# Introducing Sysrev: The Intelligent Platform For Document Review And Automated Data Extraction

July 22, 2019   by Thomas Luechtefeld

In today's technological world, data is perhaps the single most important driver of a business' success. Access to relevant data allows businesses to make a variety of informed decisions. Unfortunately, acquiring this data can be quite cumbersome as employees spend countless hours manually reviewing documents. This is especially true for more complex reviews such as journal publications, patient records, or technical specifications. Sysrev offers enterprise a platform for managing collaborative document reviews, injecting machine learning into the review process to increase accuracy and efficiency. Depending on the data source and task, Sysrev can even automate data extraction.

Sysrev, launched in June 2019, is an intelligent platform for document reviews and automated data extraction. Sysrev optimizes the review process with machine learning and adds efficiency through its intuitive, and collaborative, interface.
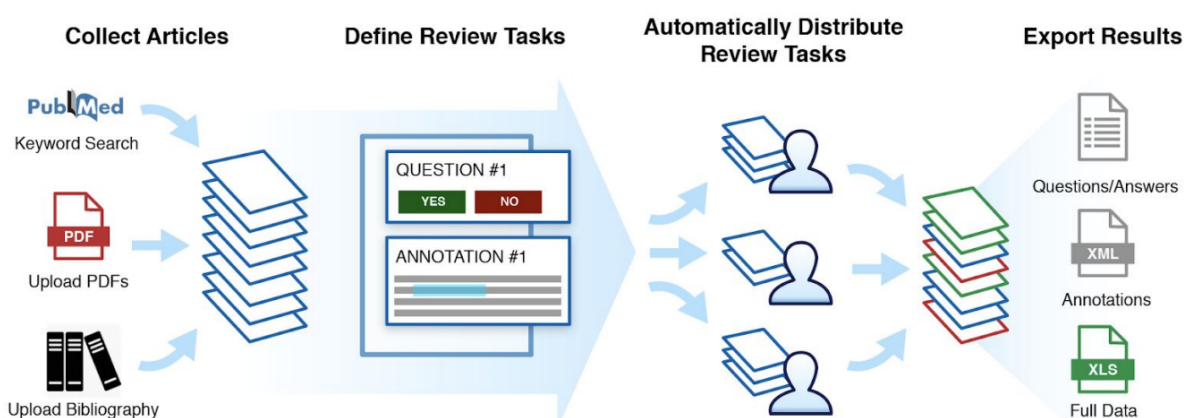
Graphic showing the document review workflow on Sysrev.com

What is a document (or systematic) review

A document review is a process of reviewing documents (pdfs, paragraphs, word files, etc.) for relevant data according to a well-defined methodology. Each systematic review has four steps:

1. Set a clear Objective - what is the question being answered

2. Search for relevant documents

3. Screen documents and Extract data

4. Report results

The purpose of any document review is defined by its Objective. Once an Objective has been defined, reviewers must collect relevant literature and/or documents. Collecting documents is by no means a simple task. If some documents are missing then the resulting conclusions are tainted but collect too many documents and you waste time and money reviewing irrelevant documents. After the collection is complete, each document is reviewed for inclusion before relevant data is extracted. Inclusion criteria must be well-defined for accurate results. After the data has been extracted, the researchers report their results.

Optimizing the Screening and Extraction Process

Syrev optimizes the screening and extraction process with the combination of machine learning, collaboration, and automated project management. To get started on Sysrev, users must first create a Project. Once a project has been created, Account Admins are able to set a number of review protocols. These protocols, such as 'how many reviewers should review each article', are important in balancing the efficiency and accuracy of the review process. The Account Admin also has access to a project management dashboard which gives insight into each reviewer's performance.

Once the protocols have been set, reviewers begin analyzing individual articles and extracting data. As documents are screened, Sysrev's artificial intelligence learns from the reviewers' decisions which articles are important. With sufficient training, Sysrev can confidently predict which articles are relevant, saving researchers countless hours. For more structured tasks, Sysrev can even learn to automate the entire data extraction process.
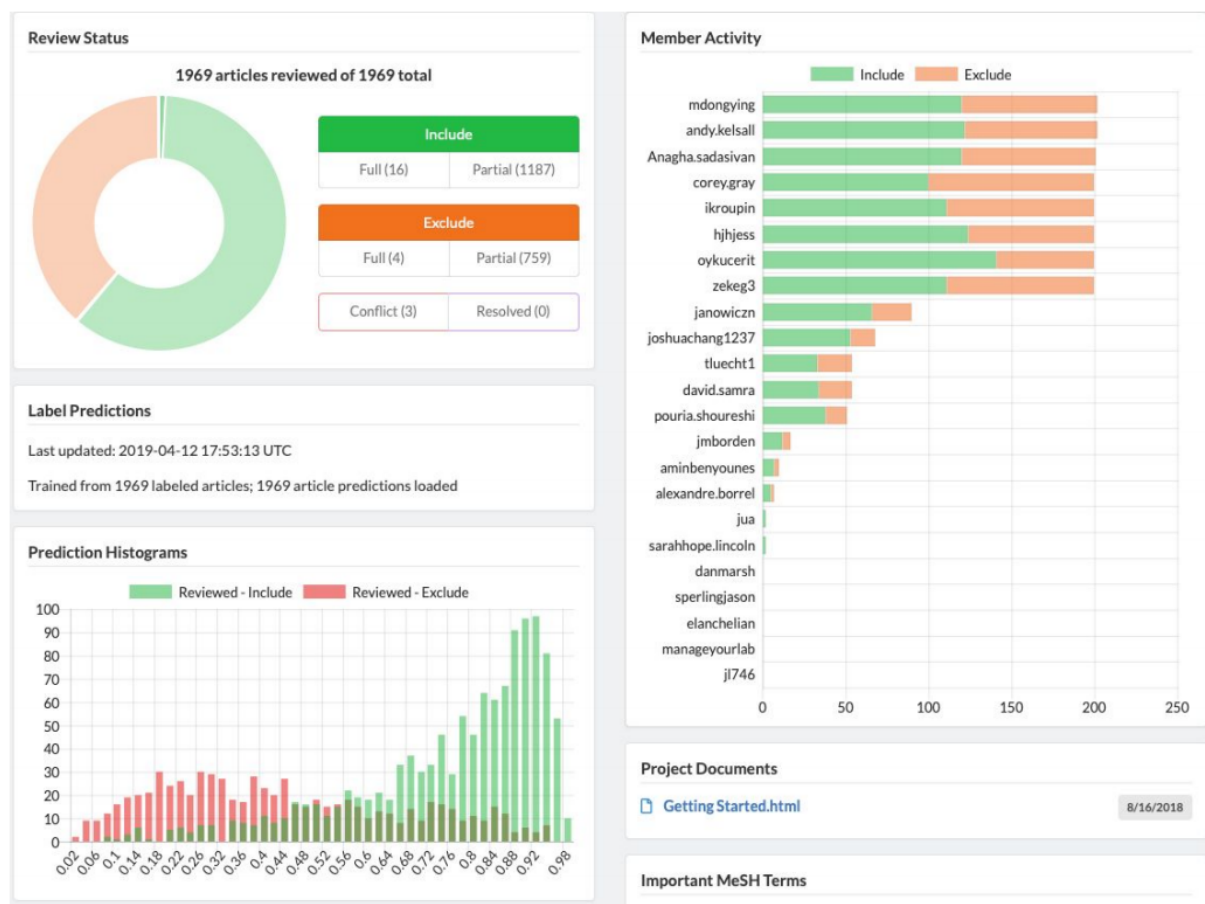
Dashboard for the Gene Hunter Project - a *Public Project* in which over 2000 articles were reviewed by 3rd party reviewers

"Living Reviews" & Data Streams

One of the major issues with document reviews is new data -- that is, data published since the last literature collection. Sysrev solves this problem through Living Reviews. Instead of closing Projects after the data extraction process, Sysrev keeps a log of the entire review --including which articles were (or were not) included. In this way, users can simply re-open an idle Project, import any missing literature, and extract the relevant data. The process of importing new literature can be automated by setting up custom data feeds.

These feeds can consist of structured or unstructured data. Whenever new literature arises in the data feed, Sysrev sends an alert to the Account Admins that there is documentation to be reviewed. For more structured tasks, this process can be further automated to create true Data Streams. That is, Sysrev will automatically identify literature and extract relevant data, aggregating the new and old information into a single database. Finding Trends and Creating Systematic Maps: A Case Study Sysrev's ability to

automate extraction tasks has applicability beyond simply literature reviews. As shown below, the Gene Hunter App utilizes data from a literature review to inform a systematic map that associates genes with a variety of medical terms.

The literature review

One of the first Public Projects on Sysrev was the Gene Hunter Project. In the Gene Hunter Project, genes were extracted from nearly 10,000 sentences. This review was conducted by a variety of paid 3rd party reviewers on the Sysrev platform.

Visualization of the extraction task - identifying genes within text

After the genes had been identified within text, the data was used to train a Named Entity Recognition (NER) model. This model, when given a block of text, has the ability to identify and extract genes. While an interesting algorithm on its own, the real opportunity begins when the model is tied into different data sources. The Gene Hunter App is such a system.

The Gene Hunter App uses the NER model developed by the Gene Hunter Project to analyze titles and abstracts within Pubmed. The result is a system that identifies which genes are associated with different 'medical terms'. As the system is analyzing natural language, the term itself can be quite varied. In fact, one can even input a researcher's name into the Gene Hunter App to see on which genes they have published. The ability to incorporate an automated learning task into a larger system has countless possibilities.

whichgenesmatter.com: depicting genes associated with the medical term 'diabetes' with specific Pubmed IDs
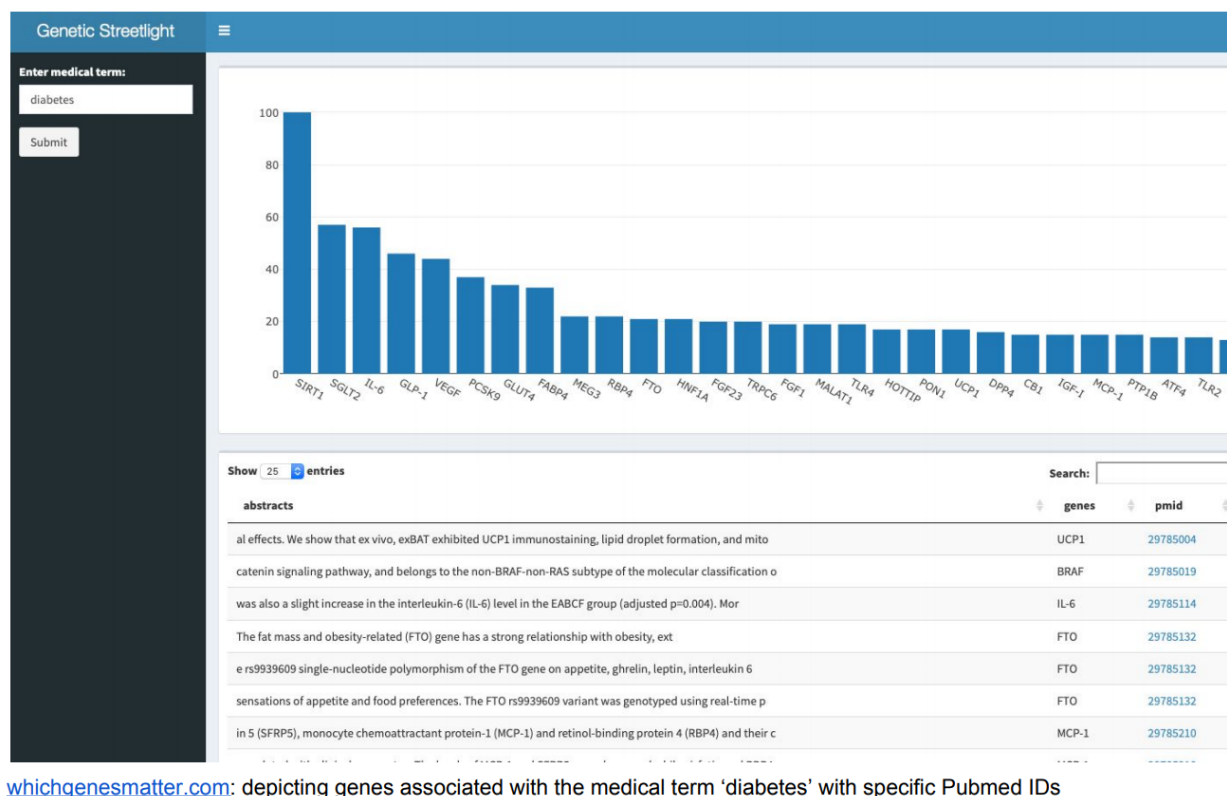
Sysrev is an intuitive and flexible platform for document reviews and associated tasks. Sysrev offers unique project management and machine learning technologies that optimize document reviews across entire teams or departments. Its ability to automate extraction tasks can drastically reduce the man-hours associated with data extraction tasks. Moreover, coupling automated data extraction with data streams or new data sources can result in powerful systems like the Gene Hunter App.