

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: Request for Comments and
Notice of Roundtable Event on the Use
of Crowdsourcing and Third-Party
Preissuance Submissions to Identify
Relevant Prior Art

Docket No. PTO-P-2014-0013
79 Fed. Reg. 15,319

COMMENTS OF PUBLIC KNOWLEDGE

Attn: Nicole Dretar Haines
Mail Stop Comments—Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450
CrowdsourcingRoundtable2014@uspto.gov

Public Knowledge respectfully submits the following comments in response to the Request for Comments on the Use of Crowdsourcing and Third-Party Preissuance Submissions to Identify Relevant Prior Art dated March 12, 2014.

Crowdsourcing is a two-way street: success requires both internal data made open to the public, and a large external population to analyze that data. Briefly, Public Knowledge provides three recommendations to maximize the effectiveness of crowdsourcing and third-party preissuance submissions.

- **Opening up more data.** In particular, file history data is viewable to humans but not reachable by the automated computer tools that are the foundation of successful crowdsourcing technologies. Thus, the PTO should find ways to make this and other data more accessible.
- **Ensuring completeness of file wrappers.** Full information on the details of prosecution, especially on interview summaries, is critical input for useful crowdsourcing and third-party submissions.
- **Partnering with the open source and startup communities.** These groups, who do not always interact with the PTO, are valuable sources of information necessary to the patent system.

I. Effective Crowdsourcing Depends on Opening Up More Data

As the PTO is no doubt aware, computer technology has opened up myriad opportunities for data processing and analysis, leading to tools, relations, and discoveries that were previously unknown. Crowdsourcing is one such innovation: the ability of the Internet to connect numerous people to work on a task has opened up a new range of activities and goals that can now be achieved.

Crowdsourcing depends on data being available for the crowd to analyze and access. Much of that data is prior art data, available outside the PTO. But crowdsourcing can, and ought to, take advantage of the volumes of data within the PTO itself. Thus, making more data available to crowdsourcing technologies is a necessary precondition to making the most of crowdsourcing.

In many ways, the PTO has gone far in making patent data accessible, by putting up patent and application data in XML form. The benefit of doing so has been clear: numerous businesses, research studies, and software tools have arisen as a result of these easily-analyzed patent data sets.¹

In one notable area, however, patent data is currently not as easily accessed: the file wrappers specifying the prosecution histories of patent applications. File wrapper data is available from Public PAIR, but the website interface is not designed for bulk data gathering, the site implements roadblocks to automated processing,² and the terms of use for Public PAIR prohibit bulk data processing.³

File wrapper information is essential to increasing the effectiveness of crowdsourcing and third-party prior art submissions. File wrappers indicate which applications are more important to the applicant (based on the length of prosecution, number of IDS references submitted, and so on), and thus indicate which applications warrant closer scrutiny by the crowd. File wrappers also identify prior art relevant to an application and, importantly, how that art is relevant to the application; this provides a

¹ Services such as Lex Machina (<https://lexmachina.com/>) and Ocean Tomo (<http://oceantomo.com>) are examples of this.

² For example, access to public PAIR is restricted by CAPTCHAs, to prevent automated programs from accessing the site.

³ See *PAIR Usage Policy* (last modified Nov. 5, 2009), http://www.uspto.gov/patents/process/status/private_pair/PAIR_Usage_Policy.jsp.

critical starting point for third parties identifying relevant art. Furthermore, file wrappers contain numerous clues about claim construction. These clues must be available to effective crowdsourcing efforts which are important to ensuring that crowdsourcing efforts, especially in view of the fact that claim language, standing alone, is often vague and impenetrable. Thus, making file wrapper information accessible to the tools of crowdsourcing and data analysis will have substantial benefits to patent examination.

If this information were made accessible to the public and to industry in a manner amenable to large-scale processing, one can imagine the number of socially beneficial tools and services that could be developed. For example, one tool might identify sets of similar applications based on the same art being cited in those applications, and then determine whether information in one of those similar applications is relevant to others. Another tool might analyze language distinguishing a patent application from prior art, to automatically identify relevant art as to the drawn distinctions.

From the advances in patent analysis that have already been made possible through bulk availability of issued patent and publication data, it is clear that accessibility of data will spur industry and researchers to develop new and useful services that will improve the patent system overall. This is the essence of crowdsourcing: by placing an privately held resource in the hands of the public, the public will put that resource to new, innovative, unexpected uses that are greater than what could have been achieved in-house or through limited partnerships.

Thus, we urge the PTO to identify ways to make more patent data, and particularly file wrapper data, available and accessible to the public for bulk processing.⁴ There are undoubtedly technical difficulties in doing so—as there presumably were with making bulk patent and publication data available—and undoubtedly there are organizations and entities that would gladly provide assistance and input.

⁴ There are at least two ongoing efforts toward making bulk application data available, namely Google's and Reed Tech's scraping of Public PAIR. These projects are limited in value to the particular crowdsourcing and third-party submission interests of the PTO, however, because they only capture snapshots of applications at the time of scraping, rather than the newest changes in applications being actively examined. Thus, the PTO should look to alternate ways of providing application data besides a lengthy external scraping process.

II. Effective Crowdsourcing Depends on Complete Prosecution Records, Especially of Interviews

In addition to making bulk file wrapper data available, the PTO should ensure that file wrappers include complete information about the prosecution of applications, as explained in our comments in a related proceeding.⁵ Effective crowdsourcing and third-party submissions depend on the availability of complete information about a patent application. For example, a potential third-party submitter might only be able to identify relevant prior art if it understands the particular meaning of a term in a claim of the application. Having a clear file wrapper that identifies any possible interpretation of that claim term, any disclaimer in scope, or other information discussed between the examiner and the applicant will help the potential third-party submitter decide whether to prepare a submission.

In particular, as we recommended in our previous comments, the PTO should make efforts to ensure that relevant information from interviews is placed in the file wrapper. Interview practice is a particular problem for transparency with regard to claim interpretations. Generally interviews are productive activities in prosecution of applications, because the examiner and applicant can have a meeting of the minds and agree on the meaning of the claims and distinctions from the prior art. That discussion is of great relevance to claim interpretation, and thus to crowdsourcing efforts as explained above, so the interview content ought to be made of detailed record in the file wrapper.

III. Effective Crowdsourcing Depends on Partnerships with the Open Source and Startup Communities

As we explained in comments in another related proceeding,⁶ the PTO should work with the software startup community and the open source software community to

⁵ See Pub. Knowledge & Elec. Frontier Found., *Comments before the U.S. Pat. & Trademark Office on Strategies for Improving Claim Clarity: Glossary Use in Defining Claim Terms* (Oct. 24, 2013), available at <http://www.publicknowledge.org/files/comments-pto-roundtable.pdf>.

⁶ See Pub. Knowledge et al., *Comments Before the U.S. Pat. & Trademark Office on Prior Art Resources for Use in the Examination of Software-Related Patent Applications* (Mar. 17, 2014), available at <http://www.publicknowledge.org/assets/uploads/blog/uspto-prior-art-comments.pdf>.

maximize the effectiveness of crowdsourcing and third-party prior art submissions. As stated in those comments: “These groups produce much of the most advanced software technology today, meaning that they are among the most valuable sources of prior art in the software field. But they often lack the resources of large companies, so they are less likely to file patent applications or otherwise generate prior art in ways traditionally expected by the PTO. Accordingly, successfully harnessing this body of prior art will require collaboration between the PTO and these communities.”

IV. Conclusion

For the foregoing reasons, we recommend that the PTO (1) make more data, especially file history data, available in an accessible form; (2) ensure that file wrappers contain a complete record of prosecution, especially exchanges between applicants and examiners during interviews; and (3) partner with the open source and startup communities to develop effective solutions to improving the patent system.

Respectfully submitted,

Charles Duan
Director, Patent Reform Project
Public Knowledge
1818 N Street NW, Suite 410
Washington, DC 20036
(202) 861-0020
cduan@publicknowledge.org

April 23, 2014