USPTO’s Challenge to Improve Patent Search
With Artificial Intelligence

Request for Information (RFI)
Table of Contents

1 Introduction ................................................................. 2
2 Purpose ........................................................................... 2
3 Objectives ........................................................................ 3
4 Information Request ......................................................... 5
5 Vendor Responses .......................................................... 6
6 Questions ......................................................................... 6

This notice is issued solely for Market Research and information and planning purposes. It is not a solicitation and does not constitute a request for quotation or proposal. Telephone calls requesting a solicitation will not be accepted or acknowledged. Participation in this effort is strictly voluntary. All costs associated with responding to this RFI will be solely at the interested respondent’s expense.

Vendors should mark all confidential and/or proprietary information in the response as applicable. Proprietary information submitted in response to this RFI will be protected from unauthorized disclosure as required by the Federal Acquisition Regulation (FAR). Information provided by vendors will only be shared with USPTO employees and contractors with signed confidentiality agreements involved with market research.

The USPTO reserves the right to review responses to this RFI and issue a Request for Proposal (RFP) / Request for Quote (RFQ) to vendors using the Alternative Competition Method in accordance with the Patent and Trademark Office Acquisition Guidance (PTAG), and the Patent and Trademark Office Efficiency Act 35 U.S.C 2(b)(4)(A).
1 Introduction

The United States Patent and Trademark Office (USPTO’s) Mission is to foster innovation, competitiveness and job growth in the United States by conducting high quality and timely Patent and Trademark examination and review proceedings in order to produce reliable and predictable intellectual property rights. USPTO accomplishes this mission by guiding intellectual property policy, improving intellectual property right protection, and delivering intellectual property information and education worldwide.

2 Purpose

The purpose of this RFI is to conduct market research and seek expertise in Artificial Intelligence (AI) capabilities as they specifically relate to “search” during patent application prosecution and to assess vendor/organization capabilities and interest. As the world's body of knowledge increases and tools expand to access information, the ability of individuals to review and filter this content is increasingly challenging. The USPTO is interested in the use of advanced technology to supplement and improve USPTO search capabilities with differentiating solutions that utilize technological advances. Quantum computing, Artificial Intelligence (AI), machine learning, natural language processing, etc. are examples of technology that might assist in narrowing the broad information base to manageable sizes for more careful individual review. This RFI will explore developing market interest-capacity in supplying a lightweight plugin(s) and/or widget technology to augment current USPTO search capabilities. As a federal agency, it is important for the USPTO to be able to explain all prosecution decisions made. Because of this, solution capabilities must be transparent to the USPTO and as well as to the general public. Black box solutions will not be accepted.

“Patent prosecution” refers to the series of interactions between inventors, practitioners, and examiners that can result in a patent. When the Applicant submits their invention, the USPTO assigns a Patent Examiner to review all of the application claims to determine whether to grant the patent. One of the most important tasks of this Patent Examiner is to substantively review the disclosures in the application and compare them to the already existing references called prior art. "Prior art" is defined as all information that has been made available to the public in any form before a given date that might be relevant to a patent application's claims of novelty and non-obviousness. Examiners are challenged with searching all of an ever increasingly complex and vast corpus of human knowledge in a limited amount of time. They must read and understand the patent application, perform an extensive prior art search, citing references to determine new contributions made or to explain the basis for rejecting the application. “Search” has become an increasingly challenging aspect of patent examination because:

- Innovators, researchers, and academia publish new ideas every day. There is an exponential growth of information for search and the examiner may be searching for related information to something that does not exist during patent prosecution. While most internet searches are similar to a “seek and find” book, prior art search for patent prosecution is more similar to looking for a needle that did not already exist in an ever growing haystack.

- Language matters and terms evolve. Patent prior art search needs categorization and context. Synonyms for keywords are technology specific. A “synonym relationship” may exist between keywords in one technology area, but no “synonym relationship” may exist between the same two keywords in another technology area.
Applicants can be their own lexicographer, defining terms used within the application. Innovation often spurs new words and phrases beyond common vocabulary.

Finally, because the patent examiner time is limited, they need better results, not more results. The components of a better prior art search often includes classification, ranking or relevancy, combined with context to help examiners make their determination of non-obviousness and novelty.

3 Objectives

The USPTO is seeking industry input and capabilities that can integrate into the technology detailed in Table 1 and tangible ideas, products or technologies that could be integrated into the existing PESE search capabilities. USPTO is looking for addresses the following topics:

a) Expanding the knowledge corpus beyond patents – “How can we effectively and consistently search further?”
   o Current search architecture relies on USPTO ingesting a copy, storing, and indexing locally or subscribing to multiple services. Is there an alternate approach or technical solution to improve information retrieval strategy?

b) Algorithmic classification of knowledge corpus – “How can we segment information more effectively?”
   o Examiners rely on faceted information to address the expanding taxonomy of innovation – if documents are not consistently classified, classification based search strategies can result in unintended ‘gaps’. Example:
     ▪ Classification and indexing of non-patent documents, literature, and references used as prior art by utilizing accepted classification categories such as the Cooperative Patent Classification (CPC)

c) Machine retrieval of representative prior art within corpus – “How can the USPTO patent search retrieve the most relevant results, not just more results from a broader corpus?
   o Examiners typically rely on tightly clustered result sets derived from Boolean search strings that can result in many similar references.
   o Machine selection/presentation can curate ‘samples’ of knowledge to increase, influence or broaden effective coverage search areas; thereby influencing examiners to look under different rocks (force multiplier). Examples:
     ▪ Capture uniquely defined vocabulary and identify related terms that further explain the context or usage
     ▪ Pinpoint keywords, phrases, synonyms, hypernyms, and/or hyponyms that capture the point of novelty

d) Summarization/knowledge extraction of retrieved information – “How can we review the search results more quickly?”
   o Given the fixed amount of time examiners need to quickly draw conclusions regarding retrieved information, in addition to ‘keyword highlighting’, can document summarization techniques help employees determine if information is or is not relevant? Examples:
- Ability to assess a patent application against prior art to identify potential “uniqueness” of the disclosure/claims, indefiniteness, double patenting, new matter, flags for sequence cases

e) Capturing search “journey” and identifying when is it complete – “When do we know when we are done?” Examples:
  o Develop stimulus to identify when a search is no longer productive to bound the search activity but supports consistency and public confidence
  o Automated recording of search activities to address perceived “gaps”, facilitate training, sharing of best practices, machine learning, etc.

In addition to the objectives in section 3, the solution response to this RFI should take into consideration the ability to scale and adapt to the following:

- Ability to continually and incrementally ingest reference disclosures (US Patents and pre-grant Publications, foreign patent abstracts or full text translations, and non-patent literature to generate a corpus of literature against which automated searches can be performed.
- Ability to train for different technical areas and patent fields such as chemical, mechanical, biological etc.
- Ability to interface with the IT solution to provide the results programmatically, for example, through an API (Application Programming Interface).
- Ability to integrate and utilize current technical stack for the PE2E search detailed in Table 1:

### Table 1 - PE2E Search Technical Stack

<table>
<thead>
<tr>
<th>#</th>
<th>Technology</th>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>jQuery</td>
<td>2.1.4</td>
<td>UI Framework</td>
</tr>
<tr>
<td>2</td>
<td>Chrome</td>
<td>V68</td>
<td>Search Application is directed to only support Chrome</td>
</tr>
<tr>
<td>3</td>
<td>RestEasy</td>
<td>2.3.1 GA</td>
<td>Restful API Implementation</td>
</tr>
<tr>
<td>4</td>
<td>Solr</td>
<td>5.4.0</td>
<td>Reviewing upgrade path to Solr 7.4.0</td>
</tr>
<tr>
<td>5</td>
<td>Java</td>
<td>1.8.0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Zookeeper</td>
<td>3.4.6</td>
<td>Zookeeper is used to manage Solr configurations. The ensemble consists of three zk nodes.</td>
</tr>
<tr>
<td>7</td>
<td>JBoss</td>
<td>6.4.1 EAP</td>
<td>Both UI and API wars leverage JBoss 7.</td>
</tr>
<tr>
<td>8</td>
<td>ImgScalr</td>
<td>4.2</td>
<td>Image Scaling library</td>
</tr>
<tr>
<td>9</td>
<td>JAI</td>
<td>1.1.3</td>
<td>Java Advanced Imaging API</td>
</tr>
<tr>
<td>10</td>
<td>Apache Tika</td>
<td>1.2</td>
<td>Data extraction and parser library</td>
</tr>
<tr>
<td>11</td>
<td>Flying Saucer</td>
<td>9.0.4</td>
<td>XML/XHTML to PDF Library</td>
</tr>
<tr>
<td>12</td>
<td>Swagger</td>
<td>1.5.3</td>
<td>Swagger JSON for API documentation</td>
</tr>
<tr>
<td>13</td>
<td>OpenCV</td>
<td>2.4.9-7</td>
<td>Image Conversion Library</td>
</tr>
</tbody>
</table>
For other information that may be useful:

- Publicly available USPTO patent data:
  - Bulk Data Storage System – [https://bulkdata.uspto.gov/](https://bulkdata.uspto.gov/)
  - Patents View – [https://www.patentsview.org](https://www.patentsview.org)

- Non-patent literature including, but not limited to, scientific journals

4  Information Request

This RFI seeks to obtain information from interested parties to arrive at a most suitable approach for meeting the objectives included in Section 3. The USPTO invites vendors and all interested stakeholders to assist with this market research by providing responses in accordance with the instructions provided throughout Section 5, Vendor Responses. Interested parties should respond in full to the topics listed in Section 3, Objectives. The overall intent is to supplement and improve the existing Patent End to End (PE2E) search with differentiating solutions.

To fully comprehend the information contained within a response to this RFI, there may be a need to seek further clarification from those respondent(s) identified as capable. Based on the requested information provided, USPTO may request a capability demonstration conducted in person or via remote testing system as arranged. The USPTO reserves the right to seek additional information from those vendors identified with unique solutions that are determined to be beneficial to the USPTO. USPTO is not paying for any costs associated with the market survey to include any submissions, follow up or demonstrations.

Based on those responses, the USPTO may determine a pool of vendors that are deemed most likely to succeed, and may invite those companies to participate in a PTAG Alternative Competition Method, in accordance with the Patent and Trademark Office Acquisition Guidelines (PTAG) and the Patent and Trademark Office Efficiency Act 35 U.S.C. 2(b)(4)(A).
5 Vendor Responses

Responses to this RFI shall be submitted by e-mail to Carlos Carter Sr., Procurement Liaison, at carlos.carter@uspto.gov and must be received at USPTO no later than 1:00 PM EST on 11/2/2018. When submitting response to this RFI include the following in the subject line. (RFI Responses – Improve USPTO’s Patent Search with AI) Interested parties shall provide a white paper describing their approach to address/solve, the objectives listed under Section 3, Objectives. In addition, provide a capability statement in accordance with bullet 1 of Section 5, Vendor Responses. The white paper submission, capability statement and any attachments shall be in sufficient detail, but not exceed 15 pages total and should address the objectives in Section 3, Objectives this will help the Government determine the experience, approach and capability of your firm. Responses must include:

- Capability Statement (1 page)
  - Entity Information
    - Business/Organization name, address, url
    - Name of representative, alternate, and their associated titles and contact information (inclusive of telephone numbers and email addresses);
    - Background and experience with subject matter

- White Paper (10 pages)
  - Objective topic area(s) being addressed
  - Solution description, including how it differentiates from any current, commonly available solutions; improvements achieved; and thought processes or methodologies utilized, e.g. libraries, open source code, APIs, etc.

- Attachments (4 pages)
  - Supporting resources and/or references utilized
  - Tables, charts, graphical visualizations

Please be advised that all submissions become Government property and will not be returned. Not responding to this RFI does not preclude participation in any future RFP, if issued. The information provided in the RFI is subject to change and is not binding on the Government. It is the responsibility of the interested parties to monitor www.fbo.gov for additional information pertaining to this RFI.

6 Questions

Questions to this RFI shall be submitted no later than 10/1/2018 01:00 PM EST. When submitting questions in response to this RFI include the following in the subject line. (RFI Questions – Improve USPTO’s Patent Search with AI). Questions shall be submitted by email only to Carlos Carter Sr., at carlos.carter@uspto.gov.

If the USPTO determines that a question received is significant enough to impact all industry responses; then the USPTO may issue an amendment to this notice with any answers to questions by no later than 10/15/2018, 2018 at 1:00 PM EST.