A Visual Tool for Displaying Annotations in BioC

Soo-Yong Shin[1], Dongseop Kwon[2][*]

[1]Department of Biomedical Informatics, Asan Medical Center, Seoul 05505, Korea
sooyong.shin@amc.seoul.kr;
[2] Department of Computer Engineering, Myongji University,
Yongin, Gyeonggi-do 17058, Korea
dongseop@mju.ac.kr

Abstract. To support data interoperability for annotation results, we developed BioC-Viewer (http://viewer.bioqrator.org/). BioC-Viewer is a web-based interactive curation tool to visualize annotation results from text mining tools and to easily curate entities and relationships with supporting BioC format. Since our focus was on a visual tool for BioGRID curators in BioCreative V BioC Track, BioC-Viewer displays annotations on full-text BioC documents, and manages protein-protein interactions and genetic interaction pairs found in the documents. BioC-Viewer has a user-friendly interface and users can perform most of operations with a minimum number of mouse clicks.

Keywords: BioC; Curation Tool; Visual Tool; Interoperability

1   Introduction

A practical problem that biocurators or text mining tools face is the lack of data exchange format. Annotation results from different text mining tools or human experts are often not interoperable since there is no standard format. To overcome this problem, BioC [1] was proposed as a simple standard for exchanging data in biomedical natural language processing. Since BioC is a simple XML format, it can be easily integrated to various tools for text mining. There are also public programming libraries in C++, Java, Perl, Python, Go and Ruby for easily reading and writing BioC documents [2]. Moreover, several tools have been developed to support BioC, which include tmBioC [3], BioQRator [4], and Egas [5].

To promote the usage of BioC format, Collaborative Biocurator Assistant Task (BioC) was organized in BioCreative V Workshop. The purpose of this track is to develop BioC-compatible systems to assists BioGRID curators [6]. As a subtask for this track, we developed BioC-Viewer (http://viewer.bioqrator.org/), which is a web-based interactive curation tool to support BioC format annotations. The primary goal of this tool is to support biocurators by visualizing annotation results from text mining tools and to make their curation process easier. Since the goal of the BioC track was to assist BioGRID curators, we focused on tasks for BioGRID curators. BioGRID curators should find protein-protein interactions (PPIs) and genetic interactions (GIs)

in a document. They also should assign experimental methods for curated interactions.

To support this, our tool provides two basic functionalities: (1) displaying annotations on full-text documents and (2) managing PPI or GI pairs found in the documents. One of the main goals of BioC-Viewer is to make an easy-to-use tool for curators. We designed our tool to perform most of operations with a minimum number of mouse clicks on a user-friendly interface.
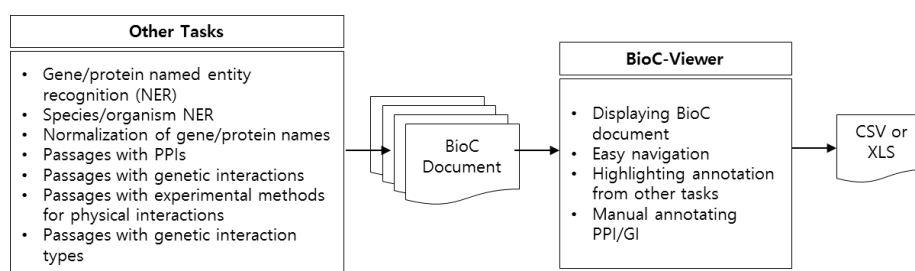
2   Design Goal

Figure 1. Workflow for the target task

The workflow of the BioCreative V BioC track is depicted in Figure 1. Results from other text mining tasks, such annotations for genes/proteins/organisms and PPI/GI passages, are imported to BioC-Viewer as BioC documents. BioGRID curators read the annotations in our viewer and record PPI and GI pairs which they think relevant. BioC-Viewer supports the following requirements for the BioGRID curation.

- BioGRID curators need to capture binary interactions only.
- There may be different interaction entries with the same pair of genes/proteins because of in the different experimental methods used.
- The order of genes/proteins in an interacting pair has a meaning. It is possible to have two interaction entries with the same interactors, but in different orders. For example, an interaction pair <A, B> is different from <B, A>.
- Annotations from other tasks may have normalized gene/protein IDs. In this case, the tool should the IDs automatically for curating PPI/GI pairs.
- The tool should provide an easy way to search gene/protein IDs, i.e. finding gene/protein IDs from external resources such as Entrez Gene[1].
- Results from other tasks may contain false predictions. Curators should easily edit interaction entries at any time.
- Some annotations from other tasks may have overlapped or nested regions between them. The tool should provide the means to display those regions.

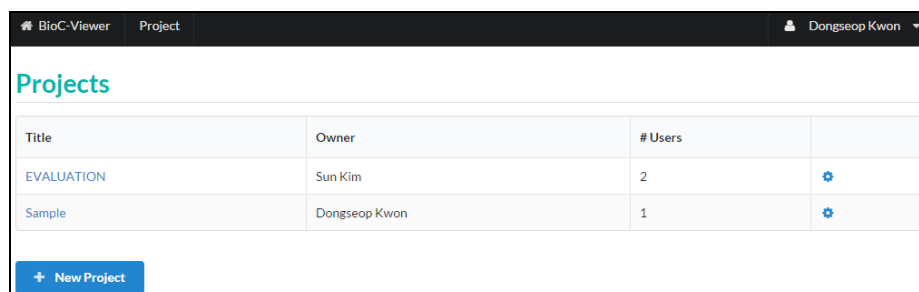[1]   http://www.ncbi.nlm.nih.gov/gene

3   Implementation

BioC-Viewer uses HTML5/CSS to support modern web browsers; hence it is compatible with recent versions of Chrome, Safari, Internet Explorer, and Firefox. It also uses 'simple_bioc' Ruby library [7] for parsing and building BioC documents. Our viewer was also implemented by using Ruby on Rails, MySQL, JavaScript, and Semantic-UI library.

BioC-Viewer consists of three main pages: (1) List of projects, (2) List of documents in a project, and (3) Document viewer. Users can sign up or log in BioC-Viewer with their email addresses. After log in to the tool, users will see a project list page by default.

3.1   List of projects

A project is the basic unit for collaboration in BioC-Viewer. A user can create a new project with an empty collection of documents, and add or remove BioC documents into the collection. A project can be shared with other collaborators (i.e. users) for review and curation. Administrators of a project can set permissions for other users by clicking a setting button in a row in the project list. The tool offers three different permissions: read permission for read-only, write permission for allowing curation, and the admin permission for managing documents or users for project. A project list page (Figure 2) shows all a logged-in user's projects including the ones that other users shared with. By clicking the title of a project, a user can go into the detail page of the project.

Figure 2. Page for the list of projects

3.2   List of documents

A project has a collection of BioC documents as showed in Figure 3. In this page, a user can upload or delete BioC documents. A user can use a mouse drag-and-drop for uploading multiple documents at once. Clicking a document ID in the list will open a document viewer for the document. By clicking the setting button in a row, users will see a popup menu for performing the tasks such as deleting a document or exporting all curated PPI/GI pairs to a file.

Figure 3. Document list in a project

3.3 Document viewer

Figure 4. BioC document viewer with PPI/GI Annotator

The document viewer consists of 4 parts (Figure 4): the top button-bar (upper-side), the outline viewer (left-side), the text viewer (center), and the PPI/GI annotator (right-side).

(1) Top-button bar has annotation-type toggle buttons on the right side. With the toggle buttons, a user can highlight only selected annotation types in the main text viewer.

(2) The outline viewer provides a bird's-eye-view of a document. The tool analyzes the structure of the BioC document, and displays the hierarchy of section titles. A user can jump to a specific position in a document by clicking an item in the outline viewer. BioGRID curators need to review all PPI/GI passages quickly. For this reason, the outline viewer highlights the sections that include PPI/GI annotations in purple. A gray-colored box indicates no PPI/GI annotations in the section.

(3) The text viewer displays annotations and other information in a full-text document. BioC-Viewer also highlights overlapped or nested annotations from different tools. For example, in Figure 4, the title of the document in the text viewer has two different types of annotations. The whole sentence is annotated as a PPI passage from a tool from other tasks. Meanwhile, some words in the sentence such as 'Aip1p', 'Cofilin', and 'Actin' are annotated as gene/protein names from another tools. Here, the gene/protein names are nested into the PPI annotation. BioC-Viewer highlights both annotation types by using underlines and background colors. In addition, BioC-Viewer can highlight overlapped annotations by dividing the overlapped region into multiple non-overlapped segments. Most visual tools based on HTML/CSS have difficulties to display such annotations because HTML does not allow overlapped HTML elements.

(4) The PPI/GI annotator, which is the right side of the text viewer, provides functionality for managing PPI or GI pairs found in a document. The upper side of the PPI/GI annotator has a form to create a new interaction pair. In addition to typing text manually, a user can input names or IDs into the form by clicking gene/protein annotations from the text viewer. A user can also select a type (i.e. PPI or GI) and an experimental method from pop-up menus. All curated PPI/GI pairs appear in the bottom of the form. Since the curated pairs are saved in the server, users of a project can see and edit all the interactions at any time in any place. Users can download curated information from the server as a form of a comma-separated file (CSV) or an EXCEL file (XLS).

4 Acknowledgment

REFERENCES

1. Comeau D.C., Doğan R.I., Ciccarese P., et al. BioC: a minimalist approach to interoperability for biomedical text processing. Database, (2013) Vol. 2013: article ID bat064; doi:10.1093/database/bat064
2. Liu,W., Islamaj Doğan,R., Kwon,D., et al. BioC implementations in Go, Perl, Python and Ruby. Database (2014) Vol. 2014: article ID bau059; doi:10.1093/database/bau059
3. Khare, R., Wei, C.-H., Mao Y., et al. tmBioC: improving interoperability of text-mining tools with BioC, Database (2014) Vol. 2014: article ID bau073; doi:10.1093/database/bau073
4. Kwon, D., Kim S., Shin, S.-Y., et al. Assisting manual literature curation for protein-protein interactions using BioQRator. Database (2014) Vol. 2014: article ID bau067; doi:10.1093/database/bau067
5. Campos, D., Lurrenco, J., Matos, S., Oliveira, J. L., Egas: a collaborative and interactive document curation platform. Database (2014) Vol. 2014: article ID bau048; doi:10.1093/database/bau048

6. BioCreative V Track 1 – Collaborative Biocurator Assistant Track (BioC), http://www.biocreative.org/tasks/biocreative-v/track-1-bioc/ (accessed August 26, 2015).
7. Simple BioC parser/builder for Ruby, https://github.com/dongseop/simple_bioc (accessed August 26, 2015)