

The ePatent project

Multilingual access To European Patent databases

Bernard Normier
Lingway
June 2002



ePatent framework

- EC eContentprogram
 - Budget 2.4M € started Jan02. 2 years.
- Partners
 - INPI, France
 - UKPO, UK
 - OEPM, Spain
 - OEPA, Austria
 - Lingway, France (natural language technology)
 - Jouve, France (database and internet infrastructure)



Using NLP to facilitate use of patent database

- Patent databases are difficult to use by non-professional users
 - Searching IPC is difficult
 - Searching boolean queries is difficult
 - Multilinguality is a key issue in Europe
- ePatent will offer
 - Easy access for non professional users (SMEs)
 - Search, filtering, reading aids and translation tools



Key technology: natural language processing

■ What is NLP

- Tools to « understand » natural languages
- Based on dictionaries, semantic networks, grammars
- Now amateur technology

■ Applications

- Natural language interfaces, mono or multi-lingual
- Text analysis : indexing, extracting, tagging
- Translation, multilingual reading aids



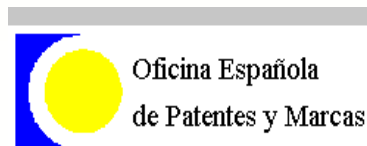
Re-using the French experience of INPI

- Using NL interface on IPC for 5 years
 - Translates NL query to IPC codes
 - Based on a 50.000 words dictionary, 35.000 concepts
 - 30.000 NL queries (in French only) per month
- Extension to 3 new languages
 - English, Spanish and German
- Extension to new features
 - Filtering, Ranking, Translation aids



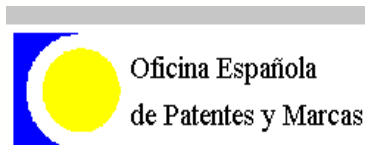
Systematic description of IP language

- An NLP application works correctly only if linguistic knowledge is adapted
- ePatent develops a linguistic database describing the patent language:
 - Terminology > used words
 - Phraseology > sentence forms, style...
 - Multilinguality > French, English, Spanish, German
- Uses this knowledge for different features
 - Search, filtering, tagging, translation




NLaccessusingIPCtopatents

- Thelanguageofpatentsisveryspecific
 - DistancebetweentheprofessionallanguageofIPandthegenera 1
technicalandscientificlanguage
- ProposeIPCcodesthroughamatchingwithIPC
textualdescriptions(Titles,cross -references,key
phrases,Notes,Outlines)
- PatentsareobtainedthroughtheIPCcodes



Natural language interface to IPCFR/EN

Query: 

Result: 10 / 376 codes Analyse

(1) [H-05-H-1/00](#) G

For
Lat

IPC Context:

- [-] [H] SECTION
- [-] [05] ELEC
- [-] [H] PLAS
- NEUTRONS ; PRODU
- [1/00]

CIBLN: Interface de Test et Diagnostic - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Refresh

Recherche: Standard Section:

Question:



Résultat : 30 / 1108 entrées

(1) [H-05-H-1/00](#) Production du plasma; Mise en oeuvre du plasma;
Scores du Formulaire (U/D/I) = 693 / 0 / 477 (entrée H-05-H-1/00-Titres)

Contexte CIB:

- [-] [H] ÉLECTRICITÉ
- [-] [05] TECHNIQUES ÉLECTRIQUES NON PRÉVUES AILLEURS;
- [-] [H] TECHNIQUE DU PLASMA ; PRODUCTION DE PARTICULES ÉLECTRIQUEMENT CHA
- OU DE NEUTRONS ; PRODUCTION OU ACCÉLÉRATION DE FAISCEAUX MOLÉCULAIRES OU ATOMI
- [1/00] **Production du plasma; Mise en oeuvre du plasma;**

Opening page http://hugo.lingway.com/lrCgi11/lrQACgi.exe?_action=lsGetPathFolder&_host=localhost&_port=122



LA PROPRIÉTÉ INDUSTRIELLE

Classical full text access to patents

- Keywords access
- Search for patents, through a boolean query and the full text engine ranking mechanism, by matching Titles and Abstracts
- No benefit of the IPC expertise
- Classical problems of full text engines: noise and silence



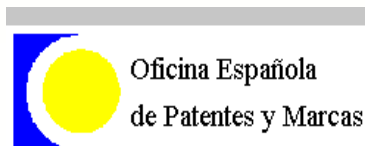
Enhancing full text with NLP

- Analyse a natural language query
- Use the semantic network to expand to related words
- Generate the appropriate boolean query to the text engine
- Reduce by a significant factor both noise and silence
- Allows a better highlighting of relevant words in text



Filtering approach

- Analyse the document structure and “rhetorical” markers in the text
 - Done by an automatic XML tagging with “semantic tags”
- In batch on large volumes (XML retro -conversion)
- In real time on small sets (result of a previous query)
- Rank relevant documents according to the query, and highlight relevant text fragments



Semantic XML tagging

BACKGROUND OF THE INVENTION

The invention relates to farm implements and in particular relates to an attachment for a bidirectional mold board plow which is adapted to engage the ground in a trailing relationship with respect to the mold boards.

Bidirectional mold board plows are known. In particular, Gomez **U.S. Pat. No. 4,800,963** describes a mold board plow which has a tool bar that supports ground engaging mold boards. The tool bar is pivotably secured in a frame and is movable into left and right hand positions behind the tractor. A hydraulic reversing mechanism is connected to the frame to set the tool bar at an angle to the left or to the right with respect to the direction of motion of the plow. **It is desirable to reverse the plow so that when the tractor direction is reversed, the ground is turned over in the same direction with each pass**

Mold board plows pick up large chunks of earth and produce clods which interfere with subsequent planting operations. One known device which conditions the soil after plowing is described in Sigmund **U.S. Pat. No. 3,574,320**. A mulching tool is pivotably mounted to a support which is carried by the plow frame. The mulching tool trails the plow at all times. **The plow is not reversible.**

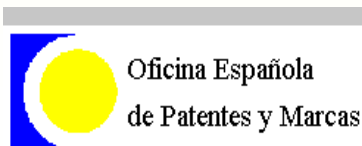


Finding patent object

The invention relates to an arrangement in a single- or multi-share plow, and more especially to a reversible plow bodies, in order to allow limited elastic compliant rearward rotation of the plow body or each plow body very heavy plowing forces. [...]

An object of this invention is to provide a plow which is lightweight and does not require large material dimensions well suited for use when the plowing conditions are as described in the foregoing paragraphs. [...]

The present invention generally relates to farm implements and, more particularly, to an on-land plow as well as adjustable plow units mounted on a common frame and which are pulled across a field by a tractor for simultaneous



Finding previous patents drawbacks

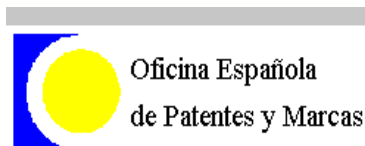
The **plow** attachment suffers from the *drawback of* having a *bulky* adjustable mounting mechanism that contains : lost. [...]

Known **plow** debris deflectors have proven to be *inadequate* for various reasons. [...]

Another *drawback of* known **plow** debris deflectors is the limited number of adjustments which can be made to th

The performance of a **plow** debris deflector is *affected by* various factors. [...]

The varying conditions of the field, the cover crop grown, the presence of wet or dry clays, sandy soils, and the d **plowed** may seriously *degrade* the performance of the **plow** debris deflector of limited adjustability. [...]



Reading aids and translation

- Help the user with foreign language
- Online dictionaries
- Term identification and translation
- “draft” translation
 - (not yet decided in ePatent)
 - Enrich existing MT systems with Patent terminology



Machine translation **before** terminology adaptation

La présente invention concerne des ceintures de sécurité pour des véhicules à particulier concernée par un système de ceinture de sécurité ce qui incorpore d'alarme adapté pour fonctionner s'il y a une tentative de conduire le véhicule sûreté d'un siège occupé **desserré**.

Une condition légale indique que le dispositif d'alarme fonctionnera quand le de la **transmission** de véhicules est dans le "**lecteur**" ou la "**dans-vitesse**" mais fonctionner quand le sélecteur de vitesse est en "**parc**" ou "point mort." C'est condition que le circuit d'allumage ne devrait pas être coupé à tout moment e manque d'attacher une ceinture de sécurité. Les termes "**se garent**" et le "lect liaison avec une **transmission** automatique et les termes "point mort" et "**dans utilisés en liaison avec une boîte de vitesse manuellement fonctionnelle**. Cep **simplicité** que les termes "**se garent**" et le "lecteur" sont ci-après employés pc les termes "point mort" et "**dans-vitesse**" respectivement.

Selon la présente invention dans un système pour **empêcher piloter** d'un véhic



Machine translation **after** terminology adaptation

[La présente invention concerne des ceintures de sécurité pour des véhicules à moteur particulier concernée par un système de ceinture de sécurité ce qui incorpore un dispositif d'alarme adapté pour fonctionner s'il y a une tentative de conduire le véhicule avec insécurité d'un siège occupé non **attaché**.

Une condition légale indique que le dispositif d'alarme fonctionnera quand le sélecteur de la **boîte de vitesses** de véhicules est dans le "**lecteur**" ou "**en prise**" mais qu'il ne fonctionnera pas quand le sélecteur de vitesse est dans le "**parking**" ou le "point mort." C'est également une condition que le circuit d'allumage ne devrait pas être coupé à tout moment à cause d'un manque d'attacher une ceinture de sécurité. Les termes "**parking**" et "**lecteur**" sont utilisés en liaison avec une boîte de vitesses automatique et les termes "point mort" sont utilisés en liaison avec une boîte de vitesses **actionnée manuellement**. Cependant, **simplicité** les termes "**parking**" et "**lecteur**" sont ci-après employés pour inclure également les termes "point mort" et "**en prise**" respectivement.

Selon la présente invention dans un système pour **empêcher le pilotage** d'un véhicule



Conclusion

- NLP is amateur technology
 - Performance and quality are acceptable
 - Dictionaries can be shared between several tasks
 - Semantic networks can be reused between languages
- NLP is a key technology for patent databases
 - To give an easy access to non-specialist
 - To develop high value added services
 - To target new market niches, in particular in SMEs

