



Intelligent Information Management
Targeted Competition Framework
ICT-2011.4.4(d)

Project **FP7-318652 / BioASQ**

Deliverable **D3.6**

Distribution **Public**



<http://www.bioasq.org>

Annotation Tool, 2nd Version

Norman Heino

Status: Final (Version 1.0)

December 2013

Project

Project ref.no.	FP7-318652
Project acronym	BioASQ
Project full title	A challenge on large-scale biomedical semantic indexing and question answering
Project site	http://www.bioasq.org
Project start	October 2012
Project duration	2 years
EC Project Officer	Martina Eydner

Deliverable

Deliverable type	Report
Distribution level	Public
Deliverable Number	D3.6
Deliverable title	Annotation Tool, 2 nd Version
Contractual date of delivery	M15 (December 2013)
Actual date of delivery	December 2013
Relevant Task(s)	WP3/Task 1.3
Partner Responsible	NCSR "D"
Other contributors	TI, UJF, ULEI, UPMC, AUEB-RC
Number of pages	6
Author(s)	Norman Heino
Internal Reviewers	A. Krithara
Status & version	Final
Keywords	project templates, latex, reports

Executive Summary

The BioASQ Annotation Tool has been improved for the second year of the project. Among bug fixes, feedback from the experts who used the first version of the tool was incorporated. Improvements in the 2nd version of the BioASQ Annotation Tool include:

- improved question list in order to provide a better overview when large numbers of questions are created,
- improved concept search that displays search rank and synonymous labels,
- change of the way RDF triples are presented to the user,
- structural editing of exact answers depending on the question type,
- improved annotation UI and workflow.

Contents

1	Introduction	1
2	Changes in Detail	2
2.1	General Improvements	2
2.1.1	Explicit finalization	2
2.1.2	Annotation storage improvements	2
2.1.3	Changes to statement search	2
2.1.4	Quick access to previously issued queries	3
2.2	Improved Question List	3
2.3	Improved Concept Search	3
2.4	Structural Editing of Exact Answers	4
2.5	Improvements to the Annotation UI	4

List of Figures

2.1	Improved question list UI.	3
2.2	Concept search UI with the search term highlighted in matched labels.	4
2.3	Structured editing of a list question.	5
2.4	Example of the grouped annotation UI	5

Introduction

The annotation tool was created with the aim of supporting the creation of the benchmark data¹ for the challenge Tasks 1b and 2b. As specified in the Description of Work of the project, this tool was specifically designed to enable the biomedical experts to create the gold standards for these tasks. In addition to features described in Ngonga Ngomo et al. (2013), the current Version 2 of the tool provides the following additional features:

- improved question list in order to provide a better overview when large numbers of questions are created,
- improved concept search that displays search rank and synonymous labels,
- change of the way RDF triples are presented to the user,
- structural editing of exact answers depending on the question type,
- improved annotation UI and workflow.

¹<http://at.bioasq.org>

Changes in Detail

2.1 General Improvements

2.1.1 Explicit finalization

Version 1 of the Annotation Tool required the experts to add a special character ('#') to the question text. This was done in order for them to be able to mark a question as finalized. A finalized question is one to be considered for benchmark inclusion. The second version of the tool provides an explicit UI element for such marking.

2.1.2 Annotation storage improvements

In Annotation Tool, Version 1 snippet annotations were stored as snippets of HTML. This led to several problems:

- For data export HTML snippets needed to be converted to index and offset values,
- in order to change an existing snippet, the old HTML needed to be stripped,
- a bug was found that caused illegal HTML when a new snippet was created overlapping an existing snippet.

The current version stores snippets always as pairs (i, k) , where i is the index of the first character of the snippet and k is the index after the last character of the snippet. The highlighted HTML to be rendered is always always computed from this representation when needed.

2.1.3 Changes to statement search

Related RDF triples are now grouped in sub graphs that can be added with a single click. An example of such related statements would be a triple that mentions a fact along with triples that associate the factual terms with human-readable labels.

2.1.4 Quick access to previously issued queries

As per request by the expert team, all queries are stored and can be retrieved from a drop-down list.

2.2 Improved Question List

The list of questions has been re-implemented as a resizable list instead of a drop-down list. This allows for arbitrarily many questions to be displayed at the same time. In addition, metadata about questions can be displayed in the list, which was not possible before. Currently, this is used to display the finalization state of a question.

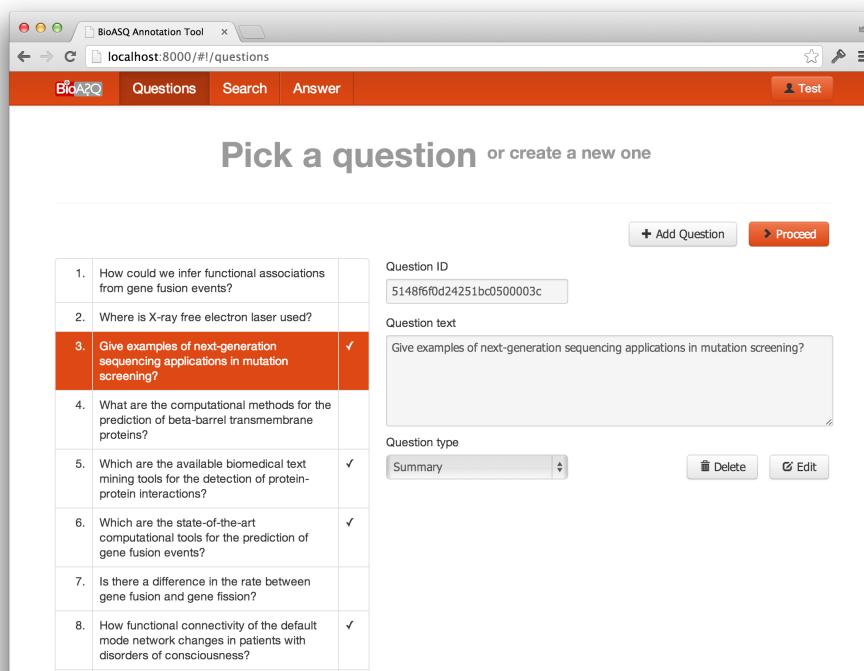


Figure 2.1: Improved question list UI.

2.3 Improved Concept Search

The concept search has been improved such that matched terms are now highlighted in the results. Concept identifiers can be associated with several string representations (so called labels), one of which is designated as the preferred or *canonical* label. In a concept that was found by matching against a non-canonical label (e. g. synonym) the matched label is now shown alongside the canonical label. In other parts of the tool only the canonical label is used.

An example for a concept query result is depicted in Figure 2.2. In the highlighted row the concept *Carbutamide* (canonical label) was found. The matched synonym for this concept was *Diabetal* (first column).

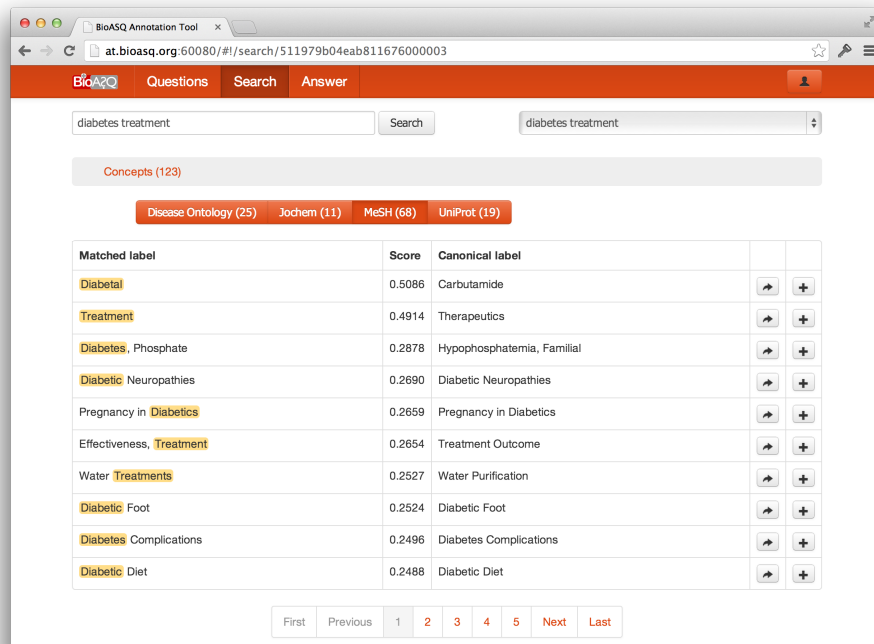


Figure 2.2: Concept search UI with the search term highlighted in matched labels.

2.4 Structural Editing of Exact Answers

Exact answers have a different structure depending on the question type. A *factoid question* has a single entity with optional synonyms as an answer. The exact answer of a *list question* is a list of entities, each with optional synonyms. For a Yes/No answer, only ‘Yes’ or ‘No’ are valid answers.

The previous version of the tool required experts to use a special syntax for each of these answer types. Unfortunately the syntax coincided with normal usage which resulted in ambiguous answers created by the experts. A manual disambiguation and curation step was thus necessary.

The current version of the tool does not require such workarounds. For each type of question a specialized user interface is provided that allows only such answers to be constructed that the question type demands. Figure 2.3 shows an example of an exact answer to a list question. The third list entry (‘InPrePPI’) is shown having the synonym ‘Integrated method for Prediction’.

2.5 Improvements to the Annotation UI

In the new version of the tool annotations are now grouped by type. Even cross-references between snippet annotations and their referenced document are available. This allows for much easier navigation of annotations and thus improves the expert’s workflow.

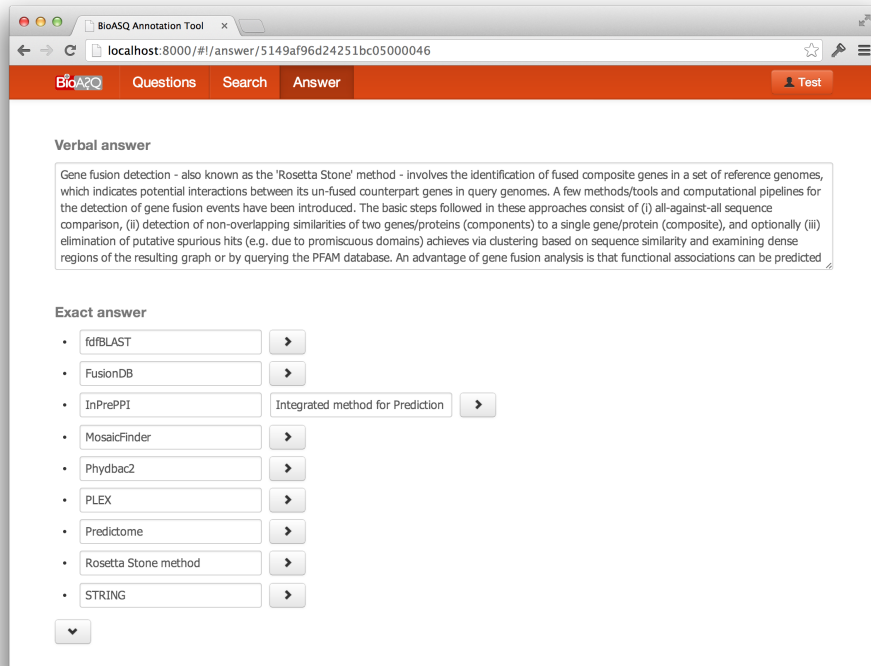


Figure 2.3: Structured editing of a list question.

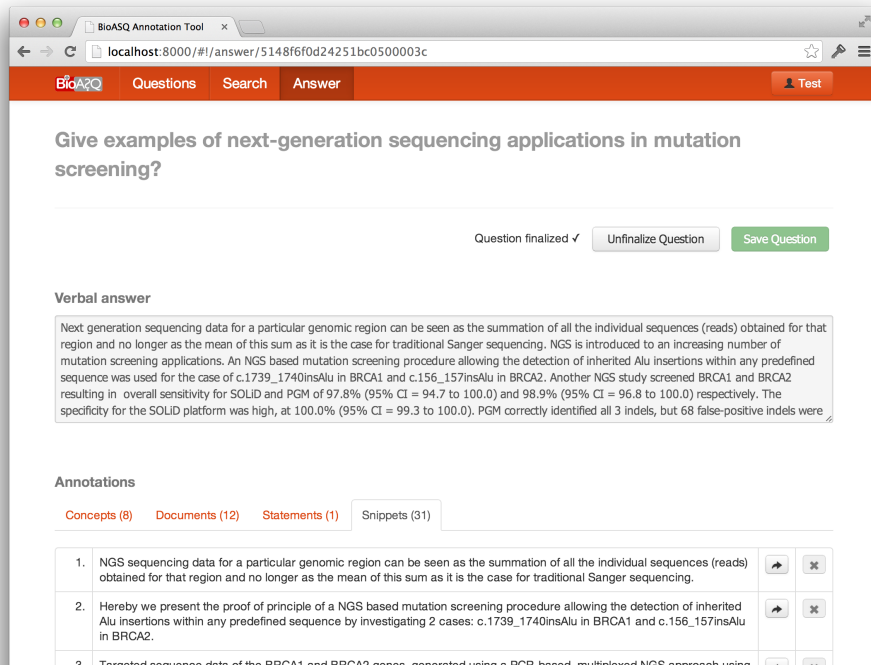


Figure 2.4: Example of the grouped annotation UI

Bibliography

A.-C. Ngonga Ngomo, N. Heino, R. Speck, T. Ermilov, and G. Tsatsaronis. Annotation Tool. Technical report, ULEI, 2013.