IP Australia Patent Backcapture Challenge

IP Australia is the Australian Government agency responsible for administering intellectual property (IP) rights and legislation relating to patents, trademarks, designs and plant breeders’ rights.

Backcapture Project Outcomes

- **390,000** historical records transformed
- **Fast delivery.** Completed in 6 weeks.
- **Reliable.** 10 x reduction in error rates
- **Accurate.** Field accuracy of 99.7%

The Patent Backcapture Challenge

IP Australia had 390,000 historic patent documents, dating back to 1904, with little or no metadata. It was impossible to search through them effectively. IPA asked Semantic Sciences to extract items of metadata using the Sintelix extraction capabilities so that these records would be accessible to clients.

Many of these documents were only available in hard copy and some of them over 100 years old, in black and white and of moderate quality. Using OCR, these documents were converted into a PDF format, creating new opportunities for storage and analysis.

Project Requirements

Capture/extract bibliographic fields from OCRed patent records and specifications from 1904 to 1979.

Provide IPA with captured/extracted data in a specified structured XML format.

The Backcapture Solution

As shown in the workflow diagram, Sintelix provided a solution to IP Australia’s challenges within 2 months by:

- Extracting and transforming existing patent specification documents into 390,000 PDF documents
- Loading those documents into Sintelix
- Normalizing and extracting information from those documents, creating 390,000 xml files
- Placing the metadata back into IP Australia databases in a searchable and easy to analyze format, making records accessible to clients.
With Sintelix, IP Australia were able to transform a significant amount of data, extracting a large amount of information, including:

- Filing date (lodging or lodged date) of patent specification
- Invention title
- Applicant(s) name
- Inventor(s) name
- Agent’s name
- OPI date
- Filing date of basic application/priority application
- IP Office of priority country
- Priority application number/number assigned to priority application
- Divisional application numbers (parent/child applications).

These examples show the metadata extracted from historic patent specifications:

Example 1.

Example 2.

“Sintelix provided a solution to IP Australia’s challenges within 2 months”

Veena Baht, IP Australia
Backcapture Project Outcomes

With Sintelix, IP Australia were able to successfully extract metadata from 390,000 patent specifications within 6 weeks, meeting the tight deadline and delivering the required level of accuracy.

The letter of recommendation below from IP Australia confirms the following project highlights:

- High consistency
- Excellent accuracy
- Rapid execution
- Low cost

Here are some of the comments from the letter of recommendation:

“The project was organised in two (2) stages: a proof of concept and a main delivery, with a decision gate in between. The results IPA received from the proof of concept were good and achieved within a very short period, so IPA authorised the main project to proceed. Its timelines were tight (6 weeks) and required high accuracy.

Semantic Sciences Research provided IPA with visibility of its progress via online access to progress reports with drill-down to the source and processed data provided from its Sintelix software platform.

Delivered results were excellent. A field accuracy of 99.7% was achieved, which is significantly greater than IPA would expect from human transcription. The project was performed on time and on budget.

IP Australia enjoyed a positive experience of working with Semantic Sciences Research and using Sintelix. The company met our procurement and performance expectations for service providers. We valued Semantic Sciences Research’s timeliness, responsiveness and proactivity.”

— Veena Bhat, Patent Search Capability Coordinator, IP Australia.

“With Sintelix, IP Australia were able to successfully extract metadata from 390,000 patent specifications within 6 weeks”

Veena Bhat, IP Australia