# Building a biomedical ontology recommender web service

Clement Jonquet[1§], Mark A. Musen[1], Nigam H. Shah[1]
[1]Center for Biomedical Informatics Research, Stanford University, CA 94305, USA

[§]Corresponding authors: Jonquet@stanford.edu, nigam@stanford.edu

Email addresses:
CJ: jonquet@stanford.edu
MAM: musen@stanford.edu
NHS: nigam@stanford.edu

## Abstract

### Background

Researchers in biomedical informatics use ontologies and terminologies to annotate their data in order to facilitate data integration and translational discoveries. As the use of ontologies for annotation of biomedical datasets has risen, a common challenge is to identify ontologies that are best suited to annotating specific datasets. The number and variety of biomedical ontologies is large, and it is cumbersome for a researcher to figure out which ontology to use.

### Methods

We present the *Biomedical Ontology Recommender web service*. The system uses textual metadata or a set of keywords describing a domain of interest and suggests appropriate ontologies for annotating or representing the data. The service makes a decision based on three criteria. The first one is *coverage*, or the ontologies that provide most terms covering the input text. The second is *connectivity*, or the ontologies that are most often mapped to by other ontologies. The final criterion is *size*, or the number of concepts in the ontologies. The service scores the ontologies as a function of scores of the annotations created using the National Center for Biomedical Ontology (NCBO) *Annotator web service*. We used all the ontologies from the UMLS Metathesaurus and the NCBO BioPortal.

### Results

We compare and contrast our Recommender by an exhaustive functional comparison to previously published efforts. We evaluate and discuss the results of several recommendation heuristics in the context of three real world use cases. The best recommendations heuristics, rated ‘very relevant’ by expert evaluators, are the ones based on coverage and connectivity criteria. The Recommender service (alpha version) is available to the community and is embedded into BioPortal.

# Introduction

## Background

Biomedical ontologies are widely used to design information retrieval systems, to facilitate interoperability between data repositories, and to develop systems that parse, annotate or index biomedical data resources. Biomedical researchers use ontologies and terminologies to annotate (or tag) their data with ontology terms for better data integration and translational discoveries [1, 2]. The number and variety (formats, locations) of biomedical ontologies is now so large that choosing one for an annotation task or for designing a specific application is a difficult challenge. Besides, re-usability is a desired practice in ontology development both because the process of building an ontology from scratch is long and hard and because the community needs to avoid the multiplication of several competing ontologies to represent similar knowledge.

However, the process to choose a set of ontologies to use is oftentimes a hard, manual and time consuming task for researchers. Members of the National Center for Biomedical Ontology (NCBO) often get requests for suggesting an appropriate ontology for a certain domain or application. There are several uses cases for ontology recommendation:

- Re-use existing ontologies when constructing new ones;
- Identify the most appropriate ontology for a given domain;
- Support an annotation workflow.

Researchers lacking an appropriate ontology may need to reprocess or re-annotate their data or redesign their application later. They may also start to develop a new ontology instead of re-using a standard shared one. They may also miss insights they might have seen had they used the right ontology when integrating their datasets with other datasets [6].

Therefore, ontology recommendation has emerged has a key issue in biomedicine. The manner in which recommendation occur depends on user settings. In some cases, the recommendation process can be long and non-automatic; the user can participate in the process (e.g., answer questions to refine the query) to enhance the accuracy of results. In other cases, a quick and fully automated approach is required, such as when ontology selection occurs at runtime in an application. For example, Sabou et al. [3] presented the requirements of a semantic browsing application called *Magpie*. Magpie needs to identify the ontologies that offer maximum coverage of a web page topic in order to identify the concepts in the web page and provide users with related information. As another example in the biomedical domain, *Reflect* [4] recognizes and highlights gene, protein and small molecule names while browsing a web page. For each recognized entity, Reflect provides a description, related information and links such as to PubMed abstracts. Both Magpie and Reflect need an ontology selection approach; however, Magpie requires a fully dynamic and automatic method that must be called at runtime, whereas Reflect requires preselection of the ontologies or vocabularies to use during application design.

This paper focuses on providing a quick automated recommendation with minimal user burden. We considered two main recommendation scenarios differentiated by the type of input provided by the user:

- *Corpus-based recommendation*: Given a corpus of textual metadata describing some elements of a biomedical dataset, our system recommends appropriate ontologies to annotate the dataset with ontology concepts.
- *Keyword-based recommendation*: Given a set of keywords/terms representative of a domain of interest, our system recommends appropriate ontologies to consider for re-use or extension for researchers building new ontologies or semantic applications.

The related works show that these two scenarios are the most frequent in the ontology recommendation literature. They have important differences. For example, the keyword-based recommendation is potentially cleaner because it avoids the introduction of spurious terms. However, it cannot account for term frequency. If a term appears several times in a corpus, one may want to give to the ontology that contains the corresponding concept a better score as this ontology is more relevant to the domain of the corpus.

## Contribution

This paper describes the *Biomedical Ontology Recommender web service*, or *Recommender*. Given textual metadata or a set of keywords describing a domain, the Recommender suggests ontologies appropriate for annotating or representing the data. Sabou et al. [3] demonstrate that a single ontology rarely provides the complete coverage or application need. Therefore, ontology selection systems need to be able to return combination of ontologies as result.Appropriateness is evaluated according to three main criteria:

- Coverage: the ontology that best covers the given data;
- Connectivity: the ontology containing the terms that are most often mapped (or referred) to by other ontologies;
- Size: the number of concepts in the ontology.

To facilitate and encourage the annotation of biomedical datasets, we created the *NCBO Annotator Web Service* [5], which annotates textual data with ontology concepts. We call annotation a mapping between a textual data and an ontology concept that declares: particular data "is associated with" a particular concept. The *Annotator* scores each annotation based on whether the term found is a preferred name, synonym, ancestor term or mapped term of a concept mentioned in the text. Biomedical researchers can use the Annotator to automatically tag datasets with ontology concepts. For example, the Gminer project (http://gminer.mcw.edu) used it to annotate rat microarray experiments. We used it to index public biomedical data resources with ontology concepts [6].

The Annotator uses one of the largest available sets of biomedical ontologies including the NCBO BioPortal ontologies and the Unified Medical Language System (UMLS) Metathesaurus ontologies. The NCBO BioPortal [7] is a web repository of biomedical ontologies. Users can browse, search, and comment ontologies both online and via a web services application programming interface. The UMLS Metathesaurus [8] is a collection of concepts, terms and their relationships from various biomedical controlled vocabularies, terminologies and ontologies.

This study describes a use of the Annotator service to implement the Recommender service. We present recommendation heuristics for suggesting ontologies in corpus & keyword-based recommendation scenarios. They aim to address the following questions:

1. Which ontologies offer maximum coverage for a set of data?
2. Which ontologies are reference ontologies for a set of data?
3. Which small ontologies are specialized for a set of data?

We evaluate and discuss recommendation results generated by each heuristics in the context of three real world use cases. The best recommendation heuristics, which address both questions 1 and 2, were rated 'very relevant' and 'relevant' by expert evaluators.

An alpha version of the Recommender is publicly available and described at [www.bioontology.org/wiki/index.php/Ontology_Recommender_Web_service](www.bioontology.org/wiki/index.php/Ontology_Recommender_Web_service). It is deployed as a REpresentational State Transfer (REST) web service for programmatic access. It can also be accessed through a user interface. The Recommender service is currently being moved into a production environment and embedded in the BioPortal web application. Additional file 1 provides documentation describing how to use the service.

# Methods

Figure 1 describes the overall workflow of the Recommender. The service accepts biomedical text data as input and suggests the most appropriate ontologies. The annotations used to generate the recommendation are produced by the Annotator summarized in the next section. Next, according to the annotations, ontologies are scored and ranked with different output values. Results can be returned as text or XML.

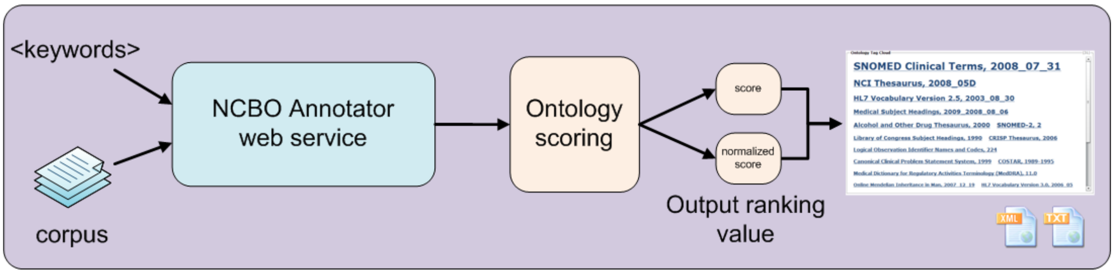


**Figure 1 – Recommender service workflow.**

### NCBO Annotator web service

The Annotator web service workflow is composed of two main steps (Figure 2) [5]. First, direct annotations are created from raw text. Annotations are based on syntactic concept recognition using a dictionary compiled from terms (concept names and synonyms) pulled from the ontologies. The Annotator enables the selection of ontologies from one of the largest sets of available biomedical ontologies. We implemented the service using the 98 English ontologies in UMLS 2008AA and a subset of the BioPortal ontologies (122 as of this writing). These ontologies provide a dictionary of 4,222,921 concepts and 7,943,757 terms. In the second step, semantic expansion components leverage the semantics in ontologies (e.g., *is_a* relations and mappings) to create additional annotations. For example, the *is_a transitive closure* component traverses an ontology parent-child hierarchy to create new annotations with parent concepts of concepts in direct annotations. The *ontology-mapping* component creates new annotations based on existing mappings between different ontologies. Point-to-point mappings that link concepts one another are defined manually or by automatic algorithms in the UMLS Metathesaurus and in BioPortal.

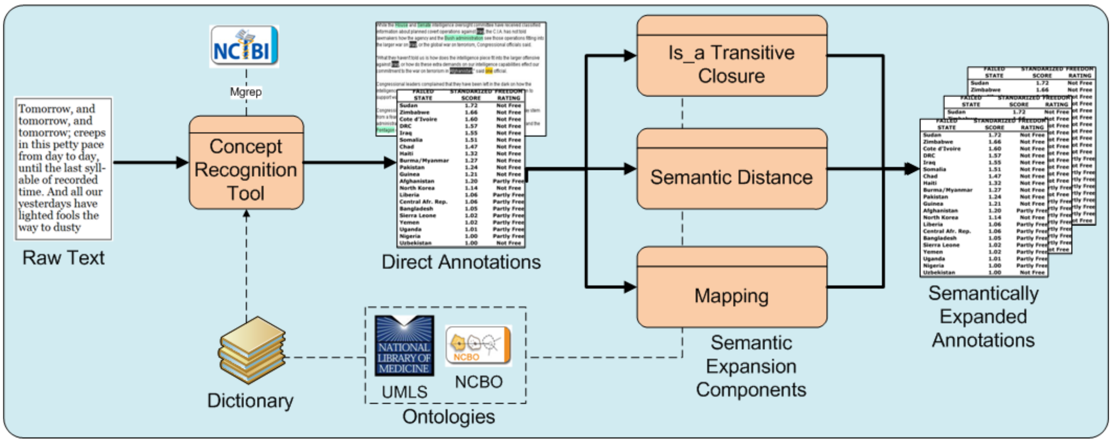


**Figure 2 – NCBO Annotator web service workflow.** *Direct annotations are created from raw text based on syntactic concept recognition (concepts names & synonyms). Next, different components expand the first set of annotations using the knowledge represented in one or more ontologies.*

We used the results of the Annotator to score ontologies. When using the Recommender, users can use the UMLS ontologies, the BioPortal ontologies, or both. By annotating user data with all available ontologies, we computed statistics and compared the ontologies to one another to recommend the most appropriate ones. We used the Annotator with two possible settings (see the appendix for exact parameters):

- Concept recognition only (CR);
- Concept recognition + mapping expansion (CR+M).

The concept recognition step (CR) allows the Recommender to evaluate an ontology following the *coverage* criterion. Ontologies containing more concepts mentioned by name or with synonyms in text data will create more annotations. Activating the mapping expansion (CR+M) has two interesting effects:

- *It extends coverage to terms defined in other ontologies*. For example, if the word 'treatment' is passed to the Annotator without mapping expansion, a direct annotation with the concept `MSH/C0087111` (*treatment* in MeSH) is identified, but no annotations are identified in SNOMED-CT. This is because the term 'treatment' does not exist in that ontology. However, with mapping expansion, a UMLS point-to-point mapping (based on CUI) `MSH/C0087111->SNOMEDCT/C0087111` can be used to generate an expanded annotation with concept `SNOMEDCT/C0087111` (*therapeutic procedure* in SNOMED-CT). Therefore, SNOMED-CT could also be considered as a potentially good ontology for text data containing the word 'treatment.'
- *It gives importance to reference ontologies i.e., ontologies that are good destinations for mappings (connectivity criterion).* The mapping to an ontology by many other ontologies shows its popularity and importance in the domain. For example, if the word 'melanoma' is passed to the Annotator without mapping expansion, two direct annotations are identified with `40644/Melanoma` (*melanoma* in NCI Thesaurus) and `40465/DOID:1909` (*melanoma* in Hunan Disease). However, with mapping expansion, an expanded annotation is also generated with `40644/Melanoma` using a mapping defined by a user in BioPortal `40465/DOID:1909->40644/Melanoma`. In this way, the Recommender gives the NCI Thesaurus more importance. In a previous study [9], we demonstrated the existence of hub ontologies in the

> network of biomedical ontologies. Hubs are connected one another by point-to-point mappings. For example: SNOMED-CT, the National Drug File Reference Terminology, MeSH, and the NCI Thesaurus are hubs. Our study demonstrated that 33% of ontologies have at least half of their concepts mapped to concepts in other ontologies.

Annotations are scored according to the context from which they were generated (direct concept recognition or semantic expansion) and returned to the user. The annotation scoring method is detailed in next section.

We did not consider *is_a* transitive closure expansion for recommendation because it gives more importance to ontologies with multiple inheritances. Examining ontology structure in order to discriminate ontologies has been suggested [10] and is discussed later in this paper.

The Annotator service provides a 'longest only' parameter to refine the matches to ontology concepts. If `longestOnly=true`, the Annotator selects only the longest term matching phrase. For example, if `longestOnly=true`, the phrase 'breast cancer' generates only ‘breast cancer.” If `longestOnly=false`, it generates three annotations: 'breast', 'cancer' and 'breast cancer.' The way the Annotator behaves with this parameter is even more useful for discriminating ontologies from one another. In fact, if it finds an annotation with the complete phrase in an ontology composing the dictionary, partial annotations with other ontologies will not be generated. For example, because 'breast cancer' exists in Human Disease and the NCI Thesaurus, if `longestOnly=true`, annotations generated with those terms will block annotations with the terms 'breast' in the Vaccine Ontology or 'cancer' in BIRNLex. This feature is interesting for the Recommender, as it allows enhancing the appropriateness of Human Disease and the NCI Thesaurus for the given phrase.

### Scoring method & output values

The score is a number assigned to an annotation to indicate its importance. Higher scores reflect more important annotations. The scoring algorithm gives a specific weight to an annotation according to its context, as well as matching terms. For instance, an annotation done by matching a concept’s preferred name gets a higher weight than one done by matching a concept’s synonym or one done with a parent-level-3 (ancestor) concept in the *is_a* hierarchy. In the previous example that considered the word 'treatment,' the Annotator would give more importance to ontologies containing the term 'treatment' than to ones containing the term 'therapeutic procedure.' Table 1 describes the weights used by the scoring algorithm.

### Table 1 – Annotation weights per context

| Annotation context | Weights |
|---|---|
| Direct annotation done with a concept preferred name | 10 |
| Direct annotation done with a concept synonym | 8 |
| Expanded annotation done with a mapping | 7 |
| Expanded annotation done with a parent level n<br>(e.g., 9 for n=1; 7 for n=2; 4 for n=5; 3 for n=8; 1 for n>12) | $1+10.e^{-0.2*n}$ |

The Recommender service uses the outputs of the Annotator to rank ontologies according to two output values:

1. *Score*: the sum of the scores of all the annotations generated with concepts from a particular ontology;
2. *Normalized score*: the score divided by the ontology size.

Each output ranking value is expected to provide different results in the scenarios considered. The score value is appropriate for the corpus-based recommendation, as it reflects the importance of terms appearing several times in the corpus. The normalized score is expected to help users distinguish between large ontologies that offer very good coverage of input data and small ontologies with correct coverage, yet more specialized to the input data's domain. Without assuming that small ontologies are better defined/formalized than larger ones, we assumed that this information could be of value to users.

Table 2 summarizes the questions we defined in the Introduction and the heuristics to address them.

### Table 2 – Recommender's heuristics and corresponding research questions.

| Annotator's method | Output value | Question |
|---|---|---|
| CR | score | Which ontologies offer maximum coverage for a set of data? |
| CR+M | score | Which ontologies are reference ontologies for a set of data? |
| CR | normalized-score | Which small ontologies are specialized for a set of data? |

#### Example

Consider the text: "*Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye.*" Sent to the Annotator, this sentence generates the following direct annotations—i.e. string matching with dictionary. (The numbers in the curly braces give the annotation weights):

- `NCI/C0025201`, *Melanocyte* in NCI Thesaurus {10}
- `NCI/C0025202`, *Melanoma* in NCI Thesaurus {10};
- `NCI/C0027651`, *Neoplasm* (synonym of *tumor*) in NCI Thesaurus {8};

- `FMA/C0015392`, *Eye* in FMA {10}
- `FMA/C0021853`, *Intestine* (synonym of *bowel*) in FMA {8}
- `40465/DOID:1909`, *Melanoma* in Human Disease {10};

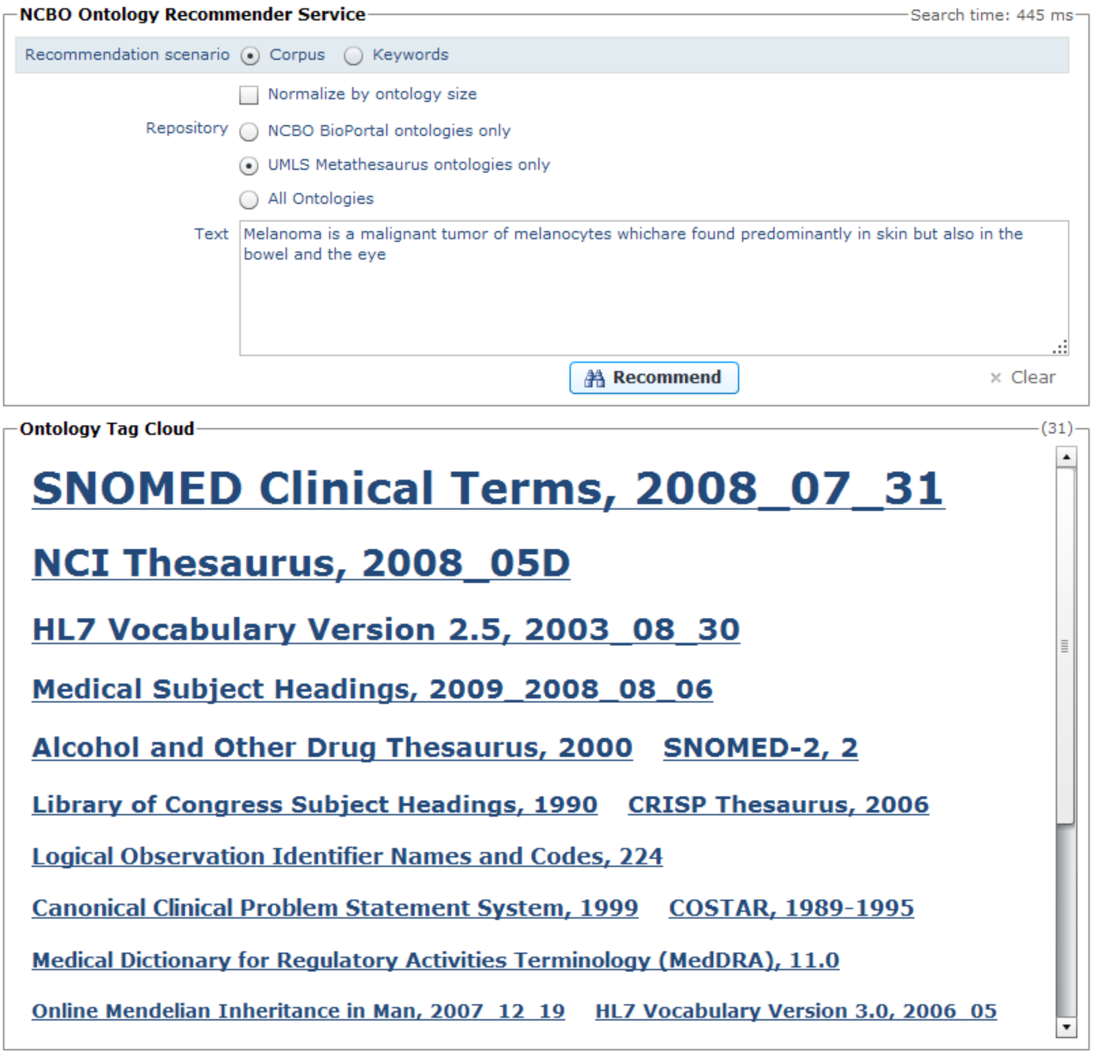
The mapping expansion generates the annotation (thanks to UMLS mappings):

- `FMA/C0025201`, *Melanocyte* in FMA, concept mapped to `NCI/C0025201` {7}.
- `NCI/C0015392`, *Eye* in NCI Thesaurus mapped to `FMA/C0015392` {7}.

The final score, using CR+M, is the sum of the annotations score per ontology:

- NCI Thesaurus (`NCI`): 35
- Foundational Model of Anatomy (`FMA`): 25
- Human Disease (`40465`): 10

Figure 3 shows results for this text in the Recommender service user interface (UMLS ontologies only). SNOMED-CT is the highest scored ontology.



**NCBO Ontology Recommender Service**

Search time: 445 ms

Recommendation scenario ● Corpus ○ Keywords

☐ Normalize by ontology size

Repository ○ NCBO BioPortal ontologies only
● UMLS Metathesaurus ontologies only
○ All Ontologies

Text: Melanoma is a malignant tumor of melanocytes whichare found predominantly in skin but also in the bowel and the eye

**Ontology Tag Cloud** (31)

SNOMED Clinical Terms, 2008_07_31
NCI Thesaurus, 2008_05D
HL7 Vocabulary Version 2.5, 2003_08_30
Medical Subject Headings, 2009_2008_08_06
Alcohol and Other Drug Thesaurus, 2000 SNOMED-2, 2
Library of Congress Subject Headings, 1990 CRISP Thesaurus, 2006
Logical Observation Identifier Names and Codes, 224
Canonical Clinical Problem Statement System, 1999 COSTAR, 1989-1995
Medical Dictionary for Regulatory Activities Terminology (MedDRA), 11.0
Online Mendelian Inheritance in Man, 2007_12_19 HL7 Vocabulary Version 3.0, 2006_05

**Figure 3 – Recommender web service user interface.** *A user can select the recommendation scenario, as well as the repository of ontologies to use, and enter the text data to recommend. A tag cloud is generated in which the score of an ontology is represented by the size of its name in the cloud.*

# Results & evaluation

In this section we present the results of a use case-based evaluation of the Recommender service. We asked three different groups to provide us with a corpus and a set of keywords representing the data they would like to use ontologies for. These evaluators were knowledgeable about biomedical ontologies and have already experienced the process of selecting the ontologies appropriate for their task. Thus, they were well-suited to evaluating the utility of the Recommender service for their datasets. Table 3 shows the source & size (number of words) of the datasets provided by each group. For each dataset, we ran the Recommender with the two methods (CR & CR+M) and generated rankings with two values (*score* and *normalized score*).

## Use cases

UC1: Researchers at the Center for Cardiovascular Bioinformatics and Modeling Johns Hopkins University developed an electrocardiogram (ECG) ontology. This ontology describes ECG data collection protocols, features of time-evolving ECG waveforms, ECG analysis algorithms, and data derived from ECG analyses. Its main role is to enable cardiovascular researchers to share and analyze primary and derived ECG annotated data. These researchers were interested in maximizing re-use of existing ontologies when building the ECG ontology.

UC2: Researchers at the European Bioinformatics Institute (EBI) developed an ontology focused on modeling experimental factors in the ArrayExpress database (http://www.ebi.ac.uk/microarray-as/ae/). The ontology was developed to "increase the richness of the annotations that are currently made in the ArrayExpress repository, to promote consistent annotation, to facilitate automatic annotation and to integrate external data." These researchers wished to map their new ontology to existing domain-specific ontologies.

UC3: Researchers at Stanford University are building a system that abstracts clinical information from two electronic medical record databases related to the care and management of breast cancer. Their goal is to assess quality of care and adherence to clinical guidelines as described in the National Comprehensive Cancer Network (http://www.nccn.org). These researchers wished to reuse ontologies that have been developed by other organizations to build their application.

## Table 3 – Source and size of the six datasets.

| **Dataset** | **Source** | **Size** |
|---|---|---|
| **UC1-keyword** | Provided by evaluator | 420 |
| **UC1-corpus** | Methods section of 3 papers about ECG-related paper | 2750 |
| **UC2-keyword** | Provided by evaluator | 9615 |
| **UC2-corpus** | Concatenated 'name', description' and 'species' sections of 30 randomly selected ArrayExpress entries | 6520 |
| **UC3-keyword** | Provided by evaluator | 72 |
| **UC3-corpus** | National Comprehensive Cancer Network (NCCN) Breast Cancer Guideline | 12540 |

### Results

To evaluate the recommendation produced for each dataset, we asked the evaluators for a reference ranking i.e., the ten ontologies that they would have liked to obtain. We also asked the evaluators to comment on our Recommender's ranking and to give it a mark between 1 and 5:

- *Very relevant* (5) – The recommendation exactly matched the results of the researcher's investigation, and the top-ranked ontologies were the ones appropriate for their data. The ranking is almost the one suggested by the researchers.
- *Relevant* (4) – The recommendation provided useful information. Most of the top ontologies were relevant. The ranking was fair.
- *Correct but not really relevant* (3) – The recommendation was technically correct but did not really help the researcher select the most appropriate ontologies. Some top ontologies were relevant others were not.
- *Few relevant* (2) – The recommendation was useless, or few relevant ontologies were identified.
- *Not relevant* (1) – The recommendation was wrong. The top-recommended ontologies were obviously not relevant.

Considering the large number of ontologies (220) that the Recommender uses and considering the presence of 5 duplicates (e.g., NCI Thesaurus being present both in UMLS (RRF format) and in BioPortal (OWL format)) we asked the evaluators to examine our top 15 results to evaluate a recommendation. Table 4 presents their evaluations.

**Table 4 – Evaluation of Recommender results.**

| Method | Output | UC1-keyword | UC1-corpus | UC2-keyword | UC2-corpus | UC3-keyword | UC3-corpus |
|---|---|---|---|---|---|---|---|
| **Concept recognition only (CR)** | Score | 5 | 4.5 | 4 | 3 | 5 | 5 |
| | Normalized score | 4.5 | 4.5 | 4 | 2 | 2 | 1 |
| **Concept recognition + mapping (CR+M)** | Score | 4 | 4 | 3 | 4 | 4.5 | 4.5 |
| | Normalized score | 3.5 | 4 | 2.5 | 2 | 1 | 1 |

### Results analysis

Overall, the evaluators were positive about the utility of the Recommender service. They all agreed that it would have helped them select ontologies for their task. On examining the results and their comments and marks, the following observations stand out:

- With the score output, large ontologies (17% of ontologies are above 20K concepts) were easily correctly identified. The high number of concepts in those ontologies makes them more appropriate for fully marking up or tagging textual descriptions. The large ontologies were usually among the top ten ranked ontologies. For example, SNOMED-CT (313K concepts & 972K terms), MeSH (291K concepts & 682K terms), Clinical Terms - Read Codes (186K concepts & 347K terms), and the NCI Thesaurus (74K concepts & 183K terms) were often in the results. We note that the NCI Thesaurus, which is not the largest ontology, was in the top 3 results for all the datasets.

- With the score output, moderately sized ontologies (36% of ontologies are between 1K & 20K concepts) and small ontologies (47% of ontologies are <1K concepts) were less often correctly identified. For example, the Recommender found Experimental Factor Ontology (2406 concepts & terms) in UC2-keyword or Human Developmental Anatomy (8340 concepts & terms) in UC1-keyword. But the Recommender missed some ontologies, regardless of the method used. For example, it never identified the Ontology for Biomedical Investigations in UC1 or RxNorm in UC3 whereas those ontologies were expected by the evaluators.
- Overall, the score was more informative than the normalized score. Normalization with ontology size placed importance on small ontologies containing a few terms in the dataset. Sometime, as in UC3, this feature was not considered relevant due to the introduction of considerable noise (e.g., Amino Acid (46 concepts & terms)). Sometime this feature is relevant, as in UC1, where it allows identifying 7 or 8 new ontologies missed by the score ranking such as Spatial Ontology (109 concepts & 168 terms), Ontology of Homology and Related Concepts in Biology (65 concepts & 132 terms). The Electrocardiography Ontology (497 concepts & terms) was also one of them. This result was particularly relevant as this is the ontology developed at Johns Hopkins University. In UC2, the Recommender correctly indentified Common Anatomy Reference Ontology (46 concepts & 50 terms) and Experimental Factor Ontology (2406 concepts & terms). This result was also particularly relevant as this ontology was specifically developed at EBI to annotate ArrayExpress experiments.
- The intersection of the top ten ontologies obtained with the score output and the top ten ontologies obtained with normalized-score output was small. We note that, the ontologies in this intersection are generally very relevant to the dataset; for example, Experimental Factor Ontology or Uber Anatomy Ontology in UC2 or Mass spectrometry in UC1.
- The normalized score is more informative for keyword-based recommendations than with corpus-based recommendation. This result is not surprising, because in a keyword-based recommendation, the score is supposed to be directly proportional to the number of unique concepts identified in the keywords. Therefore, ranking after normalizing the score is equivalent to ranking based on percentage of overlap between the keywords and the ontology.
- The influence of the mappings was noted by the evaluators as resulting in giving disproportionate importance to reference ontologies. Indeed, the use of the mappings will slightly change the ranking of the top ontologies that can help distinguishing between large ontologies (which are often actually the ones with many mappings). However, this was not the preferred ranking for most uses cases (except UC2-corpus). Furthermore, the use of mappings does not help in identification of average size and small ontologies.

- The influence of mappings was in fact detrimental in the keyword scenario. In this scenario, where a user provided the exact terms to map to ontology concepts, the activation of mappings introduced noise.
- The influence of mappings was not significant after results were normalized by ontology size.
- The size of the dataset influenced recommendation quality. Indeed, the smallest datasets received the lowest marks (e.g., UC1-keyword and UC3-keyword). This finding is expected because the more data the Recommender has, the better the recommendation will be.
- We note that when a significant and a large enough set of keywords is provided e.g., UC2, the recommendations generated based on keywords are preferred over the ones generated from corpus data. This is also an expected behavior because corpus data introduces spurious text phrases that bias the results. Especially, we note that the rankings based on the normalized score are significantly better with a large set of keywords.
- When used without the mappings expansion (CR), we obtained an average response time of 15 seconds for 1000 words with the service's current implementation. However, activating the mapping expansion (CR+M) slows its performance to ~1 hour for 1000 words.

The original datasets used as well as the recommendations generated are available at: http://www.bioontology.org/wiki/index.php/Ontology_Recommender_Web_service

# Related work

Much work has been done in the semantic web community on evaluating the quality of an ontology for a particular criteria or for a specific application. In this section we present a summary of the recent literature about ontology recommendation. We identified 20 tools or methods and compared them from high level functional perspective described below and summarized in Figure 4 (see additional file 2 for a detailed description of each tool and method). Only a few approaches (6) are applied to the biomedical domain.

## Recommendation criteria

We performed a functional comparison for the different tools or methods identified in the literature based on the following criteria:

- **Automation**: *Is the tool or the method fully automatic or does it requires an interaction with the user for whom a recommendation is necessary?* For examples, [30, 20, 31] are non-automatic method in which users are involved in the recommendation process. In OntoKhoj the user is requested for the disambiguation process.
- **Dynamic**: *Is the tool quick enough to be dynamically invoked by client applications and accurate enough to avoid requiring human intervention to clean the results?* For example, AKTiveRank can be dynamically invoked only if the ontology returned by the underlying search engine has already been processed once; otherwise the response time is strongly affected. Non-automatic approaches are excluded from this criterion. Part of this criterion is called 'run time performance' in [11].

- **Term matching**: *Is the tool or method based on any kind of matching between the query terms (directly submitted, or expanded, or from a corpus) and the class and property names of the ontologies?* This kind of matching could be exact match, or fuzzy match (e.g., contains, stemming) as in [3]. This criterion is called ‘class match measure’ in AKTiveRank or ‘coverage’ in OntoSelect or ‘topic coverage’ in [11].
- **Property matching**: *Does the tool or method exploits any kind of matching between the query terms and the property values of the classes e.g., definition, synonyms?* For instances, [24], BioPortal Search and OLS do matching using the concept names and synonyms as the Recommender does. Being restricted to a specific domain does facilitate such feature implementation as it is easier to specify which property values to look up into.
- **Query expansion**: *Does the tool performs any form of query expansion to retrieve a more representative set of terms to match with the ontologies?* For examples [20, 25, 3, 31] use WordNet to expand the query terms with hypernyms, hyponyms or synonyms. As another example, [24] expands the user query by automatically obtaining a corpus for the given the keywords via Google or Wikipedia.
- **Structure measure**: *Is the tool or method based on some formal measures of the ontology structure?* This criterion is similar to ‘richness of knowledge’ in [11] or ‘structure’ in OntoSelect. AKTiveRank proposes two of such measures: the ‘centrality measure’ based on the position of the matching concepts in the hierarchy (i.e., middle level concepts are given more importance; the ‘density measure’ based on the number of relation for a concept (i.e., concepts with high numbers of subclasses, superclasses and instances are given more importance). [23] and [28] also propose a granularity measure based on relation richness.
- **Connectivity**: *Does the tool or the method exploits the possible references (e.g., import, instantiation) or link (e.g., mappings) between ontologies in order to give more importance to reference ontologies?* This criterion is called ‘popularity’ in [11] or ‘connectedness measure’ in OntoSelect. For example the PageRank approach of Swoogle and OntoKhoj.
- **Disambiguation**: *Does the tool or the method performs any kind of disambiguation or semantic matching using the semantics of the ontology when doing the matching with terms?* For example, AKTiveRank measures the ‘semantic similarity’ between matching concepts and ontologies in which matching concepts are semantically close to one another are better ranked. OntoKhoj includes a disambiguation process involving the user.
- **Reasoning**. *Does the tool or method uses any kind of reasoning?*
- **Popularity**: *Does the tool or method uses any kind of users direct (e.g., reviews, notes) or indirect (e.g., usage logs) assessments to rank ontologies?* For examples, reviews & notes entered in BioPortal by users or assements used in WebCORE.

The notion of *coverage*, previously mentioned in the paper can now be described as the conjunction of the term matching and property matching.

### Summary of the limitations of current tools for application to biomedical ontologies

In their study Sabou et al. [11] identified two major shortcomings of ontology selection approaches:

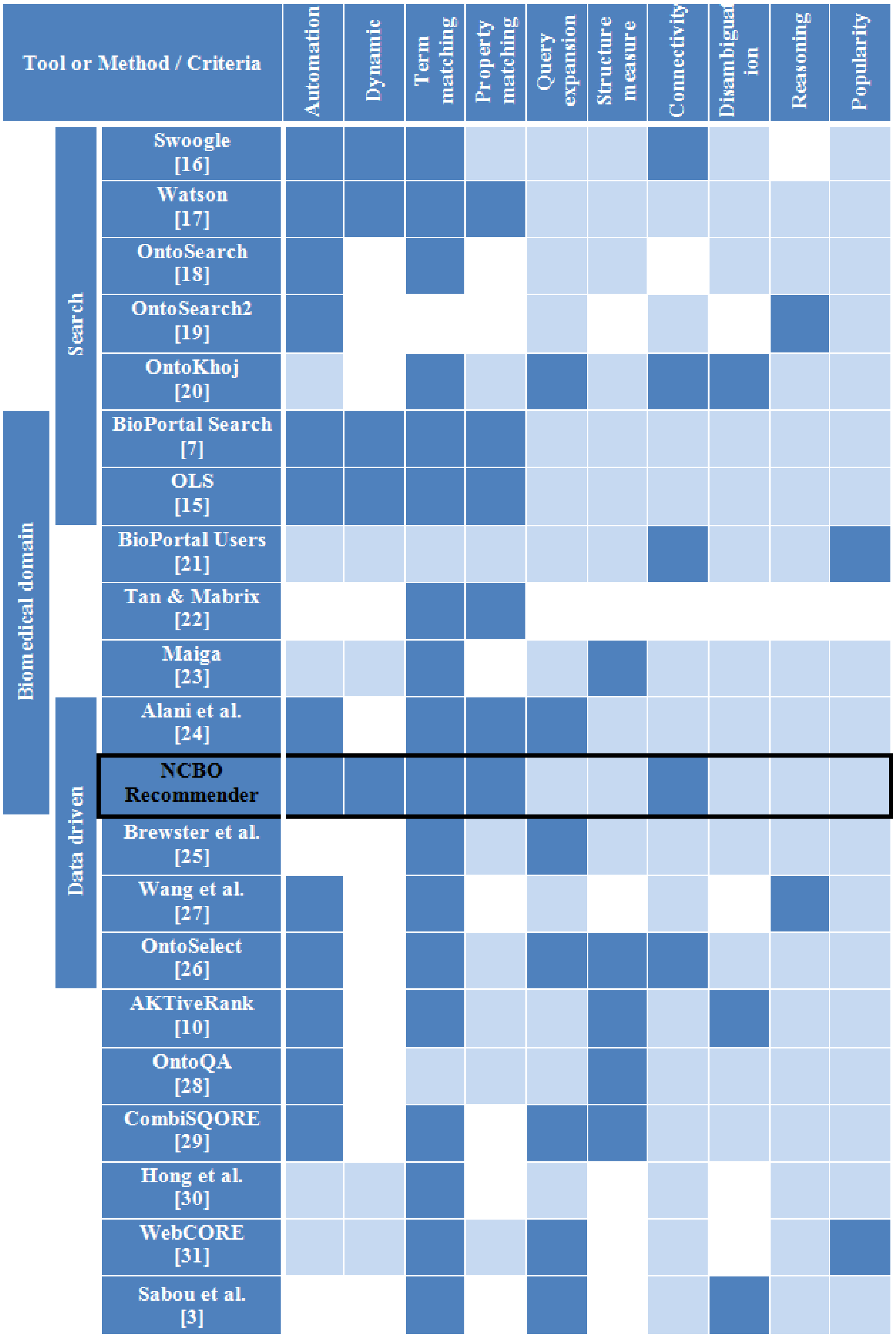
- Relations between concepts are ignored most of the time;
- The meaning of concepts is ignored most of the time.

In light of the comparison done in **Error! Reference source not found.**, the former shortcoming is not valid anymore as eight methods address the ‘structure measure’ or ‘connectivity’ criteria. However, the latter shortcoming still stands because ‘property matching’ or ‘disambiguation’ criteria are not well addressed; probably because correctly identifying ontology concepts in keywords or unstructured text is still a hard task. For example, if a user submits only the term ‘cold’ to an ontology selection system, it is impossible to figure out if the intent of the user is to get results for the disease or for the sensation. This is an issue that some methods propose to address either by interaction with the users or by having a corpus-based approach. The analysis of the literature explains the need of a new tool for biomedical ontology recommendation; particularly because:

- Many methods are not implemented into a concrete application or service that can actually be used by the biomedical community;
- The number of available ontologies is often limited;
- Few of the tools handle biomedical ontologies in non semantic web standard format (e.g., OBO, RRF);
- Despite the limited use of meaning of concepts, simple properties such as synonyms, which are often not available in a general context, are usually explicitly defined in biomedical ontologies;
- The used of mappings is missing. Resources like the UMLS Metathesaurus or NCBO BioPortal are now very rich in point-to-point mappings [9];
- Only one method suggests using the size of the ontology which according to the results presented before could sometime be relevant.

All these limitations are addressed by the Recommender in a simple web service application that can recommend from over 220 biomedical ontologies.

**Figure 4 – Comparison of ontology selection approaches.** *Bold blue filled cells mean a positive value for the given criterion whereas light blue cells mean a negative value. White cells are undetermined or not applicable.*



| Tool or Method / Criteria | | | Automation | Dynamic | Term matching | Property matching | Query expansion | Structure measure | Connectivity | Disambiguation | Reasoning | Popularity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Search | Swoogle [16] | ■ | ■ | ■ | □ | □ | □ | ■ | □ | | □ |
| | | Watson [17] | ■ | ■ | ■ | ■ | □ | □ | □ | □ | □ | □ |
| | | OntoSearch [18] | ■ | | ■ | | □ | □ | | □ | □ | □ |
| | | OntoSearch2 [19] | ■ | | | | □ | | □ | | ■ | □ |
| | | OntoKhoj [20] | □ | | ■ | □ | ■ | □ | ■ | ■ | □ | □ |
| Biomedical domain | | BioPortal Search [7] | ■ | ■ | ■ | ■ | □ | □ | □ | □ | □ | □ |
| | | OLS [15] | ■ | ■ | ■ | ■ | □ | □ | □ | □ | □ | □ |
| | | BioPortal Users [21] | □ | □ | □ | □ | □ | □ | ■ | □ | □ | ■ |
| | | Tan & Mabrix [22] | | | ■ | ■ | | | | | | |
| | | Maiga [23] | □ | □ | ■ | | □ | ■ | □ | □ | □ | □ |
| | Data driven | Alani et al. [24] | ■ | | ■ | ■ | ■ | □ | □ | □ | □ | □ |
| | | **NCBO Recommender** | ■ | ■ | ■ | ■ | □ | □ | ■ | □ | □ | □ |
| | | Brewster et al. [25] | | | ■ | □ | ■ | □ | □ | □ | □ | □ |
| | | Wang et al. [27] | ■ | | ■ | | □ | | □ | | ■ | □ |
| | | OntoSelect [26] | ■ | | ■ | □ | ■ | ■ | ■ | □ | □ | □ |
| | | AKTiveRank [10] | ■ | | ■ | □ | □ | ■ | □ | ■ | □ | □ |
| | | OntoQA [28] | ■ | | □ | □ | □ | ■ | □ | □ | □ | □ |
| | | CombiSQORE [29] | ■ | | ■ | | ■ | ■ | □ | □ | □ | □ |
| | | Hong et al. [30] | □ | □ | ■ | | □ | | □ | | □ | □ |
| | | WebCORE [31] | □ | □ | ■ | □ | ■ | | □ | | □ | ■ |
| | | Sabou et al. [3] | | | ■ | | ■ | | □ | ■ | □ | □ |

## Discussion & future work

The task of recommending an ontology is hard because of the variety of user requirements and expectations: “Good ontologies are the ones that serve their purpose” [25]. The perfect recommendation method, automatic, easy-to-use, and completely accurate does not probably exist. The related work illustrates the importance of being able to address the two scenarios of corpus & keyword based recommendation. The analysis of the results demonstrates that the Recommender successfully addresses both scenarios. Human intervention is needed to clean the noise introduced in the recommendation when normalizing by ontology size (i.e., third question). Moreover, the response times obtained with the CM method are good enough to envision embedding the Recommender into software application e.g., semantic browsing. Table 5 summarizes questions addressed by the Recommender.

**Table 5 – Addressing of each questions (automation, speed & accuracy).**

| Question – Recommender's method | Automation | Fast enough | Accuracy |
|---|---|---|---|
| Which ontologies offer the maximum coverage for my data? – (CR – score) | Yes | Yes | Yes |
| Which ontologies are reference ontologies for my data? – (CR+M – score) | Yes | No | Yes |
| Which small ontologies are specialized for my data? – (CR – normalized score) | Yes | Yes | Not enough |

When analyzing the results, one can say that normalizing by the size is not enough to identify specialized ontologies. However, discriminating ontologies based on the ontology structure (as suggested in related work) is hard and subjective i.e., how to decide to give more importance to formalism rather than to another one. This is the reason why we do not use the Annotator is_a transitive closure expansion.

Also, the mapping expansion appeared useful but its influence was offset by coverage. In the future, we would like the Recommender to abstract on service configurations (e.g., context weights, parameters, criteria) in order to enable further control on the scoring routine. Each user can therefore select the 'scenario' that matches the best his needs. For instance, a user with a small set of keywords would prefer to use the corpus-based recommendation; whereas a user with a large set of keywords would rather go for the keyword recommendation scenario.

We also note that the results of the Recommender are dependent on the accuracy of the NCBO Annotator, which uses lexical matching for concept recognition and the limitations that go with it [5]. Matching text to ontology concepts is a hard task. A major drawback of ontology selection approaches is to be based on some kind on lexical matching between concepts and keywords. The matching methods do not take into consideration the semantics of the ontologies when doing the look up. The Recommender service provides two useful features to partially address this issue:

- The services uses synonyms, because synonyms are the first step to accomplish a semantic match.
- The service uses mappings between ontologies to leverage semantics that has been implied by connecting ontologies together.

We note that the recommendation could be greatly enhanced if the annotations could be scored according to a degree of specificity of the concept forming the annotation. Indeed, concepts like 'cancer', 'cell' or 'disease' are less specific than 'pheochromocytoma' or 'appendicectomy'. One might want to see the ontologies that contain the specific terms, ranked higher than the ones containing the less specific ones. Evaluating concept specificity is indeed a challenging task. Within the context of BioPortal, we are investigating three different approaches: (i) based on the usage (e.g., to mine the user logs); (ii) based on the usefulness (annotated data); (iii) based on the ontology structure (e.g., to use the neighborhood/relationships of a concept to determine a degree of importance).

Furthermore, the recommendations are often linked to the size of ontologies. Large ontologies are easier to identify than small ones. The ontology recommendation challenge has to reconcile two conflicting effects related to the size of ontologies: On one hand, large ontologies have the advantage of a large coverage that allows good reusability and data integration. However, these ontologies are hard to manipulate as they introduce noise (e.g., numbers or country in SNOMED-CT) and are sometime hard to process (e.g., memory loading, reasoning, search). On the other hand, small ontologies are easily usable and processable and are adequate for precise tasks. However, these ontologies are sometime too specialized in order to be re-usable and lack links to other ones that will facilitate data integration. The result is that the most often users want to deal with a small piece of a large ontology i.e., an ontology view (or module). This is, indeed, the hardest thing to do. To the best of our knowledge there is no tool or method that has addressed the question of recommending only an ontology view or a subpart of an ontology. This challenge is also identified in Sabou et al. in [11, 3]. With the introduction of some type of ontology views (e.g., hierarchy branch) in BioPortal [21], it would be interesting to extend the Recommender service with the possibility to recommend views.

Finally, the evaluators mentioned the requirement to use the Recommender with a customized list of ontologies. It appears that in some cases, users already know the ontologies to use in their task. In those cases, users are not interested in the “recommendation/selection” aspect but in the “evaluation” aspect provided by the Recommender (cf additional file 2). Undeniably, the noise introduced in the recommendations by the other ontologies makes the evaluation process a bit harder. We are currently implementing a way to filter the recommendations per ontology in order to address those uses cases.

# Conclusions

Biomedical ontologies have been identified as a crucial means for representing knowledge and annotate biomedical data in order to create a biomedical semantic web. Ontologies facilitate data integration and translational discoveries[6]. In this paper, we have discussed the need of ontology recommendation in order to design new ontologies and annotate data. We have presented the *Biomedical Ontology Recommender web service*, which – given textual metadata (corpus or keyword describing the domain of interest) – suggests appropriate ontologies relevant for annotating the given data. Our approach uses both a syntactic concept recognition step (string matching with concept names & synonyms) and a mapping expansion step to enforce reference ontologies (expand annotations with point-to-point mappings). The Recommender uses over 220 ontologies from the UMLS Metathesaurus and the NCBO BioPortal, which is to the best of our knowledge, unique in biomedicine. We have demonstrated the Recommender’s performance on several recommendation heuristics via a use case-based evaluation. Overall, evaluators agreed on the utility of the recommendations provided both for their keyword and corpus datasets. We have compared and contrasted our approach with other methods and tools that have been published in the literature. We have also identified the key outstanding issues in ontology recommendation that point to fruitful research directions; such as: (i) selecting small and specialized ontologies; (ii) semantically matching free text and ontology concepts; (iii) recommending according to a specific desired criterion; (iv) recommending ontology views. In the future, we expect further work on the Recommender service to address those issues.

## Competing interests

None

## Authors' contributions

CJ designed the Recommender workflow, implemented the web service and wrote the manuscript. NHS conceived of the project, participated in technical discussions as well as system design, and edited the manuscript. MAM supervised and mentored CJ and NHS.

## Acknowledgements

This work is supported by NIH grant U54 HG004028 in support of the National Center for Biomedical Ontology. We acknowledge the assistance in software development of Chris Callendar (University of Victoria, CA) for the user interface. We also thank Helen Parkinson (European Biomedical Institute), Stephen Granite (Johns Hopkins University) and Wei-Nchih Lee (Stanford University) for the help in the Recommender evaluation.

## Appendix

### Parameters to give to the NCBO Annotator web service

The following NCBO Annotator parameters are used to implement the Recommender service (non specified parameters are set to default values):

| | | |
|---|---|---|
| `longestOnly` | `= true` | [for keyword-based recommendation] |
| | `\| false` | [for corpus-based recommendation] |
| `withDefaultStopWords` | `= true` | |
| `minTermSize` | `= 3` | |
| `localSemanticTypeIDs` | `= T000` | [for UMLS Metathesaurus repository] |
| | `\| T999` | [for NCBO BioPortal repository] |
| | `\| nothing` | [for both repositories] |
| `activateMapping` | `= true` | [for CR+M method] |
| | `\| false` | [for CR method] |

## Additional Files

File name: Recommender web service documentation.doc

File format: Microsoft Word 97-2003

Title of data: Recommender web service documentation

Description of data: Contains the detailed documentation on how to use the recommender web service.

File name: Detailed comparison of existing recommendation tools.doc

File Format: Microsoft Word 97-2003

Title of data: Detailed comparison of existing recommendation tools

Description of data: Contains a detailed comparison of the existing ontology recommendation tools that form the basis of the related work section.

## References

[1] Butte AJ, Chen R: **Finding disease-related genomic experiments within an international repository: first steps in translational bioinformatics**. In *American Medical Informatics Association Annual Symposium, AMIA'06*, Washington DC, USA 2006:106–110.

[2] Bodenreider O, Stevens R: **Bio-ontologies: Current Trends and Future Directions**. *Briefings in Bioinformatics* 2006, **7**(3):256–274.

[3] Sabou M, Lopez V, Motta E: **Ontology Selection on the Real Semantic Web: How to Cover the Queens Birthday Dinner?** In *15th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks, EKAW'06, Volume 4248 of* Lecture Notes in Computer Science. Edited by Staab S, Svátek V, Podebrady, Czech Republic: Springer 2006:96–111.

[4] Pafilis E, O'Donoghue SI, Jensen LJ, Horn H, Kuhn M, Brown NP, Schneider R: **Reflect: augmented browsing for the life scientist**. *Nature Biotechnology* 2009, **27**:508–510.

[5] Jonquet C, Shah NH, Musen MA: **The Open Biomedical Annotator**. In *AMIA Summit on Translational Bioinformatics*, San Francisco, CA, USA 2009:56–60.

[6] Shah NH, Jonquet C, Chiang AP, Butte AJ, Chen R, Musen MA: **Ontology-driven Indexing of Public Datasets for Translational Bioinformatics**. *BMC Bioinformatics* 2009, **10**.

[7] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith NB, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA: **BioPortal: ontologies and integrated data resources at the click of a mouse**. *Nucleic Acids Research* 2009, :4. [Advance Access published online].

[8] Bodenreider O: **The Unified Medical Language System (UMLS): integrating biomedical terminology**. *Nucleic Acids Research* 2004, **32**:267–270.

[9] Ghazvinian A, Noy N, Jonquet C, Shah N, Musen M: **What Four Million Mappings Can Tell You about Two Hundred Ontologies**. In *8th International Semantic Web Conference, ISWC'09, Volume 5823 of* Lecture Notes in Computer Science. Edited by Bernstein A, Karger DR, Heath T, Feigenbaum L, Maynard D, Motta E, Thirunarayan K, Washington DC, USA: Springer 2009:229–242.

[10] Alani H, Brewster C: **Ontology Ranking Based on the Analysis of Concept Structures**. In *3rd International Conference on Knowledge Capture, K-Cap'05*, Banff, AB, Canada: ACM 2005:51–58.

[11] Sabou M, Lopez V, Motta E, Uren V: **Ontology Selection: Ontology Evaluation on the Real Semantic Web**. In *4th International EON Workshop, Evaluation of Ontologies for the Web, EON'06, Volume 179 of* CEUR Workshop Proceedings. Edited by Vrandecic D, Suarez-Figueroa MC, Gangemi A, Sure Y, Edinburgh, Scotland, UK 2006.

[12] Lozano-Tello A, Gomez-Perez A: **ONTOMETRIC: A Method to Choose the Appropriate Ontology**. *Database Management* 2004, **15**(2):1–18.

[13] Brank J, Grobelnik M, Mladenic D: **A survey of ontology evaluation techniques**. In *Conference on Data Mining and Data Warehouses. SiKDD'05*, Ljubljana, Slovenia 2005:4.

[14] Gangemi A, Catenacci C, Ciaramita M, Lehmann J: **Modelling Ontology Evaluation and Validation**. In *3rd European Semantic Web Conference, ESWC'06, Volume 4011 of* Lecture Notes in Computer Science. Edited by Sure Y, Domingue J, Budva, Montenegro: Springer 2006:140–154.

[15] Côté RG, Jones P, Apweiler R, Hermjakob H: **The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries**. *BMC Bioinformatics* 2006, **7**(97):7.

[16] Ding L, Finin T, Joshi A, Peng Y, Cost RS, Sachs J, Pan R, Reddivari P, Doshi V: **Swoogle: A Semantic Web Search and Metadata Engine**. In *13th ACM Conference on Information and Knowledge Management, CIKM'04*. Edited by Grossman DA, Gravano L, Zhai C, Herzog O, Evans D, Washington DC, USA: ACM 2004.

[17] D'Aquin M, Baldassarre C, Gridinoc L, Angeletou S, Sabou M, Motta E: **Watson: A Gateway for Next Generation Semantic Web Applications**. In *Poster & Demonstration Session at the 6th International Semantic Web Conference, ISWC'07*, Busan, Korea 2007:3.

[18] Zhang Y, Vasconcelos W, Sleeman D: **OntoSearch: An Ontology Search Engine**. In *24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, AI'04*. Edited by Bramer M, Coenen F, Allen T, Cambridge, UK: Springer 2004:58–69.

[19] ZPan J, Thomas EJ, Sleeman DH: **OntoSearch2: Searching and Querying Web Ontologies**. In *IADIS International Conference WWW/Internet*, Murcia, Spain 2006:211–219.

[20] Patel C, Supekar K, Lee Y, Park EK: **OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification**. In *5th ACM International Workshop on Web Information and Data Management, WIDM'03*, New Orleans, LA, USA: ACM 2003:58–61.

[21] Noy NF, Dorf M, Griffith NB, Nyulas C, Musen MA: **Harnessing the Power of the Community in a Library of Biomedical Ontologies**. In *Workshop on Semantic Web Applications in Scientific Discourse, SWASD'09, Volume 523 of* CEUR Workshop Proceedings. Edited by Clark T, Luciano JS, Marshall MS, Prud'hommeaux E, Stephens S, Washington DC, USA 2009:11.

[22] Tan H, Lambrix P: **Selecting an Ontology for Biomedical Text Mining**. In *Human Language Technology Conference, BioNLP Workshop*, Boulder, CO, USA: Association for Computational Linguistics 2009:55–62.

[23] Maiga G: **A Flexible Biomedical Ontology Selection Tool**. In *Strengthening the Role of ICT in Development, Volume 5*. Edited by Kizza JM, Lynch K, Nath R, Aisbett J, Vir P, Fountain Publishers 2009:171–189. [Proceedings of 5th International Conference on Computing and ICT Research, ICCIR'09].

[24] Alani H, Noy NF, Shah N, Shadbolt N, Musen MA: **Searching Ontologies Based on Content: Experiments in the Biomedical Domain**. In *4th International Conference on Knowledge Capture, K-Cap'07*, Whistler, BC, Canada: ACM 2007:55–62.

[25] Brewster C, Alani H, Dasmahapatra S, Wilks Y: **Data Driven Ontology Evaluation**. In *4th International Conference on Language Resources and Evaluation, LREC'04*, Lisbon, Portugal 2004:4.

[26] Buitelaar P, Eigner T, Declerck T: **OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection**. In *Demonstration Session at the 3rd International Semantic Web Conference, ISWC'04*, Hiroshima, Japan 2004.

[27] Wang X, Guo L, Fang J: **Automated ontology selection based on description logic**. In *12th International Conference on Computer Supported Cooperative Work in Design, CSCWD'08*, Xi'an, China: IEEE 2008:482–487.

[28] Tartir S, Arpinar IB, Moore M, Sheth AP, Aleman-Meza B: **OntoQA: Metric-Based Ontology Quality Analysis**. In *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, Houston, TX, USA 2005:9.

[29] Ungrangsi R, Anutariya C, Wuwongse V: **CombiSQORE: a combinative-ontology retrieval system for next generation semantic web applications**. *IEICE Transactions on Information and Systems* 2008, **E91-D**(11):2616–2625.

[30] Hong TP, Chang WC, Lin JH: **A Two-Phased Ontology Selection Approach for Semantic Web**. In *9th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, KES'05*. Edited by RKhosla, RJHowlett, LCJain v 2005.

[31] Cantador I, Fernandez M, Castells P: **Improving Ontology Recommendation and Reuse in WebCORE by Collaborative Assessments**. In *Workshop on Social and Collaborative Construction of Structured Knowledge, CKC'07, Volume 273 of* CEUR Workshop Proceedings. Edited by Noy N, Alani H, Stumme G, Mika P, Sure Y, Vrandecic D, Banff, Canada 2007.

[32] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Consortium TO, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah NH, Whetzel PL, Lewis S: **The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration**. *Nature Biotechnology* 2007, **25**(11):1251–1255.

[33] Rosch E: *Principles of Categorization*, New Jersey, USA: Lawrence Erlbaum 1978 :27–48.

# Figures

<Figures to be moved here for final version>

# Tables

<Tables to be moved here for final version>

## Recommender web service documentation

This section describes how to use the Recommender service. An *alpha* version of the Recommender service is deployed as a REpresentational State Transfer (REST) web service and return results in several formats as text or XML. URLs given in this section are subject to change as the web service is currently being deployed in production environment.

### REST Application Programming Interface (API)

In order to use the Recommender web service REST API a user needs to implement a client application that will send an HTTP POST request to the Recommender base service URL:

`http://ncbolabs-dev2.stanford.edu:8080/OBS_v1/recommender1.1/`

A simple HTML client to test the Recommender is available here:

http://obs.bioontology.org/recommender/Recommender1.1.html

The following parameters are available for the API:

- `method` – Identifies the method used for recommendation. Each method will request the NCBO Annotator with specific parameters in order to achieve the Recommender heuristics. 4 methods are available:
    - `1` – CR with `longestOnly=false`
    - `2` – CR+M with `longestOnly=true`
    - `3` – CR with `longestOnly=true`
    - `4` – CR+M with `longestOnly=true`
- `output` – Identifies the output ranking value considered for ranking ontologies in the results. 4 outputs values are available:
    - `score` – The sum of the scores of all the annotations generated with concepts from a particular ontology.
    - `nb-annotating-concepts` – The number of unique concepts annotating the input text for a particular ontology.
    - `normalized-score` – The ratio between the ontology score and the ontology size.
    - `overlap` – The ratio between the number of annotating concepts and the ontology size.
- `repository` – The set of ontologies to consider for recommendation:
    - `ncbo` – Only the NCBO BioPortal ontologies.
    - `umls` – Only the UMLS Metathesaurus ontologies.
    - `all` – Both NCBO & UMLS ontologies.
- `text` – The text data used for recommendation.
- `format` – The output format of the results. Ontologies are sorted according to the `output` parameter. 3 formats are available:
    - `asText` – Returns the results in text document.
    - `asTextSimple` – Returns the results in a simplified text document without the output ranking values.

    - `asXML` – Returns the result in a XML document. Each ontology is represented by a `<obs.common.beans.RecommendedOntologyBean>` for example:

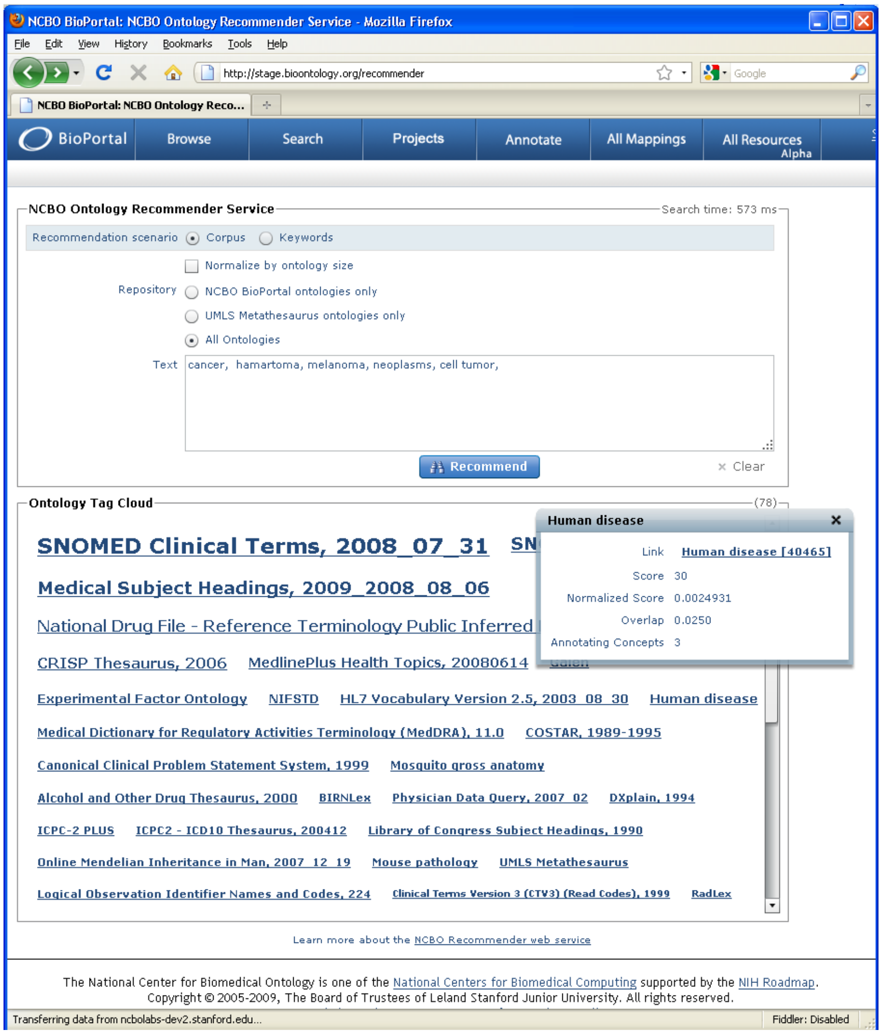
```
<obs.common.beans.RecommendedOntologyBean>
      <localOntologyID>40644</localOntologyID>
      <ontologyName>NCI Thesaurus</ontologyName>
      <ontologyVersion>09.07</ontologyVersion>
      <virtualOntologyID>1032</virtualOntologyID>
      <nbAnnotation>5</nbAnnotation>
      <score>42.0</score>
      <nbAnnotatingConcept>4</nbAnnotatingConcept>
      <normalizedScore>5.656794E-4</normalizedScore>
      <overlap>0.0050</overlap>
</obs.common.beans.RecommendedOntologyBean>
```

Remark: The number of annotations for a particular ontology is also generated however this value cannot be used for ranking.

### User interface

The Recommender is currently being embedded in BioPortal (Figure 1). A simplified user interface allows a user to test some of the Recommender heuristics in a web browser. The two discussed scenarios are available in the user interface (corpus & keyword-based recommendation). An additional parameter allows normalizing the ranking result by ontology size. The `methods` considered by the user interface are always `1` for corpus-based recommendation and `3` for keyword-based recommendation. The `outputs` considered are `score` and `normalized-score`. Therefore, the user interface addresses the first and third question indentified in Introduction. The user interface allows also selecting the `repository` to use.

As illustrated in Figure 1, a tag cloud is generated in which the score (or normalized-score) of an ontology is represented by the size of its name in the cloud. When clicking on an ontology name in the cloud, a pop-up window provides the values of each output ranking value for this ontology (Figure 1). A link to the original ontology in BioPortal is also provided when it is appropriate.



**NCBO Ontology Recommender Service** Search time: 573 ms

Recommendation scenario ● Corpus ○ Keywords

☐ Normalize by ontology size

Repository ○ NCBO BioPortal ontologies only

○ UMLS Metathesaurus ontologies only

● All Ontologies

Text: cancer, hamartoma, melanoma, neoplasms, cell tumor,

**Ontology Tag Cloud** (78)

SNOMED Clinical Terms, 2008_07_31   SN

Medical Subject Headings, 2009_2008_08_06

National Drug File - Reference Terminology Public Inferred

CRISP Thesaurus, 2006   MedlinePlus Health Topics, 20080614   Galen

Experimental Factor Ontology   NIFSTD   HL7 Vocabulary Version 2.5, 2003_08_30   Human disease

Medical Dictionary for Regulatory Activities Terminology (MedDRA), 11.0   COSTAR, 1989-1995

Canonical Clinical Problem Statement System, 1999   Mosquito gross anatomy

Alcohol and Other Drug Thesaurus, 2000   BIRNLex   Physician Data Query, 2007_02   DXplain, 1994

ICPC-2 PLUS   ICPC2 - ICD10 Thesaurus, 200412   Library of Congress Subject Headings, 1990

Online Mendelian Inheritance in Man, 2007_12_19   Mouse pathology   UMLS Metathesaurus

Logical Observation Identifier Names and Codes, 224   Clinical Terms Version 3 (CTV3) (Read Codes), 1999   RadLex

**Human disease**

| | |
|---|---|
| Link | Human disease [40465] |
| Score | 30 |
| Normalized Score | 0.0024931 |
| Overlap | 0.0250 |
| Annotating Concepts | 3 |

Learn more about the NCBO Recommender web service

The National Center for Biomedical Ontology is one of the National Centers for Biomedical Computing supported by the NIH Roadmap.
Copyright © 2005-2009, The Board of Trustees of Leland Stanford Junior University. All rights reserved.

**Figure 1 – Recommender web service user interface in BioPortal.**

## Detailed comparison of existing recommendation tools

Much work has been done in the semantic web community on evaluating the quality of an ontology for a particular criteria or for a specific application. Existing efforts make a distinction between ontology *recommendation* (also called ontology *selection* or ontology *ranking* in the literature) and ontology *evaluation* approaches. We first explain the distinction, and then present an overview of the recent literature on ontology recommendation. Only a few approaches are applied to the biomedical domain.

### Ontology evaluation vs. Ontology selection

Sabou et al. [11] define ontology evaluation and ontology selection as two different things that closely related one another: *Ontology selection is the process that allows identifying one or more ontologies or ontology modules that satisfy certain criteria. The actual process of checking whether an ontology satisfies certain criteria is, an ontology evaluation task.* Indeed, ontology evaluation consists in evaluating an ontology, independently from other ones, following a specific criterion. There are different approaches to evaluate an ontology. Some evaluate the logical correctness of the ontology, some compare metrics, some compare technologies, some compare structure & representation, some evaluate the appropriateness of an ontology for a specific tool or task, some compare ontologies against common benchmarks, some evaluate the appropriateness of an ontology against a corpus of data, some evaluate according to user assessments. The reader might refer to [12, 13, 14].

Sabou et al. [11] illustrate an interesting paradox: few of the existing ontology evaluation techniques are employed by the selection algorithms described after. Most of the approaches for ontology selection are based on some kind of matching (syntactic or semantic) between the query terms (or the processed corpus of data) and the ontology content. Then some measures are applied to the identified concepts and the aggregation of these measures for each concept per ontology usually gives the ranking. The Recommender follows the same general principle. As a consequence, in the following, we only detail ontology evaluation approaches if they have been explicitly used in a tool or method for ontology selection. We review the possible evaluation criteria (actually used) and the properties of the ontology selection approaches that we have identified in the literature.

### Search ontologies approaches

The expression ‘ontology search’ is also sometime used in the literature but is ambiguous as it can have two different meanings: (i) *Search for ontology* which is the process, similar to ontology selection, of finding ontologies corresponding to a given query with or without ranking the results; (ii) *Search in ontology* which is the process of finding concepts or other ontological entity that forms an ontology (e.g., property) across ontologies. Often, the former is based on a method that uses the latter.

Search in ontologies is a feature often present in ontology repository or on the web. BioPortal Search [7] and the Ontology Lookup Service (OLS) [15] are specifically restricted to biomedical ontologies. They are described in the next section. Swoogle [16], Watson, [17], OntoSearch [18], OntoSearch2 [19] and OntoKhoj [20] are tools to search in ontologies. Those tools usually crawl the semantic web documents from the web and index them. Then they allow a keyword-based search of their index and they usually return only the ontologies that have the keyword in their class or property names (Watson does actually use also property values). Watson & OntoSearch2 propose a more formal query approach by allowing users to pass SPARQL queries. Swoogle & OntoKhoj use ranking algorithms similar to page rank used by Google to give more importance to ontologies that are highly connected (references to and from). For example, OntoKhoj uses instantiation, sub classes and domain/range links. However, neither OntoKhoj nor Swoogle do use mappings between ontologies which in the biomedical domain are the most frequent way of connecting ontologies (note also that neither BioPortal Search nor OLS use mappings). The Recommender service can use these mappings to discriminate ontologies as previously explained.

Search in ontologies approaches alone are not sufficient for ontology recommendation as they do not provide the final abstraction to say to a user "these are the ontologies you should use" for your data. Those approaches fail addressing the scenarios identified previously (i.e., they cannot deal directly with a large set of keywords or corpus datasets). However, those methods can be used to implement ontology selection workflow as in AKTiveRank [10] detailed later.

### Approaches applied to biomedicine

Biomedical ontology repositories like the NCBO BioPortal [7] and the EBI Ontology Lookup Service [15] usually proposes a search functionality that allows a user to query the repository with keywords. Concept names and synonyms are indexed (e.g., using Lucene) to benefits from the classic search features e.g., exact match or contains. BioPortal proposes also to search in all concepts properties values allowing for example to search with different type of concept identifiers. None of the two mentioned search engines rank their results (as Swoogle and OntoKhoj). Those search in ontologies approaches, even if specialised for biomedical ontologies, failed addressing the need of recommendation previously explained.

In the context of BioPortal, NCBO is experimenting a community-based evaluation approach [21]. BioPortal does provide the infrastructure to enable users to give information such as reviews of ontologies for specific projects or notes on ontology concepts or view on ontologies. The underlying questions that the community assessments help to resolve are for examples: Has anyone used this ontology successfully for a task that is similar to mine? Is this ontology currently active (i.e., discussed, used, and maintained)? The Recommender service complements this informal community-based approach with an automatic approach that asserts coverage and connectivity. In addition, we are currently working on implementing different recommendation heuristics that will leverage BioPortal direct and indirect user assessments.

Quite recently, Tan & Mabrix [22] describes a list of criteria to consider when selecting an ontology for a biomedical text mining task. The criteria are for example: ontology type (e.g., top-level ontology, domain ontology) supporting technologies (e.g., ontology format, tools available), types of representation accessible in the ontology (e.g., concepts, relations, instances), coverage of the domain, and evaluation in a real system or against a benchmark (e.g., BioNLP, BioCreative). The high level criteria elicited by Tan & Mabrix are indeed useful, and one should say that users have to be aware of those criteria when looking for an ontology for their needs. However, the proposed framework does not provide an automated way to abstract all these criteria and recommend an appropriate ontology. A user has to manually or semi-manually evaluate all of them to decide which ontology to (re)use.

Also in the biomedical domain, Maiga [23] proposes a tool for biomedical selection. The semi-automatic tool allows a user to specify task requirements as a query that will be matched against a set of summarized ontologies. Then measures (as described in two sections) based on the coverage between query and ontology content as well as based on the ratio and density of relations in the ontologies are used to rank ontologies. Maiga's method is therefore more specific than the Recommender but unfortunately only a couple of biomedical ontologies are used where as the Recommender proposes a generic and automatic method over 220 biomedical ontologies.

Alani et al. [24] address the scenario where a user wants to search for ontologies relevant for a particular domain (e.g., anatomy). Simple keyword search approaches often do not work in that scenario because of the lack of metadata about ontologies e.g., the term 'anatomy' often does not exists in important anatomy ontologies e.g., Foundational Model of Anatomy. Therefore, they propose a corpus-based approach, experimented in the biomedical domain: when looking for ontologies on a particular topic they retrieve, from the web (e.g., Google or Wikipedia), a collection of terms that represent the given domain (e.g., body, brain, skin, etc. for anatomy). The terms are then used to expand the user query and search in the existing biomedical ontologies (for label and synonyms). Their results show an improvement in retrieval results by 113%, compared to the classic search in ontologies approaches that search only for the user query terms (cf. previous section). However, the method does rank the returned ontologies. Alani et al. Address a specific scenario that is not addressed by the Recommender. We do not expand the user query (keyword or corpus) and assume that the given keywords or corpus data are relevant enough. The rest of the Recommender method is similar to Alani et al. method i.e., lookup for label and synonyms but the Recommender does rank the results and allow using mappings in the process. Other corpus or data-based approaches are described hereafter ([25, 26, 27]) however, Alani et al. and the Recommender distinguish as they are fully automatic (including the corpus selection for Alani et al.).

## Data driven approaches

Data driven approaches are interesting as they allow considering the first scenario identified in introduction: which ontology 'fit' the best for this data? Brewster et al. [25] was the first data driven ontology evaluation approach which consisted in evaluating the appropriateness of an ontology against a corpus. The method consists in (i) identify the keywords & terms of the corpus; (ii) expand the terms using a knowledge base such as WordNet; (iii) map back the indentified terms to the ontology concepts.

OntoSelect [26] is also a corpus-based ontology selection approach and their method is similar to [25]. OntoSelect is an ontology library that helps selecting an ontology for annotation where a given web page is considered as the corpus. OntoSelect's ontology selection method also relies on how well an ontology is connected to other ontologies (using importation mechanism) as well as on structure measures (as described in next section). In [27], the authors proposed a data driven selection approach based on description logic. They use reasoning on ontology module to go beyond simple syntactic approach to achieve topic coverage.

There are two main risks in those approaches: the recommendation is dependent on both the identification of keywords that can miss some important terms and the expansion (in the case of [25] and [26]) of those keywords than can introduce noise.

The Recommender avoids those two drawbacks and gives more importance to concept recognition step (using preferred name and synonyms) assuming the input data is representative enough. The scoring algorithm of the Recommender takes into consideration term frequency as suggested by these methods (in the keyword extraction step).

### Measures based and other miscellaneous approaches

AKTiveRank [10] was the first tool for ontology recommendation that ranks candidate ontologies following several measures based on the ontology structure. The task of selecting candidate ontologies is delegated to an external ontology search engine. AKTiveRank takes keywords as input, and queries Swoogle for the given keywords in order to extract candidate ontologies. Then AKTiveRank applies measures based on the coverage (i.e., finding concepts matching with class and property names) and on the structure of the ontologies to rank them. AKTiveRank is not restricted to any particular domain but, as many other tool mentioned, only deals with ontologies compliant with semantic web standards (Ontology Web Language (OWL) and Resource Description Framework (RDF)). This is a major issue in the biomedical domain where many ontologies are in the OBO or RRF format. The Open Biomedical Ontology format is the format used by the OBO Foundry initiative [32]. The Rich Release Format (RRF) is the format of ontologies embedded in the UMLS Metathesaurus [8]. Moreover, not being restricted to the biomedical domain prevents AKTiveRank to use abstractions such as the 'synonym' property provided by BioPortal or UMLS which significantly enhance the matching step.

OntoQA [28] proposes another set of measures to evaluate the quality of an ontology. Whereas AKTiveRank measures are at the 'concept level' (i.e., measures are applied on (or between) concepts after the first step matching, OntoQA measures are at the 'ontology level' and are independent of any previously identified concepts. For instance, OntoQA looks at the diversity of relations in an ontology schema or the average number of attributes per concepts or the distribution of concepts across the hierarchy. OntoQA originality is that it proposes measures to evaluate the quality of the knowledge based associated with an ontology schema (i.e., instances).

CombiSQORE [29] proposes a method similar to [24]. Assuming a 'semantic query' (decomposed in sub queries) as input, the tool expands the query with WordNet and then uses the result of ontology search engines i.e., Swoogle and Watson to retrieve ontologies. Then measures are applied to rank ontologies based on the similarity between the sub-queries and the ontologies. An originality of CombiSQORE measures that is similar to the Recommender is the use of the ontology size to compare ontologies together.

Hong et al. [30] propose a non automatic interactive approach for selecting an ontology that fits the best a formalized domain. This method is similar to the ontology evaluation approach based on gold standard.

WebCORE [31] (similarly to [21]) introduces another criterion which is the collaborative assessments provided by users in an ontology library platform. They work aims to enhance ontology selection by combining automatic evaluation technique with explicit users' opinions and experiences. Their method consist in a first step in taking a set of initial terms and expanding it through WordNet and then use classic ontology coverage metric to elicit a list of candidate ontology. In a second step, this list is ranked taking into consideration previous feedback and evaluation of the users. The authors are mainly interested in the use case of selecting (non automatically) an ontology for re-use. They do not pay attention to the scenario where automatic quick recommendation is needed.

Sabou et al. [3] present a method for ontology selection that has two interesting originalities: (i) a semantic match that goes beyond the simple syntactic match when doing the matching between keywords and concepts; (ii) the identification of the appropriate ontology combination that will address a complete coverage of the query terms. The method also uses WordNet for query expansion if necessary. However, the method is not fully implemented or available as a service. We note that the Recommender service, in the keyword scenario for which this is relevant, does not provide the smallest set of ontologies that offers 100% coverage of the given keywords. This is another abstraction on the Annotator results that could be interesting to look at.