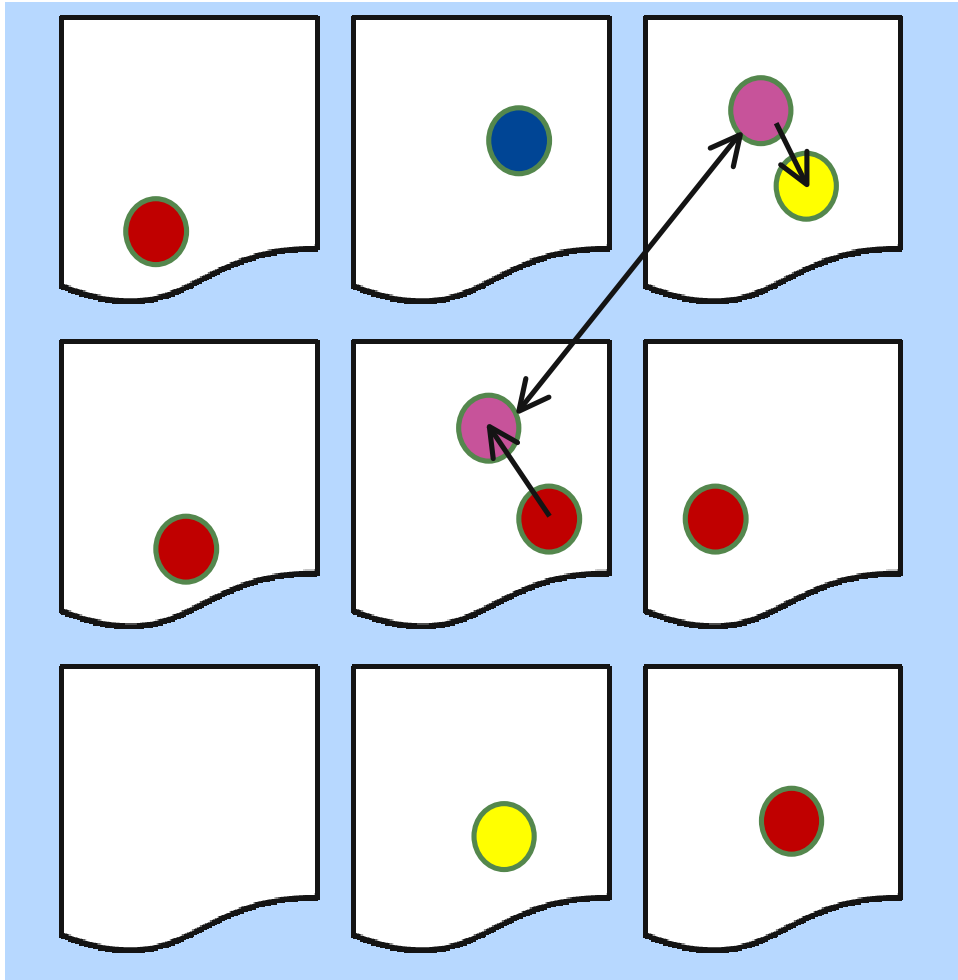| By catgegory		
Organization		**Doc**
Corporation	▶ 267	US7625612
University	▶ 42	US7625560
Institute	▶ 35	US7625201

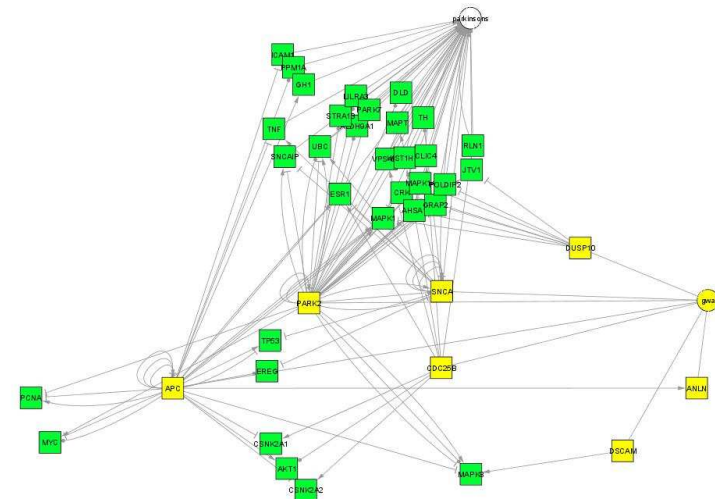| By mention of topic	
▶ Diabetes	25-hydroxy vitamin D 3 compositions
▼ Diabetes Type 1	Animal genotyping method
	DNA markers for meat tenderness

Linguamatics

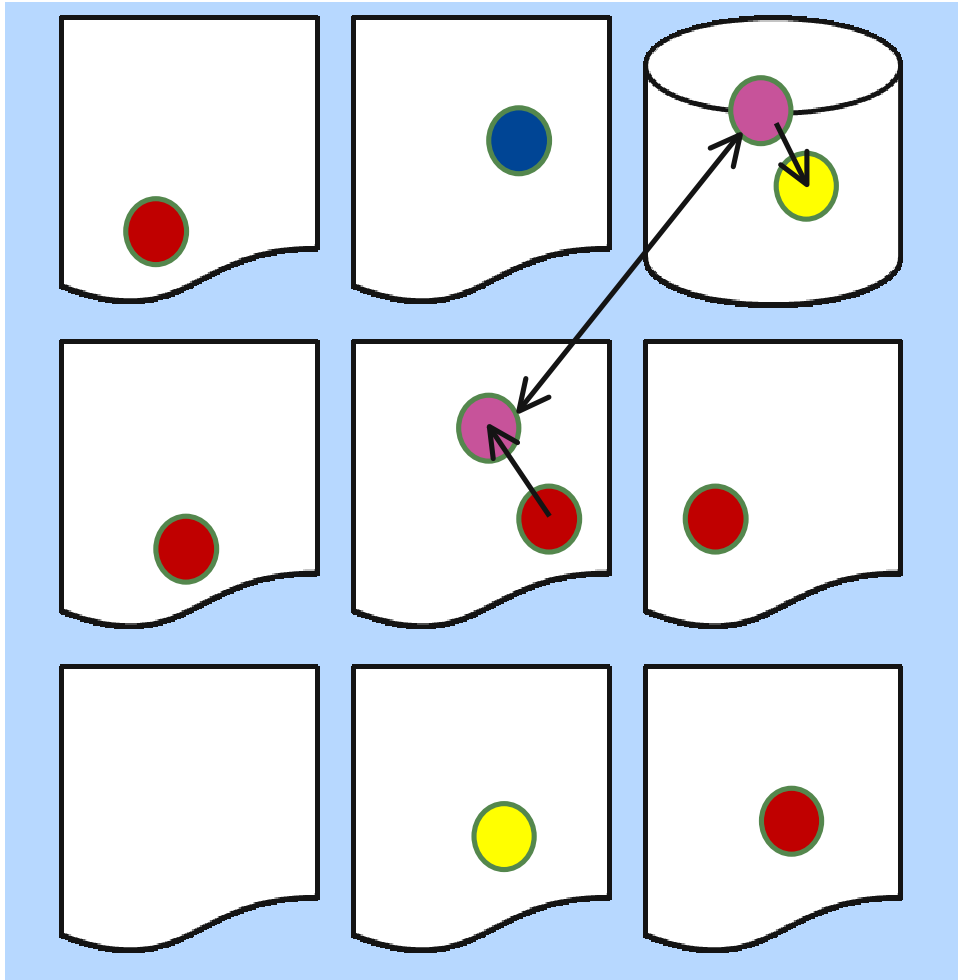# Summarize Information from Multiple Documents
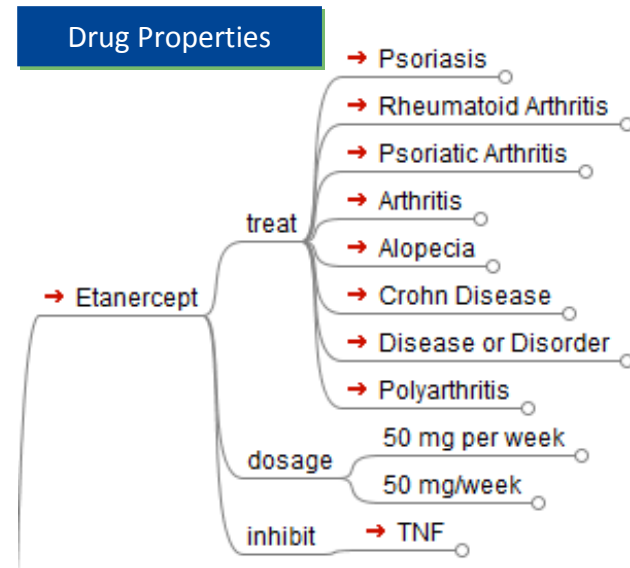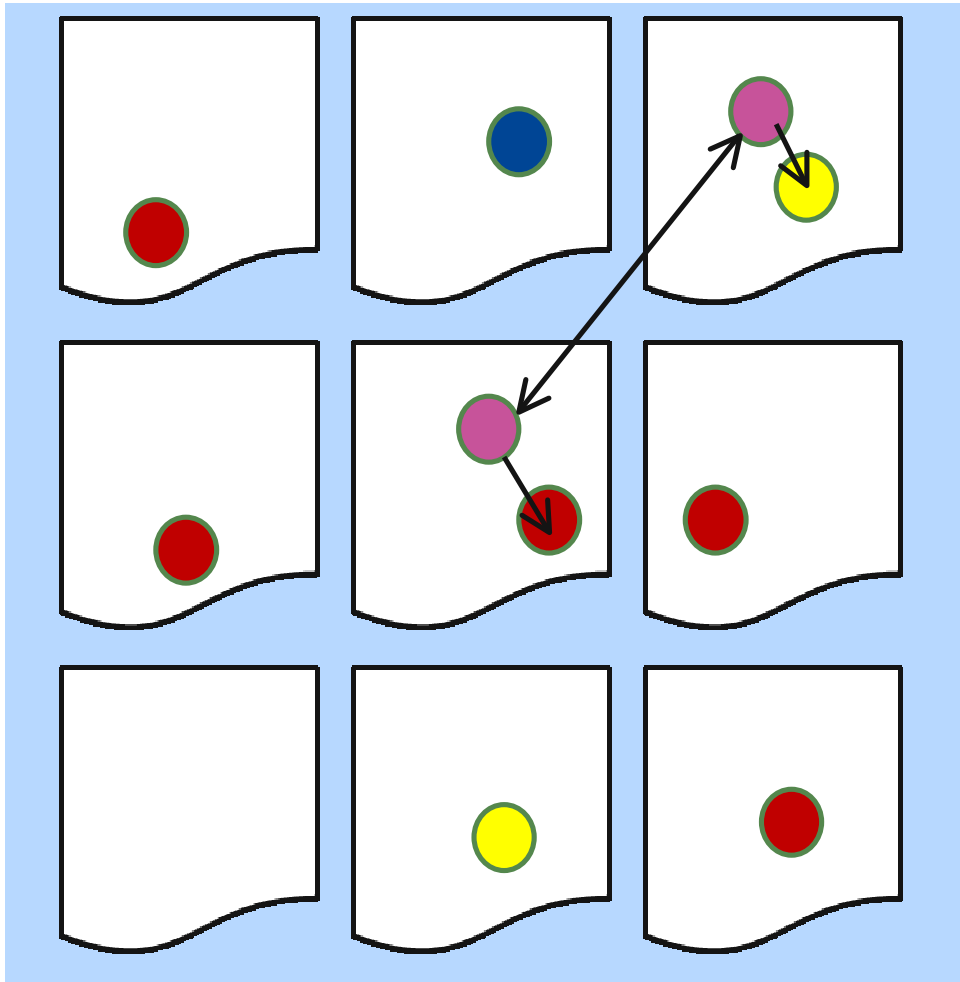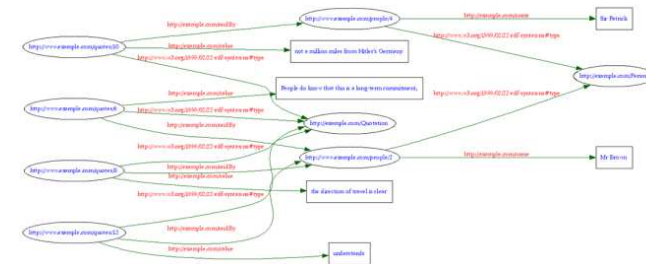
Discover New Knowledge from the Documents

Indirect Relationships

# Discover New Knowledge from Multiple Sources

Indirect Relationships

Demonstration

- 3718 USPTO patents, 2009

- Life Science area

Linguamatics

Raw Source from USPTO

- Patents are downloaded in three formats:

  - 1976-2001: tagged text

  - 2001-2004: XML (INID-like)

    - 2001-2002: DTD version 2.4

    - 2003-2004: DTD version 2.5

  - 2005 onwards: XML (verbose)

    - 2005: DTD version 4.0

    - 2006: DTD version 4.1

    - 2007 onwards: DTD version 4.2

- All formats are mapped to XML with a single consistent set of fields

Configuring Helps the End User
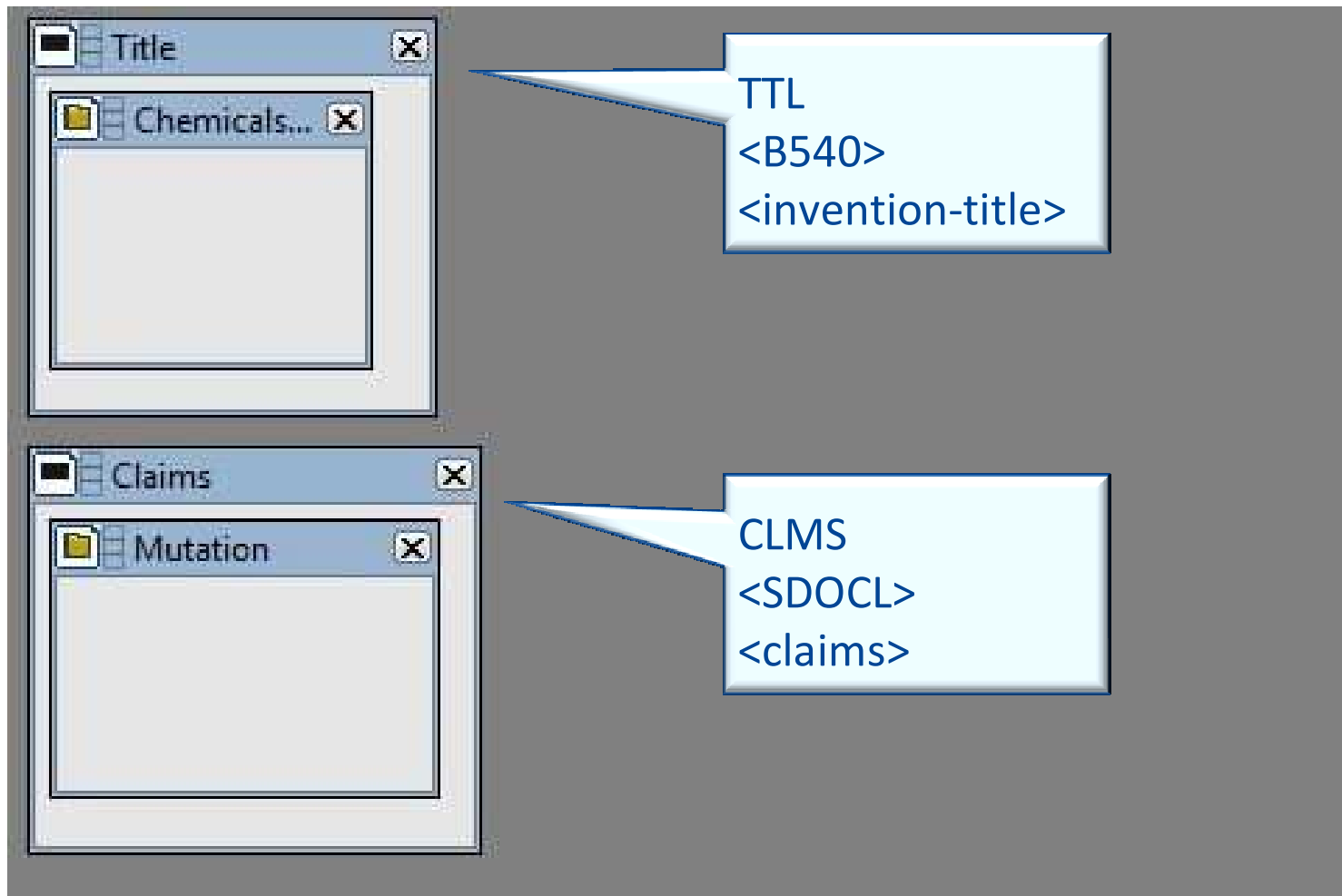
Conclusions

- **Different uses of patents**

  – Patents as a knowledge source for R&D and Commercial Intelligence

  – Use of NLP and terminologies to help find patents, and review them

- **I2E provides value in**

  – Getting to the right information faster

  – Cutting down time to review

  – Connecting between patents and other document source or knowledge bases

Linguamatics