# PerFedPat IEF Marie Curie Project (Project Number 275522)
# Executive Summary

In this executive summary we describe the work and the achievements at the end of the PerFedPat project. PerFedPat is an IEF Marie Curie project aimed at to research into a new generation of advanced patent search systems for the patent related industries and the whole spectrum of patent users by designing a new exciting framework for integrating multiple patent data sources, patent search tools and UIs. The work involved -in the first year of the project- the elaboration and development of core PerFedPat components and the initial evaluation of the PerFedPat prototype. The work which has been done in the second year of the project integrated more components and search tools. Also, a final evaluation of the PerFedPat system has been conducted in a national patent office. In this executive summary, we also present the training activities of the fellow researcher as well as knowledge transfer and integration activities. Important scientific results that were achieved are highlighted. Finally, impact and exploitation opportunities of the PerFedPat system are analyzed and discussed.

There is an abundance of systems today to search for patents. Some of them are free and have become available from patents offices and Intellectual Property (IP) organizations the last ten years (e.g. Espacenet and Patentscope), as the growth of the internet and the development of search technologies facilitated the provision of powerful web-based search systems of patent databases. Other systems are free –but developed by search technology providers (Google Patents)-, or are based on subscription and are provided from other independent producers (e.g. Delphion). All web-based patent search systems allow searches using the simple "search box" paradigm. Other free or commercial systems may have better capabilities, for example for structural searching in particular fields, term proximity operations or to leverage domain semantics, but essentially they all operate on the same centralized index paradigm. According to this paradigm, patent documents need to be periodically crawled or otherwise collected, afterwards they are analyzed and eventually become part of the centralized index.

**PerFedPat** is an interactive patent search system that follows a different approach based on Federated Search. Federated Search represents a Distributed Information Retrieval (DIR) scenario and allows the simultaneous search of multiple searchable, remote and physically distributed resources. PerFedPat provides core services and operations for being able to search, using a federated method, multiple online patent resources (currently Espacenet, Google patents, Patentscope and the MAREC collection), thus providing unified single-point access to multiple patent sources while hiding complexity from the end user who uses a common query tool for querying all patent datasets at the same time. Wrappers are used which convert the PerFedPat internal query model into the queries that each remote system can process. "Translated" queries are routed to remote search systems and their returned results are internally re-ranked and merged as a single list presented to the patent searcher. PerFedPat is developed upon ezDL therefore, in addition to the patent resources which are provided in PerFedPat, there are other resources already provided by ezDL, most of them

offering access to online bibliographic search services (e.g. ACM DL, DBLP, Springer, PubMed) for non-patent literature.

The second idea that we explored in the PerFedPat project is based on the general model of integrating multiple tools and methods for professional search. More precisely we explored the efficiency of an open framework for patent search using a variety of patent search tools and User Interfaces (UIs). To achieve this goal PerFedPat uses a pluggable and extensible architecture, thereby providing multiple patent search tools and UIs. Consequently in PerFedPat federated search is used beyond the way that it is used in traditional Distributed IR, i.e. to provide a single merged list of multiple ranked results. Hence, the second innovative feature of PerFedPat is that it enables the use of multiple search tools which are integrated in PerFedPat. Based on this architecture PerFedPat is a pluggable system which puts together the following components: retrieval, selection, integration, presentation and adaptation (Figure 1). Currently the search tools which are integrated are: a) an International Patent Classification (IPC) selection tool, b) a tool for faceted navigation of the results retrieved based on existing metadata in patents, c) a tool producing clustered views of patent search results d) a Machine Translation (MT) tool for translating queries for cross lingual information retrieval e) a tool for patent image retrieval.
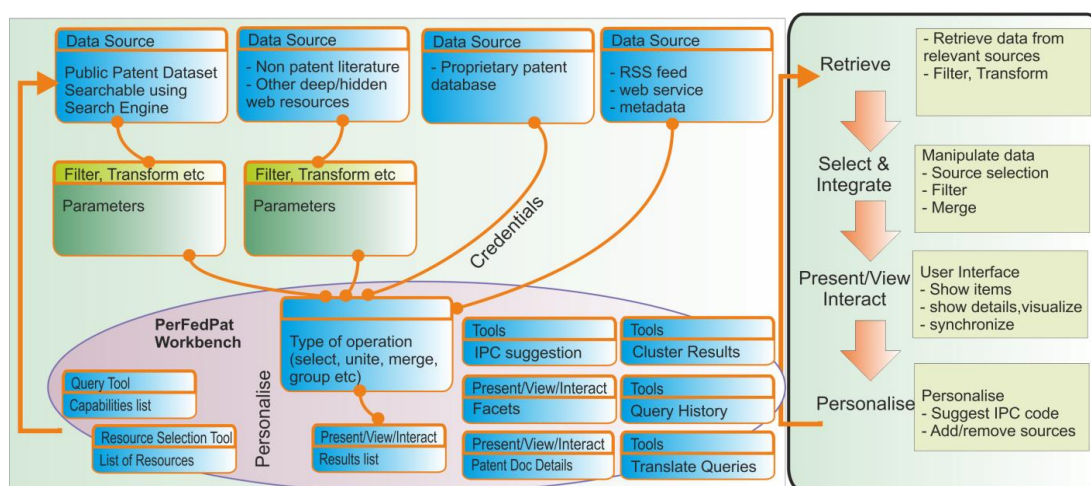


Figure 1: PerFedPat architecture and component overview

Although PerFedPat relies on existing patent search systems to execute the core retrieval task, from an architectural point of view PerFedPat is innovative using the Federated Search approach and goes beyond the state of the art in patent search systems in terms of scale, heterogeneity as well as extensibility as it is based on a service-oriented, message-centric architecture able to integrate data sources into new, more useful ways. From that perspective the PerFedPat system is the first open architecture data aggregator for patent information, and its contribution is to show that the sum of the utilities provided by each search tool could be really bigger than the single utilities and enabling possibilities lie in an integrated approach for patent data delivery and intelligent processing and presentation.

- *Scale:* The PerFedPat patent search system is in principle more scalable than other systems since it is based on a highly distributed method for accessing patent information sources, each using its own storage, processing, and searching capabilities.
- *Heterogeneity:* Different data sources, search tools and UIs can be combined in a

non-predetermined way, non over-engineered method, as far as they abide by the PerFedPat framework.

- *Extensibility:* The PerFedPat framework is not developed be a single turnkey one-size-fits-all solution, but instead is designed as a pluggable architecture in which it is easy to develop and deploy new components. The ezDL framework on which PerFedPat is based is easy to extend and is based on a service-oriented architecture.

In the first iteration of the PerFedPat project we conducted one-to-one interviews with patent search experts. Also after a stable prototype has been developed we conducted expert review studies with patent examiners and task-based evaluations using beta versions of the PerFedPat system (Figure 2). We were primarily seeking to check the effectiveness, efficiency and usability of the system. Also to seek and document expert opinion and advice about the match between the system and the real world, confidence in the system, user control in using tools, etc. This first evaluation work has indicated a promising degree of acceptance of the usefulness of a federated search system like PerFedPat.
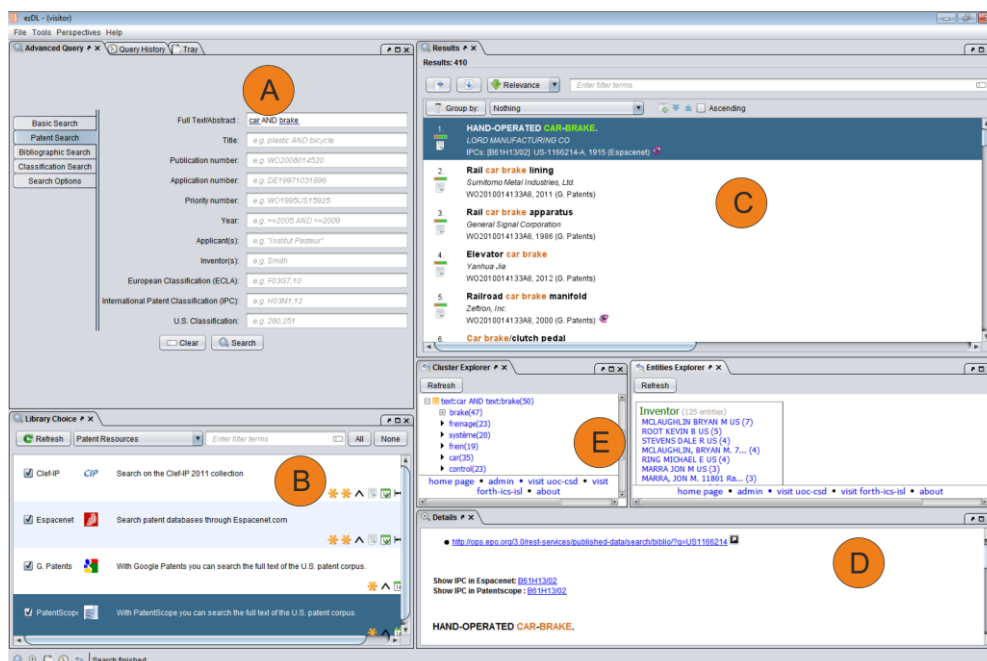


Figure 2: PerFedPat overview with some core and patent search tools open

Another large user study was done in the final third iteration of the PerFedPat project. The aim of this study was two-fold. First, we examined if patent examiners could learn PerFedPat easily and also if they would like and could be positively affected and well engaged using a federated patent search system comprising multiple resources and search tools. Also, we examined if the patent examiners could efficiently use the federated search system and if the integration of multiple tools can assist them to attain the effectiveness required in prior-art patent search tasks. Overall the aim of the user study was to explore the overall opinion of patent professionals as well as the usefulness and the effect of a new interactive patent search system, PerFedPat.

Our work on PerFedPat was inspired by the idea of providing an integrated patent search system which will be able to provide a rich, personalized information seeking experience for different types of patent searches, potentially exploiting techniques from different IR/NLP/MT technologies. We believe we demonstrated the feasibility of producing such an integrated system. First, we have demonstrated the applicability of federated search in patent search systems. PerFedPat provides core services and operations for being able to search multiple online patent resources (currently Espacenet, Google patents, Patentscope

and the MAREC collection), thus providing a unified single-point access to multiple patent sources while hiding complexity from the end user. More patent resources can easily be made part of the PerFedPat federation to increase coverage or for reasons of specialized searches that some patent systems may provide.

The PerFedPat produced new and innovative knowledge of substantial quantity and quality in the area of professional search in general and that of patent domain in particular. Patents have direct economic value: many organizations make significant revenue by licensing others to manufacture or use their patented invention; patents may be traded – i.e. the rights sold to new organizations; they are used to estimate some company's value; and so on. Patent protection is an extremely important way in which individuals and organizations protect and exploit their intellectual property. During the project different disciplines were investigated and search tools and search systems for patent professionals and for patent search in general were produced. The success of this project could potentially have a very positive impact on patent search industry. The PerFedPat project illustrated an innovative way for building patent search systems and applications, which goes well beyond existing patent search systems and search methodologies.