



Proficient Annotation Recommendation in a Biomedical Content Authoring Environment

Asim Abbas, Steve Mbouadeu, Avinash Bisram, Nadeem Iqbal, Fazel Keshtkar,
and Syed Ahmad Chan Bukhari^(✉)

Collins College of Professional Studies, Division of Computer Science, Mathematics and
Science, St. John's University, 8000 Utopia Parkway, Queens, NY 11439, USA
{abbasa, iqbaln, keshtkaf, bukharis}@stjohns.edu,
{steve.mbouadeu19, avinash.bisram19}@my.stjohns.edu

Abstract. Given the ubiquity of unstructured biomedical data, significant obstacles still remain in achieving accurate and fast access to online biomedical content. Accompanying semantic annotations with a growing volume biomedical content on the internet is critical to enhancing search engines' context-aware indexing, improving search speed and retrieval accuracy. We propose a novel methodology for annotation recommendation in the biomedical content authoring environment by introducing the socio-technical approach where users can get recommendations from each other for accurate and high quality semantic annotations. We performed experiments to record the system level performance with and without socio-technical features in three scenarios of different context to evaluate the proposed socio-technical approach. At a system level, we achieved 89.98% precision, 89.61% recall, and an 89.45% F1-score for semantic annotation recollection. Similarly, a high accuracy of 90% is achieved with the socio-technical approach compared to without, which obtains 73% accuracy. However almost equable precision, recall, and F1- score of 90% is gained by scenario-1 and scenario-2, whereas scenario-3 achieved relatively less precision, recall and F1-score of 88%. We conclude that our proposed socio-technical approach produces proficient annotation recommendations that could be helpful for various uses ranging from context-aware indexing to retrieval accuracy.

Keywords: Annotation recommendation · Automate semantic annotation · Biomedical semantics · Biomedical content authoring · Peer-to-peer · Annotation ranking

1 Introduction

The timely dissemination of information from the scientific research community to peer investigators and other healthcare professionals requires efficient methods for acquiring biomedical publications. The rapid expansion of the biomedical field has led researchers and practitioners to a number of access-level challenges.

This work is supported by the National Science Foundation grant ID: 2101350.

Due to the lack of machine-interpretable metadata (semantic annotation), essential information present in web content is still opaque to information retrieval and knowledge extraction search engines. Search engines require this metadata to effectively index contents in a context-aware manner for accurate biomedical literature searches and to support ancillary activities like automated integration for meta-analysis [1]. Ideally, biomedical content should include machine-interpretable semantic annotations during the pre-publication stage (when first drafting), as this would greatly advance the semantic web's objective of making information meaningful [2]. However both of these procedures are complicated and require in-depth technical and/or domain knowledge. Therefore, a cutting-edge, publicly available framework for creating biological semantic content would be revolutionary.

The generation or processing of textual information through a semantic amplifying framework is known as semantic content authoring. Primary elements of this process include ontologies, annotators, and a user interface (UI). Semantic annotators are built to facilitate tagging/annotating their encompassing ontology concepts using pre-defined terminologies, whether it being through a manually, automatically, or through a hybrid approach [3]. As a result, users create information that is more semantically rich when compared to typical writing utilities such as word processors [4]. Furthermore, it is categorized that the two semantic content writing methodologies used today can be categorized as either bottom-up or top-down. In a bottom-up approach, a collection of ontologies is utilized to semantically enrich or annotate the textual content of a document [5]. Semantic MediaWiki [6], SweetWiki [7], and Linkator [8] are a few examples of bottom-up-designed semantic content production tools. However, these tools have significant drawbacks. The bottom-up approach is an offline, non-collaborative, and application-centric way of content authoring. Additionally, it cannot be used with the most recent version of Microsoft Word because it was created more than eleven years ago. Top-down approaches were developed to add semantic information to existing ontologies, each of which being extended or populated using a particular template design. Hence, this approach is sometimes referred to as an ontology population approach to content production. Top-down approaches do not improve the non-semantic components of text by annotating them with the appropriate ontology keywords. Instead, they use ontology concepts as fillers while authoring content. Examples of top-down approaches are OntoWiki [9], OWiki [10], and RDFAuthor [11].

Over the years, the development of biomedical semantic annotators has received significant attention from the scientific community due to the importance of the semantic annotation process in biomedical informatics research and retrieval. Biomedical annotators can be further classified into a) general-purpose annotators for biomedicine, which assert to cover all biomedical subdomains, and b) use case-specific biomedical annotators, which are developed for a specific sub-domain or annotate specific entities like genes and mutations in a given text. Biomedical annotators primarily use term-to-concept matching with or without machine learning-based methods, in contrast to the general purpose non-biomedical semantic annotators, which combine NLP (Natural Language Processing) techniques, ontologies, semantic similarity algorithms, machine learning (ML) models, and graph manipulation techniques [12]. Biomedical annotators such as NOBLE Coder [13], ConceptMapper [14], Neji [15], and Open Biomedical Annotator

[16] use machine learning and annotate text with an acceptable processing speed. However, they lack a strong disambiguation capacity i.e., the ability to distinguish the proper biomedical concept for a particular piece of text among several candidate concepts. Whereas NCBO Annotator [17] and MGrep services are quite slow. Similarly Rysan-nMd annotator asserts that it can balance speed and accuracy in the annotation process. On the other hand, its knowledge base is restricted to specific UMLS (Unified Medical Language System) ontologies and does not fully cover all biomedical subdomains [18].

To solve the aforementioned constraints, we designed and developed “Semantically” a publicly available interactive systems that enables users with varying levels of biomedical domain expertise to collaboratively author biomedical semantic content. To develop a robust Biomedical Semantic Content Editor, balancing between speed and accuracy is the key research challenge. Finding the proper semantic annotations in real time during content authoring is particularly challenging since a single semantic annotation is frequently available in multiple biomedical ontologies with various connotations. To balance the efficiency and precision of the current biomedical annotators, we present an unconventional socio-technical method for developing a biomedical semantic content authoring system that involves the original author throughout the annotation process. Our system enables users to convert their content into a variety of online, interoperable formats for hosting and sharing in a decentralized fashion. We conducted a series of experiments using biomedical research articles obtained from Pubmed.org [19] to demonstrate the efficacy of the proposed system. The results show that the proposed system achieves a better accuracy when compared to the existing systems used for the same task in the past, leading to a considerable decrease in annotation costs. Our method also introduces a cutting-edge socio-technical approach to leveraging semantic content authoring to enhance the FAIRness [18] of research literature.

2 Proposed Methodology

This section introduces “Semantically,” a web-based, open source, and accessible systems for creating biomedical semantic content that can be used by authors with varying levels of expertise in the biomedical domain. An authoring interface resembling the Microsoft Word editor is provided for end users to write and compose biomedical semantic contents including research papers, clinical notes, and biomedical reports. We leveraged Bioportal [17] endpoint APIs to cater the initial layer of semantic annotation and automate the configuration process for authors. Subsequently, the annotated terms or concepts are highlighted to improve visibility. The system provides a social-collaborative environment in “Semantically Knowledge Cafe” which allows authors to receive assistance from domain experts and peer reviewers to assist in the appropriate annotation of a specific phrase or the entire text. The proposed methodology consists of two primary sections: Base Level Annotations and Annotations generated through Socio-Technical Approach (as shown in Fig. 1).

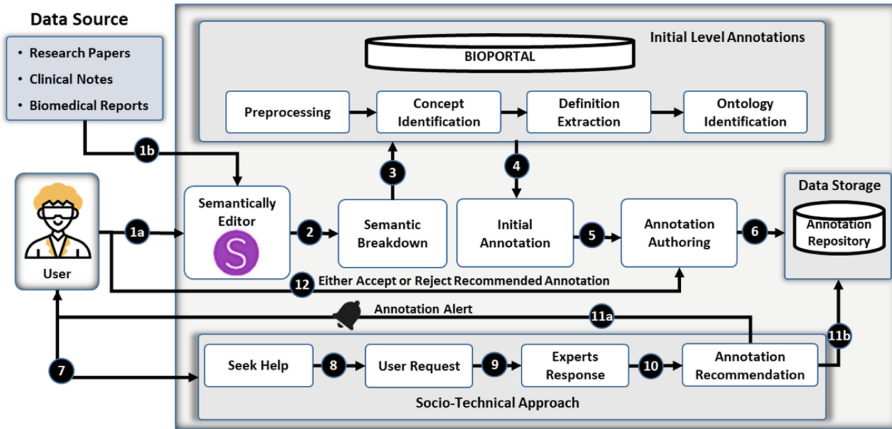


Fig. 1. Proposed methodology workflow

2.1 Base Level Annotation

This semantic annotation approach, a top-down approach, is meant to annotate existing texts using a collection of established ontologies. A biomedical annotator is a key component of a content-level semantic enrichment and annotation process [3]. These annotators leverage publicly accessible biomedical ontology systems like Bioportal [17] and UMLS [5] to assist the biomedical researcher in structuring and annotating their data with ontology concepts that enhance information retrieval and indexing. However, the semantic annotation and enhancement process cannot be easily automated and often requires expert curators. Furthermore, the lack of a user-friendly framework makes the semantic enrichment process more difficult to non-technical individuals. To address this barrier, we used an NCBO Bioportal [17] web-service resource which analyzes the raw textual data, tags it with pertinent biomedical ontology concepts, and provides a basic set of semantic annotations without the requirement of technical expertise.

- Semantically Workspace:** The Semantically framework was developed for a wide range of users, including bench scientists, clinicians, and casual users involved in medical journalism. The users initially have the choice to start typing directly in the Semantically text editor or import pre-existing content from research articles, clinical notes, and biomedical reports. Afterward, based on their level of expertise and knowledge with particular ontologies, authors are provided a few annotation options to choose from. Users without a technical background may readily traverse a simplified interface, while more experienced users can employ advanced options to exert more control over the semantic annotation process, as depicted in Fig. 1.
- Semantic Breakdown:** Semantic breakdown is the process of retrieving semantic information for a biomedical content utilizing the NCBO Bioportal [17] annotator at a granular level. Term based matching as an ontology approach is supported by the NCBO web services and a set of semantic information is returned. The semantic breakdown process consists mainly of (i) Biomedical concept or terminology

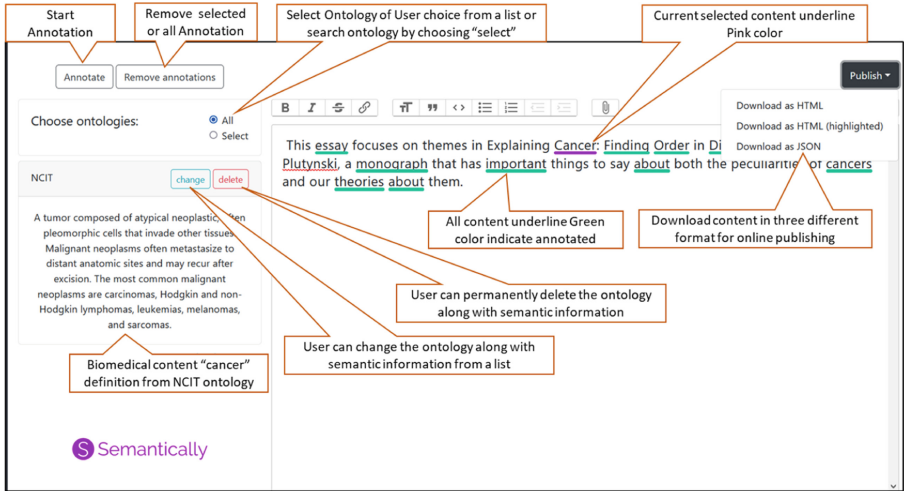


Fig. 2. Proposed biomedical semantic content authoring interface

identification in a series of biomedical data, (ii) Ontological semantic information extraction.

- i. *Biomedical Concept Identification:* We make use of the NCBO BioPortal Annotator API [17] to identify biomedical terms. The NCBO annotator accepts the user's free text which is then mapped using a concept recognition tool. The concept recognition tool leverages Mgrep services developed by NCBO and an ontology-based dictionary created from UMLS [5] and Bioportal [17]. The tool identifies the biomedical terms or concepts following a string matching approach and the result is the collection of biomedical terminologies and ontologies to which the terms belong.
 - ii. *Ontological Semantic Information Extraction:* An ontology is a collection of concepts and the semantic relationships among them. In the proposed methodology, we leverage the Bioportal Ontology web service [17], a repository containing around 1018 ontologies in the biomedical domain. The biomedical text provided to "Semantically" is routed to Bioportal [17] to retrieve the relevant ontologies, acronyms, definitions, and ontology links for individual terminologies. This semantic information is displayed in an annotation panel for user interpretation and comprehension. "Semantically" allows users to author this semantic information based on their knowledge and experience by picking the appropriate ontology from the provided list, suitable acronyms, or eliminating semantic information and annotations for specific terminology.
- C. **Semantic Annotation Authoring Process:** During the semantic breakdown process, a set of base level semantic information is acquired from Bioportal. The associated terminology is then underlined with a green color as illustrated in Fig. 2. The

author is able to click on it to observe and examine the obtained semantic information against recognized biomedical terms or concepts. When an author clicks on a recognized concept, the underlining color changes over to pink to signify that the author has selected it; the semantic information about that concept, including its definition and associated ontology, appear on the left panel, as is in Fig. 2. Additionally, the author is permitted to use the authoring function for specific biomedical terminology that is provided on the left side panel, such as removing annotations, changing underlying ontology, and deleting the ontology all together. To change the selected ontology, the author can click on the “change” button in the left panel, where a list of ontology along with their definitions is given to choose from. Users also have the option to permanently erase an annotation by pressing the “delete” buttons on the left panel, or remove all annotations in the document by pressing “Remove annotations” as shown in Fig. 2. In order to make the application annotation and authoring process independent of external ontology repositories like Bioportal [17], the annotated data is stored in our database for future use.

2.2 The Socio-Technical Approach

After achieving base level semantic annotation, “Semantically” offers a novel socio-technical environment, where the author is able to discuss and receive recommendations from peers to achieve more precise and high quality annotations. Three different scenarios, as given in the Table 1, are used to assess the proposed approach. In order to assess and recommend the appropriate annotation to the author, we use a statistical technique upon gathering the features as given in Table 2. The setting is known as “Semantically Knowledge Cafe,” where the author may post their query, peers or domain experts can react with some self-confidence score, and other community users can credit their reply by up-voting or down-voting in a collaborative and constructive manner. The post author is notified of any recommended annotations with the option of accepting or rejecting the

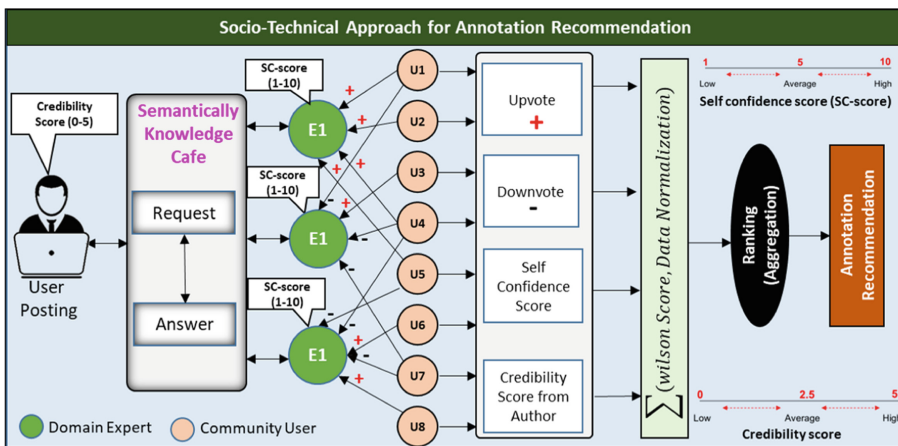


Fig. 3. Proficient annotation recommendation evaluation process

suggestion; only the finally accepted annotation is recorded as an answer in the database. The statistical process of the annotation recommendation is described in Fig. 3.

To evaluate the suggested socio-technical approach, we have devised three scenarios as illustrated in Table 1. These scenarios are related to the query regarding semantic annotation for biomedical content. If we examine the existing question answering platform such as Stackoverflow, such kind of scenarios or query can be found. An example is presented for a Case:1 as below:

The author is asked to identify an accurate ontology annotation from experts for the biomedical term “breast-conserving surgery”. The “Semantically” framework provides an interface where users can input such a query.

Table 1. Semantically annotation recommendation scenarios

Scenario no.	Scenario description
Scenario: 1	Which ontology should i use?
Scenario: 2	What is the suitable ontology vocabulary?
Scenario: 3	Does this ontology best describe this terminology?

For Example:

“Which ontology should I use for “breast-conserving surgery”?”

Submitting the above query will create a new post on the “Semantically Knowledge Cafe” forum for experts E_i response where $E_i = e_1, e_2, e_3 \dots e_n$. Similarly, the platform also provides an interface for the expert E_i to streamline the response process in the forum. For example the expert can describe the suggested annotation, rate their level of self-confidence, and quickly find the precise ontology they mean by utilizing NCBO otology tree widget tool. Whenever the expert E_i submits their response to the post author for Case.1, other community users $U_i = u_1, u_2, u_3 \dots u_n$ do respond to the expert reply in the form, upvote as $+V$ and downvote as V as shown in Fig. 3. A statistical measure is taken for expert self-confidence score, upvote ($+V$), downvote ($-V$) and credibility score from the author by applying Wilson formula and data normalization techniques. Finally an optimal recommendation of annotation is suggested to the author.

2.3 Semantically Recommendation Features (SR-FS)

In order to find more optimal and high quality annotation recommendations, we addressed several features in the proposed socio-technical approach as given in Table 2. These features are produced by and gathered from community users, who actively participated in the socio-technical environment. We presented features as upvote ($+V$), downvote ($-V$), expert confidence score (ECS) and user credibility score (UCS) in the following way.

Table 2. Semantically recommendation features (SR-f) descriptions

Features no.	Features name	Feature descriptions
SR.f1	Upvotes (+), Downvotes (-)	Votes from community users
SR.f2	Self confidence score	Score from responder
SR.f3	Credibility score from author	Score from authors

- (i) **Upvote and downvote (SR-f1):** In the social network environment upvotes (+V) and downvotes (-V) play a crucial role, whereas +V implies the usefulness or quality of a response or answer while -V point to irrelevance or low quality. This feature measures the quality of domain expert responses to the post. A higher amount of up-votes and small amount of down votes indicates better quality annotation recommendations. These features are further processed by Wilson's score confidence interval for a Bernoulli parameter see Eq. 1 to determine the expert suggested annotation quality score.

$$\text{Wilsonscore} = \left(\hat{p} + \frac{Z_{\alpha/2}^2}{2n} \pm Z_{\alpha/2} \sqrt{\left[\hat{p}(1 - \hat{p}) + \frac{Z_{\alpha/2}^2}{4n} \right] / n} \right) / \left(1 + \frac{Z_{\alpha/2}^2}{n} \right) \quad (1)$$

where,

$$\hat{p} = \left(\sum_{n=1}^N +V \right) / (n) \quad (2)$$

$$n = \sum_{i=0}^N \sum_{j=0}^M (+V_i, -V_j) \quad (3)$$

and, $Z_{\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of the standard normal distribution.

In Eq. 1, \hat{p} is the sum of upvotes (+V) of a community user's U_i to the expert E_i response for a post from an author for correct annotation divided by overall votes (+V, V). Likewise n is the sum of number of upvote and downvote (+V, V) and α is the confidence refers to the statistical confidence level: pick 0.95 to have a 95% chance that our lower bound is correct. However the z-score in this function never changes.

- (ii) **Self Confidence Score (SR-f2):** In Psychology a self-confidence score is defined as an individual's trust in their own skills, power, and decision, that they can effectively make. In the proposed approach we allow the expert to give a confidence score for their decision made for annotation recommendation. As shown in Fig. 4, the experts can choose a self confidence level in a range between 1 to 10, indicate how

they feel about annotation recommendation. For instance, if an expert believes that their recommendation is slightly above average, they might rate them as a 6, but if an expert feels more confident, rate them as an 8.

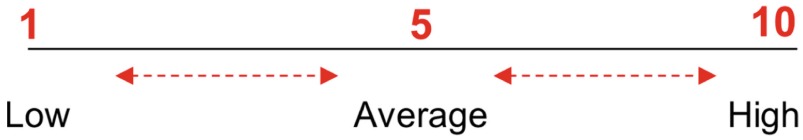


Fig. 4. Self confidence score selection level

$$z_i = (x_i - \min(x)) / (\max(x) - \min(x)) * Q \tag{4}$$

where z_i is the i_{th} normalized value in the dataset. Where x_i is the i_{th} value in the dataset, e.g. the user confidence score. Similarly $\min(x)$ is the minimum value in the dataset, e.g. the minimum value between 1 and 10 is 1, so the $\min(x) = 1$ and $\max(x)$ is the maximum value in the dataset, e.g. the maximum value between 1 and 10 is 10, so the $\max(x) = 10$. Finally Q is the maximum number wanted for normalized data value, e.g. we normalized the confidence score between 0 and 1, the maximum value between 0 and 1 for Q is 1.

- (iii) **Credibility Score from a User (SR-f3):** The credibility is defined as the ability to be trusted or acknowledged as genuine, truthful or honest. As an attribute, credibility is crucial since it helps to determine the domain expert knowledge, experience and profile. Therefore if a domain expert profile is not credible, others are less likely to believe what is being said or recommend. Subsequently annotation recommendation is received to the author post from “Semantically Knowledge Cafe”, the author is allowed to either accept or reject the recommended annotation. Whenever an author accepts a user’s recommended annotation, a credibility score between 1–5 is added to the user’s profile. Similarly, if the author rejects the annotation recommendation, a negative credibility score of 0 is added to their profile (Fig. 5).

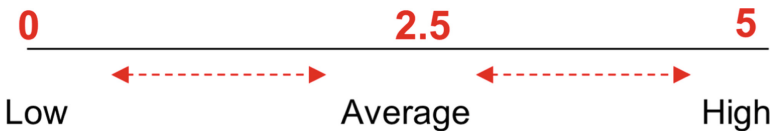


Fig. 5. Credibility score from an author selection level

All the features (SR-f1, SR-f2 and SR-f3) are equally contributed and deeply correlated with one another for final annotation recommendation. Though the final output of SR-f1 is between 0 and 1, Therefore kept the process consistent and features dependencies, we normalize the self-confidence score (SR-f2) between 0

and 1 utilizing Eq. 4. Finally all the SR-FS (Semantically Ranking Feature Score) for each expert E_i annotation recommendation is computed and aggregated using Eq. 5.

$$Sr - Fs = \sum_{j=1}^m \sum_{i=1}^n \sum_{k=0}^p (F_j, E_i, A_k) \quad (5)$$

$$final - score = arg \max \left[\sum_{i=1}^N (Sr - Fs) \right] \quad (6)$$

where F_i is feature score for Annotation A_i and Expert E_i . The final decision or ranking happens based on maximum feature scoring gained by the Expert E_i response to the author post or query see Eq. 6.

3 Experiments

3.1 Datasets

In the proposed methodology a total of 30 people participated. A set of 30 biomedical articles were chosen from Pubmed.org [19] and distributed to the participants randomly. The participants were split into three groups with ten members being assigned to each scenario as we already discussed in proposed methodology Sect. 2.2. Though each user is guided and instructed to process the given documents on ‘‘Semantically’’. Subsequently processing the individual documents by participants a set of base annotations is obtained as outlined in proposed methodology Sect. 2.1. Meanwhile each participant is encouraged to post the query or question on the ‘‘Semantically Knowledge Cafe’’ forum for the biomedical content, which they have doubts of for their initial annotation and a number of 645 total posts is generated. Whereupon, all the participants are allowed to respond to the posts with a confidence score ranging from 1 to 10. Cumulatively 2845 number of answers are given by the expert E_i from the participant and other users participate to weigh the reply post in the form of upvotes and downvotes, an average number of 6056 upvotes and 7942 downvotes are recorded. Subsequently, received the recommended annotation, the author is enabled to give credibility score between 0 and 5. Finally, we reviewed the results without socio-technical and with socio-technical approach by three domain specific experts from academia at professor level.

3.2 Results and Analysis

Precision, Recall, and f1-Score where the accuracy measures used in our proposed methodology. Precision counts the number of valid instances (TP) in the set of all retrieved instances (TP + FP). Recall measures the number of valid instances (TP) in the intended class of instances (TP + FN). Finally the F1-score is the harmonic mean between precision and recall, used to obtain the adjusted F measure. The results were computed using the following equations:

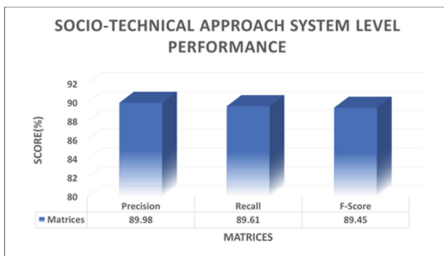
$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

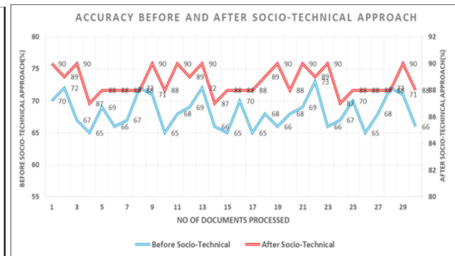
$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{9}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{10}$$

A. **System Level Performance:** A system level performance is constructed by averaging the scenario level results. Initially we find the results at the document level for each scenario. We than combine the results at scenario level and finally averaged the results. After obtaining the results for each scenario, a domain expert from academia, all being domain professors, were engaged to manually assess the results. Thereafter we manually compared the results of domain experts to the socio-technical approach and calculated the precision, recall and f1-score to find the system efficacy. As a result, the system demonstrated almost identical performance of 90% for an annotation recommendation in a socio-technical environment to the manual expert review as shown in Fig. 6a. Additionally, we used Eq. 10 to compare the performance of the system with and without socio-technical approach. A document’s level accuracy is determined with or without the socio-technical approach. In Fig. 6b, the X-axis represents the number of documents processed, while on Y-axis at left represent the level of accuracy without socio-technical approach and on Y-axis at right represent the level of accuracy with socio-technical approach. Consequently, having socio-technical approach is more effective compared not having it at the individual document level. An accuracy of over 90% was achieved by nine documents and the majority of the rest had an accuracy ranging between 87% and 90% using the socio-technical approach. In contrast, only a single document yielded an accuracy above 73% and the rest of the documents had an accuracy between 65% to 73% without the socio-technical approach see Fig. 6b. Overall the proposed socio-technical approach remains the winner by obtaining 89.98% precision, 89.61% recall, and an 89.45% F1-score.



(a) System Level Performance



(b) Without and with socio-technical

Fig. 6. (a) presented the System Level Performance of a Socio-technical approach, (b) presented the accuracy of system without and with socio-technical approach

B. Scenarios Level Performance: This section presented the scenario level performance of the proposed socio-technical approach, where distinguishing results have been produced by these scenarios taking into account the above three matrices precision, recall and f-score. However scrutinizing the results, immense performance is shown by scenario-1 with a 91.14% precision, 90.85% recall and an F-score of 90.53%, and lower efficiency is achieved by scenario-3 with precision of 88.17%, recall 87.82% and f1-score 87.89%. Similarly, an acceptable performance is earned by scenario-2 upon a precision score of 90.63%, recall 90.15% and f1-score 89.9%, which is near to the performance of scenario-1 see Fig. 7. As we have already discussed in the proposed methodology the different working scenarios. However, after carefully examining the outcomes of each scenario, we came to a conclusion: why are distinctive results produced? Whenever the author posted taking scenario-1, the Expert E_i or responder is open and free to suggest the appropriate ontology, and multiple options are available. On the other hand taking scenario-3 author is bounded to choose the appropriate ontology, only from the list suggested by the automatic annotator of NCBO [7]. Also the author explicitly mentioned a list of ontology to the expert in the post to suggest to ontologies only from the given list. Similarly, considering scenario-2, the expert is able to suggest the appropriate ontology vocabulary class by utilizing the NCBO ontology tree widget tool [4] and the expert E_i is not bounded to recommend the ontology vocabulary class from a user specific vocabulary list. Consequently due to bounding the expert E_i to recommend me ontology from author defined vocabulary list scenario-3 achieve less accuracy compared to scenario-1 and scenario-2 that gain an acceptable level of accuracy as shown in Fig. 7.

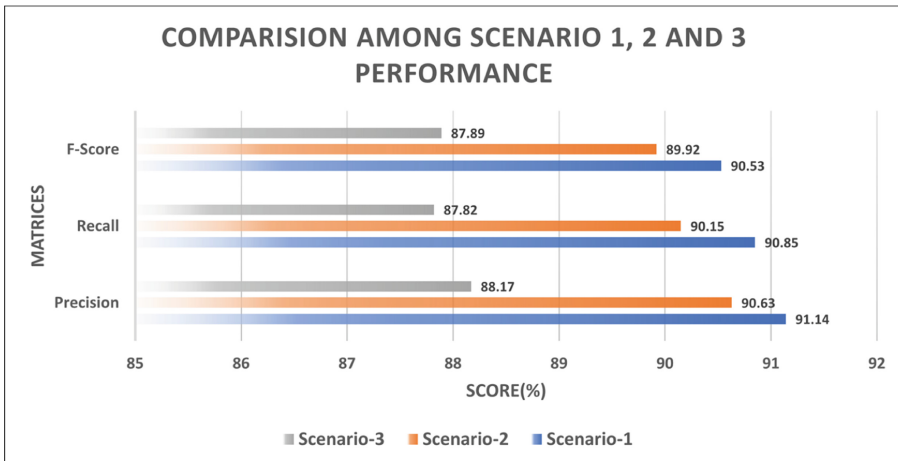


Fig. 7. Scenario level performance comparison

4 Conclusion

One of the main reasons why semantic content authoring is still in its infancy and researchers have not been able to achieve its desired objectives is because researchers did not realize the importance of the original content creator (author). Involvement them rather than heavily focusing on technological sophistication is extremely consequential. The objective of this study is to develop an open source interactive system that empowers individual authors at different levels of expertise in the biomedical domain to generate accurate annotations. Therefore, we proposed a novel socio-technical approach to develop a biomedical semantic content authoring system that balances speed and accuracy by keeping the original author in loop through the entire process. Similarly the “Semantically Knowledge Cafe” is a forum style extension authors can post their query for annotation recommendations. Our work provides a stepping stone towards optimizing biomedical information on the web for search engines, making it more meaningful and useful. The application is available at <https://gosemantically.com>.

References

1. Bukhari, S.A.C.: Semantic enrichment and similarity approximation for biomedical sequence images. University of New Brunswick, Canada (2017)
2. Warren, P., Davies, J.: The semantic web– from vision to reality. *ICT Futures*, pp. 53–66. <https://doi.org/10.1002/9780470758656.ch5>
3. Abbas, A., et al.: Clinical concept extraction with lexical semantics to support automatic annotation. *Int. J. Environ. Res. Public Health* **18**(20), 10564 (2021)
4. Abbas, A., et al.: Biomedical scholarly article editing and sharing using holistic semantic uplifting approach. *Int. FLAIRS Conf. Proc.* **35** (2022)
5. Abbas, A., et al.: Meaningful information extraction from unstructured clinical documents. *Proc. Asia Pac. Adv. Netw.* **48**, 42–47 (2019)
6. Laxström, N., Kanner, A.: Multilingual semantic MediaWiki for Finno-Ugric dictionaries. *Septentrio Conf. Ser.* **2**, 75 (2015). <https://doi.org/10.7557/5.3470>
7. Buffa, M., et al.: SweetWiki: a semantic Wiki. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3199377>
8. Araujo, S., Houben, G.-J., Schwabe, D.: Linkator: enriching web pages by automatically adding dereferenceable semantic annotations. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) *ICWE 2010. LNCS*, vol. 6189, pp. 355–369. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13911-6_24
9. Auer, S., Dietzold, S., Riechert, T.: OntoWiki – a tool for social, semantic collaboration. In: Cruz, I., et al. (eds.) *ISWC 2006. LNCS*, vol. 4273, pp. 736–749. Springer, Heidelberg (2006). https://doi.org/10.1007/11926078_53
10. Iorio, A.D., et al.: OWiki: enabling an ontology-led creation of semantic data. In: Hippe, Z.S., Kulikowski, J.L., Mroczek, T. (eds.) *Human – Computer Systems Interaction: Backgrounds and Applications 2. AISC*, vol. 99, pp. 359–374. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23172-8_24
11. Tramp, S., Heino, N., Auer, S., Frischmuth, P.: RDFauthor: employing RDFa for collaborative knowledge engineering. In: Cimiano, P., Pinto, H.S. (eds.) *EKAW 2010. LNCS (LNAI)*, vol. 6317, pp. 90–104. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16438-5_7

12. Jovanović, J., et al.: Semantic annotation in biomedicine: the current landscape. *J. Biomed. Semant.* **8**(1), 44 (2017)
13. Tseytlin, E., et al.: NOBLE – flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinform.* **17**(1) (2016). <https://doi.org/10.1186/s12859-015-0871-y>
14. Funk, C., et al.: Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. *BMC Bioinform.* **15**, 59 (2014)
15. Campos, D., Matos, S., Oliveira, J.L.: A modular framework for biomedical concept recognition. *BMC Bioinform.* **14**(1) (2013). <https://doi.org/10.1186/1471-2105-14-281>
16. Shah, N.H., et al.: Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinform.* **10**(S9) (2009). <https://doi.org/10.1186/1471-2105-10-s9-s14>
17. Jonquet, C., et al.: NCBO annotator: semantic annotation of biomedical data. In: *International Semantic Web Conference, Poster and Demo session*, Washington DC, USA, vol. 110, October 2009
18. Mbouadeu, S.F., et al.: Towards structured biomedical content authoring and publishing. In: *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*. IEEE (2022)
19. PubMed Help: PubMed Help. National Center for Biotechnology Information, USA (2020)