

## The 2011 ICIC Patent Panel

Chemistry has been, and still is, the favorite area for new developments in the electronic information industry. This is because, by its nature, chemistry is more structured than other areas of expertise and chemistry research has always demanded high-quality information. To detect trends in the (chemical) patent information industry was, therefore, the goal of this year's patent panel at the ICIC conference. Even more so, since the content of scientific information in patents has drastically increased over the last twenty years.

Chaired and animated by [Pierre Buffet](#) of Questel, the organizers had invited five expert panelists from various areas: [Rahman Hyatt](#) from Minesoft (a service provider), Alfred Elmaleh from the French Petroleum Institute (a patent professional), [Monika Hanelt](#) from Agfa Graphics (a patent information expert), [James Ryley](#) from SumoBrain (a service and software provider) and [Alexandro Campana](#) from WIPO.

The organizers of the patent panel had solicited questions from the user community prior to the conference. These questions were then grouped into five sections: (1) information professionals, (2) tools and software applications, (3) the rise of Asia, (4) economics of the patent information industry, and (5) the role of patent offices. All questions were made available to the panelists before the conference. During the patent panel, the audience had the chance to ask even more questions.

### ***Information professionals***

A new generation of scientific researchers has arrived on the scene who grew up with the internet and are used to satisfying their information needs by searching themselves (Google generation). This is not an entirely new situation for the patent information community because end-user tools such as SciFinder or Reaxys have been available for a long time. This has definitely had an impact on the role of the information professional in research institutions.

Simple questions are now being answered by end-user tools and no longer require the information professional. Therefore, searches have become more complicated. Clients address the information professional whenever answer sets are too large or they cannot handle the type of search. Typically, searches on chemical structures (simple molecules) can be done by the end-user. Searches for reactions or processes, on the other hand, ask more frequently for the specialist.

The reason for contacting the search specialist also depends on the nature of the client. Scientists who have worked for a long time in a special area require less assistance from the search specialist than novices in a given field. In the latter case, data sets are often rather incomplete. In general, end-user search provides for more informed clients, which is seen as an advantage to the professional searcher.

Most panelists agreed that a professional search is required whenever patents are to be filed – in several institutions they are even mandatory. Some members of the panel observed that scientists do not have sufficient time for doing simple searches themselves, let alone complicated ones. A close cooperation between patent departments and information professionals is regarded as being beneficial to the efficiency of patent filing.

In many institutions, information professionals proactively participate in the process of identifying new areas of research by preparing alerting services. The delay of professional searches – up to 2–4 weeks – is considered a problem by the panelists and, of course, by the end-users.

## ***Tools and software applications***

The discussion on new tools concentrated on methods such as semantic search, linguistic methods, and automatic index extraction – a clear distinction between those terms obviously not being possible. Although considerable progress has been made on semantic search tools, the patent panel was not convinced that these methods have been established as reliable routine tools. Many agree that this is due to the type of documents and especially the language in chemistry patents.

Semantic tools are regarded by some as a black box, and the trust in them supplying reliable results is not large. Although it can be made transparent why a given answer set is found after the search, it is difficult to predict the result of a search. The devil seems to reside in the detail. However, many agree that a combination of semantic indexing and human curation will eventually produce results more efficiently.

Semantic methods are regarded as helpful in processing search results for the customer, but they do not provide more relevant answers. They can also assist in locating essential information in a patent (the standard structure of patents being another help in that case).

In certain areas, the sheer volume of the original material makes its processing by deep human indexing impossible. In such cases there is no way around semantic methods, for example the new flood of Asian patents (see next paragraph) or mass data from the internet, such as Twitter or even e-mails.

## ***The rise of Asia***

The number of Asian patent documents, especially those from China, is anticipated to increase enormously (for 2015, the SIPO projects 750,000 patent and even more utility model applications.) This will be a challenge for all producers of patent information. Although the panelists agreed that the state of the art of machine translation is anything but perfect, all participants saw no way around this technology in the future.

Currently, 40% of the knowledge generated worldwide comes from Asia. As of July 2012, Chinese and Korean patents will be included in the [minimum PCT documentation](#) set. Therefore, searches in Chinese (and other Asian) documents will have to be made and information professionals will have to rely on machine translation.

Many efforts are currently on the way to increase the quality of automated translation. The original trilateral cooperation between the U.S., European, and Japanese patent offices has been extended to [five parties](#) by including the Chinese and Korean offices. One of its ten [projects](#) aims at increasing the quality of computer translations – and great advances have been made. In addition, the EPO cooperates closely with Google on that topic. Other prominent projects in this area are the EU-sponsored [Project PLuTO](#) (Patent Language Translations Online) and the [CLIR activities](#) (cross-language information retrieval).

Although the discussion focused on machine translation, other approaches for coping with the problem were suggested. If patent documents were provided in XML (instead of PDF), the cost of database building could be greatly decreased. The WIPO has, therefore, formed a [task force on XML](#). However, the WIPO can only issue recommendations to the national patent offices. A completely different approach to the language problem was suggested: Why can't well-trained Asian experts do searches directly in Asian documents? Obviously, several patent information vendors have Chinese interfaces for searches. In addition, CAS employs Chinese experts to extract titles, keywords, and abstracts (whereas the indexing of substances is done in the U.S.).

## ***Economics of the patent information industry***

Both the patent panel and the audience agreed that there is and will be a strong demand for value-added services, such as Thomson-Reuters or Chemical Abstracts. Expensive joint deep

indexing services, such as PDG and IDC, were not terminated in the past because of a lack of demand but because of the availability of services such as Derwent (Thomson) or CAS.

The question of offshoring services such as patent indexing or professional searching provided several interesting answers. The reason for offshoring such services to Asian countries cannot be merely financial: the amount of management involved, e.g., for quality control, is far too large to justify the cost saved. Offshoring the technology of database producing is not so problematic, but offshoring the contents is. In addition, intellectual property is regarded as much too valuable to outsource. However, access to Asian languages by employing well-educated native speakers is a major reason for offshoring.

## ***Role of patent offices***

An enormous amount of searches are performed in the patent offices using highly sophisticated tools, such as EPOQUE. Several applicants have asked for access to EPOQUE in order to have the same tool as the examiners at the patent office. After complete re-engineering of EPOQUE, this may legally be feasible in 2012. Although EPOQUE is considered to have some charm, searchers in industry may not be able to subscribe to all databases the EPO does. One participant in the discussion suggested that the patent offices should concentrate on improving their workflow instead of making their services commercially available.

To support the patent offices in providing high-quality information is considered to be the major role of the WIPO, preferably in cooperation with the information providers. This is also apparent in a public-private partnership initiated by WIPO: the [ASPI program](#) (Access to Specialized Patent Information). This program will provide advanced tools and services for retrieving and analyzing patent data to key centers in least developed countries. Six major database vendors have joined this partnership.

## ***Digestif***

The patent panel concluded by asking the rather provocative question: to which organization, person or product would the gold medal go for the biggest improvement in patent searching over the past twenty years? Being good sports, the panelists refrained from having the same vote twice. The five answers were: (1) to the EPO for its policy of disseminating patent information, (2) to the development of methods of artificial intelligence and, thus, linguistic retrieval, (3) to the availability of chemical structure searching, (4) to Google for generously sharing data processing software for further development (not for Google patents though!), and (5) to the international cooperation between governmental institutions (e.g. EPO, WIPO) and the private information industry.

*Wolfgang Gerhartz*

# Patent panel - questions

## 1 - Information professionals

- The new generation is growing up with searching, they are from the "Google generation" and they search themselves. I can imagine that this will have influence on the search profession. What will be the future of the Patent Information Specialist/Professional? Will they be then more a consultant to give advice and to handle more complicated issues? **Guido Vang (DSM)**
- Has the number of information professionals changed significantly and do you anticipate more or less in the future? Do you see trends within the user community? **PDG**
- In the past, we, the information professionals, have seen, that more and more products have been developed for end users, often with more functionalities. Do you think, that professional information retrieval will be replaced within some years by highly sophisticated end-user tools? How do you balance these two user groups and direct your resources? **PDG**

## 2 – Tools and Software Applications

**2.1.** How do you estimate the development in semantic technologies with respect to patent information? **Peter Kallas (BASF)**

*&together*

**2.1bis.** There is no breakthrough in automated data mining and mapping of patent data. Is the hype over and what does the future look like? Are there any new developments with regard to automated mining on the horizon? **PDG**

*& [si ontologies non couvertes]*

**2.1ter.** When will semantic indexing (using text mining and ontologies) be a routine procedure to provide high value databases? How will database producers differ in quality when using more or less the same technologies? In this scenario: should every database producer concentrate on every topic, or can single producers concentrate on industries (e.g. agrochemistry) with sophisticated automated indexing? **Heinz-Gerd Kneip**

*& optional if time allows*

**2.2.** Over the past ten years, ICIC has featured and discussed many, many intelligent new software applications for the patent world -- including search, mining, analysis and visualisation. Do the panellists feel that precise patent searching is now easier and more efficient than it was back in 2001? **Harry Collier**

**2.3.** Twenty years ago at this conference Edlyn Simmons and Nancy Lambert gave a memorable paper entitled “ Patent statistics: comparing grapes and watermelons” . Are patent statistics still flawed as a source of competitive intelligence? How have things changed (if at all) over 20 years? **Wendy Warr [ref to OECD works]**

*Transition*

**2.4.** How do you see the future of indexing /coding and preparing added-value abstracts with all the challenges connected with the growing numbers of patents /journals (e.g. from China) ? **PDG**

### **3. The rise of Asia**

**3.1.** The rise in documentation in an Asian language is very important. China had 400 000 patent (invention) applications in 2010 and even more Utility Model applications. SIPO projects 750 000 patent applications in 2015: Are we properly planning for this huge increase in Asian documentation? **Niclas Morey - EPO**

**3.1bis.** How will patent database providers cope with the huge and increasing number of patents e.g. from China without dramatically increasing their pricing? Answers should mention planned measures in the workflow to reduce operational costs. **Heinz-Gerd Kneip – Basf [cf Economics?]**

**3.2.** What is your opinion about the necessity, the development and futered inclusion of literature for CN and other “ emerging” countries? **Pater Kallas - BASF**

### **4. Economics of the Patent Information Industry**

**4.1.** Will added-value services like TR (Thomson-Reuters) and CAS survive? **Peter Kallas - Basf**

**4.2.** More and more large companies are experimenting with off-shoring their search groups or parts of the search-groups (to Asian or East-European countries) and dismantling their patent information departments - how do our providers see the future as far as their customer base is concerned? Is this off-shoring a trend all over the world or only a “ European phenomenon” ? **PDG**

**4.3.** Although the information on the internet is not covered by any of the classical databases, it is considered in patent law suits as a source of prior art if the information satisfies the same criteria as oral disclosures and instances of prior use. This Internet information can even be found in search reports and/or cited during examination procedure. Are there any plans for database producers and hosts to cover this information by e.g. partnering with Google? **PDG**

## 5. Role of Patent Offices

5.1. In order to increase the quality of patents filed by applicants, major patent offices are seriously considering to make available, without limitation and to everybody their search tools. **Jacques Michel**

&

5.1bis. Patent Offices are more and more active in delivering search solutions. There was an enormous fear in the end 90s that this free offering would kill commercial providers. In the meantime, some of them had double digit growth rates over several years.

Retrospectively, do you have an explanation why those fears were not justified?

**Anonymous**

5.2. What will be and what should be the role of organizations like WIPO and EPO in the future information provision? **Peter Kallas - Basf**

## 6. Digestif

If the members of the panel had to award a gold medal for the organisation, person or product that has made the biggest improvement in patent searching over the past twenty years: who or what would they choose? **Harry Collier**