There is no API for Google Scholar, however, python code exists that parses the Google Scholar output.

<http://www.icir.org/christian/scholar.html>

* “Can extract publication title, main online URL, number of citations, number of online versions, link to Google Scholar's main cluster for the work, and Google Scholar's cluster of all works referencing the publication.
* Can print entries in CSV format or plain text.”

While this exact solution won’t work for us and is missing several features we could use, it may be a good starting point for working with Google Scholar.

There is a free product that seems to do exactly what we’re looking for by querying Google Scholar and Microsoft Academic Search:

<http://www.harzing.com/pop.htm>

Whether or not this product is a viable option (I’m going to assume it’s not, given that the crawler is a significant portion of the project), the [FAQ](http://www.harzing.com/pophelp/faq.htm) provides a lot of valuable information regarding Google Scholar query limits and things of that nature.

Microsoft Academic Search, unlike Google Scholar, does provide an API. I’m not sure if it will be as effective (I don’t really have data to test it with) but it could prove useful:

http://academic.research.microsoft.com/About/Help.htm#4

We may want to ask if <http://libguides.mit.edu/apis> references anywhere that could contain information that might be useful to this aspect of the project.