

# QuickGraph: A Rapid Annotation Tool for Knowledge Graph Extraction from Technical Text

Tyler Bikaun<sup>(✉)</sup>, Michael Stewart and Wei Liu

The University of Western Australia  
35 Stirling Highway, Crawley, Western Australia  
tyler.bikaun@research.uwa.edu.au  
michael.stewart@uwa.edu.au  
wei.liu@uwa.edu.au

## Abstract

Acquiring high-quality annotated corpora for complex multi-task information extraction (MT-IE) is an arduous and costly process for human-annotators. Adoption of unsupervised techniques for automated annotation have thus become popular. However, these techniques rely heavily on dictionaries, gazetteers, and knowledge bases. While such resources are abundant for general domains, they are scarce for specialised technical domains. To tackle this challenge, we present **QuickGraph**<sup>1</sup>, the first collaborative MT-IE annotation tool built with indirect weak supervision and clustering to maximise annotator productivity.

QuickGraph’s main contribution is a set of novel features that enable knowledge graph extraction through rapid and consistent complex multi-task entity and relation annotation. In this paper, we discuss these key features and qualitatively compare QuickGraph to existing annotation tools. A demonstration of our system is available at: <https://youtu.be/ZlzH-AAoGXs>.

## 1 Introduction

Hand-labelling is still the most reliable means to obtain quality training data to support deep learning applications; however, it is time-consuming and resource-intensive (Pustejovsky and Stubbs, 2012). Unsupervised approaches such as weak/distant supervision (Craven and Kumlien, 1999; Mintz et al., 2009) and data programming (Ratner et al., 2017) thus are usually attractive alternatives or starting points to human-annotation.

Leveraging unsupervised techniques, however, is predicated on the availability of relevant external resources such as semantically aligned knowledge bases, and a priori knowledge of phenomena/concepts in the corpus of interest.

In general domains, these are widely available, e.g. YAGO (Suchanek et al., 2007), Freebase (Bollacker et al., 2008), Wikidata (Vrandečić and Krötzsch, 2014), DBPedia (Lehmann et al., 2015), whereas they remain scarce for emerging and specialised domains, such as engineering, industrial, medical, biological, law enforcement (Neves and Leser, 2012; Dima et al., 2021) which we refer to as technical domains. Due to their close real-world applications, technical domains are often more impactful and likely to have formal ontologies of engineered knowledge. Consequently, human-annotation remains critical and essential for obtaining quality training data for technical information extraction and instance population.

Numerous annotation tools exist, supporting many NLP tasks (Neves and Ševa, 2019). However, few tools support large-scale, hierarchical, multi-task, multi-label, entity and relation annotation that is required for translating NLP research to real-world industry applications in technical domains (Stenetorp et al., 2012; Yimam et al., 2013; Klie et al., 2018; Stewart et al., 2019; Abrami et al., 2019; Islamaj et al., 2020; Tang et al., 2020). Moreover, these tools lack features to optimise annotator productivity and return-on-time-invested. Such features are particularly essential for technical domains, as annotators are frequently subject matter experts, who are typically time-poor and costly.

Weak/distant supervision (Craven and Kumlien, 1999; Mintz et al., 2009) is a powerful paradigm for large-scale (potentially low-quality) automatic annotation. Surprisingly, the integration of this paradigm into annotation tools to accelerate labelled sample acquisition for deep learning applications remains unexplored.

To fill these gaps, we introduce **QuickGraph**, the first collaborative annotation tool for multi-task IE that is designed to be:

- *Fast*: Accelerates annotation via entity and relation propagation, and semantic clustering.

<sup>1</sup>QuickGraph. <https://quickgraph.nlp-tp.org>

- *Powerful*: Supports complex multi-task entity and open/closed relation annotation and knowledge graph extraction.
- *Intuitive*: Simple to set-up and use.
- *Efficient*: Optimises annotation through easily-configurable relation constraints.
- *Insightful*: Builds real-time knowledge graphs from annotations<sup>2</sup>, and provides three dimensions of inter-annotator agreement.

## 2 Related work

Many NLP annotation tools have been developed in recent years (Neves and Ševa, 2019), however few support entity and relation annotation as a single integrated task, nor do they contain purposely designed features to enhance annotator productivity. Here we discuss the most notable entity and relation annotation tools.

Historically, *brat* (Stenetorp et al., 2012) has been the most popular but routinely receives criticism for its antiquated technology and set-up difficulty (Kummerfeld, 2019; Neves and Ševa, 2019). Similarly, *WebAnno* (Yimam et al., 2013), *INCEpTION* (Klie et al., 2018) and *TextAnnotator* (Abrami et al., 2019) are feature-rich and multi-purpose tools, but are challenging to use. *SLATE* (Kummerfeld, 2019) is a command-line-based tool, but lacks multi-task functionality and is restricted to technical end users because of its command-line design. *TeamTat* (Islamaj et al., 2020) is a powerful tool, but is oriented for small-batch complex annotation of large documents such as scholarly articles. *Redcoat* (Stewart et al., 2019) is a feature-rich entity typing tool and has been demonstrated to support MT-IE (Stewart and Liu, 2020), but is not purpose-built for relation annotation. *SALKG* (Tang et al., 2020) is a unique knowledge-graph annotation tool, yet lacks features for collaborative annotation and adjudication. Despite this, each of the aforementioned tools contain powerful elements, but universally lack features to support rapid large-scale, complex, annotation.

Recent tools enhance annotator productivity using active and proactive learning, including *APLeNTy* (Nghiem and Ananiadou, 2018), *Paladin* (Nghiem et al., 2021), *FitAnnotator* (Li et al., 2021) and *ActiveAnno* (Wiechmann et al., 2021). Inadequately, these tools cannot perform multi-task entity and relation annotation. Moreover, their performance using large ontologies remains unproven (Nghiem and Ananiadou, 2018; Li et al., 2021). Reliance on

<sup>2</sup>When both entity and relation annotation is performed.

active learning may also result in unsatisfactory corpus quality due to sample acquisition and reliability concerns (Attenberg and Provost, 2011; Lowell et al., 2019).

## 3 System highlights

### 3.1 Key capabilities

QuickGraph is a multi-task document-level hierarchical entity and relation annotation tool. Our tool supports annotations that are: i) hierarchical, ii) multi-label, iii) multi-class, and iv) nested. These attributes enable annotation of tasks such as named entity recognition, coreference resolution, entity typing, part-of-speech tagging, relation extraction, semantic role labelling, and triple annotation. One of QuickGraph’s novelties is its support for open<sup>3</sup> and closed<sup>4</sup> relation annotation, permitting open relation extraction tasks (Niklaus et al., 2018; Stanovsky et al., 2018).

Additional key contributions of QuickGraph are its novel features for indirect weak supervision through annotation propagation (Section 3.2.2), semantic clustering of documents to promote annotator consistency (Section 3.4.1), and real-time knowledge graph construction from text (Section 3.6.1). Each of these contributions enable rapid annotation of corpora to support deep learning applications without the need of external resources such as knowledge bases, dictionaries, or gazetteers.

### 3.2 Why is QuickGraph fast?

#### 3.2.1 Get started - quick

QuickGraph is available for free online and takes only minutes to create an account and set-up a project for rapid annotation. Our tool provides preset ontologies for popular entity and relation annotation tasks including ConceptNet-5.5 (Speer et al., 2017), CoNLL03 (Tjong Kim Sang and De Meulder, 2003), FIGER (Ling and Weld, 2021), SemEval-2007 Task 04 (Girju et al., 2007), and SemEval-2010 Task 8 (Hendrickx et al., 2010).

#### 3.2.2 Entity and relation propagation

QuickGraph’s novel entity and relation propagation features enable annotators to make *a click worth a thousand annotations* (Figure 1B-D), analogous to the adage *a picture is*

<sup>3</sup>Relations are unbounded surface form linguistic expressions.

<sup>4</sup>Relations are bounded and predefined.

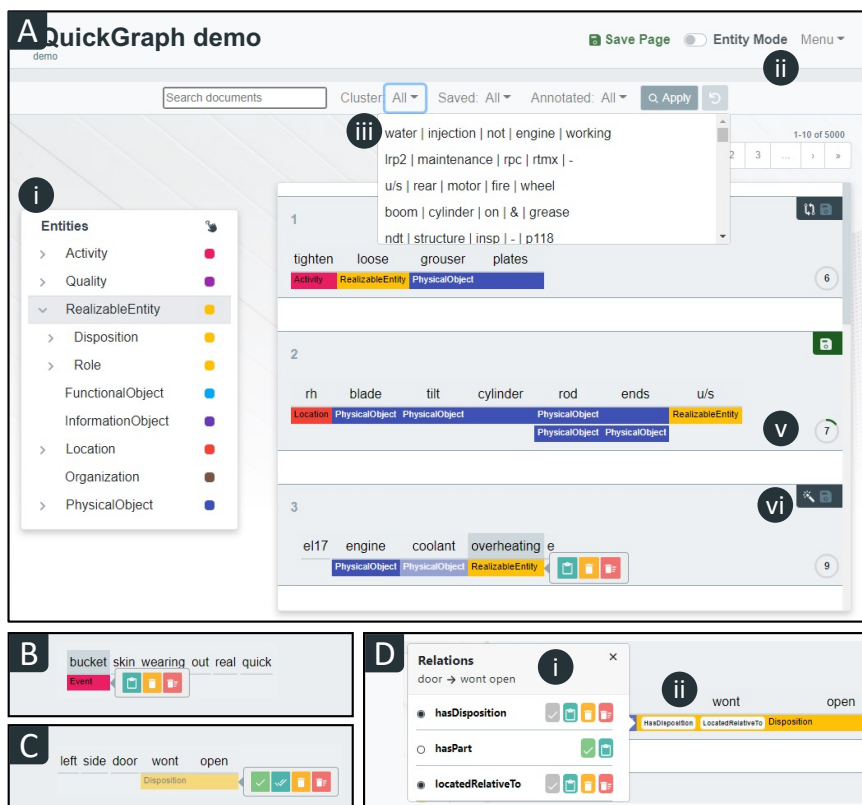


Figure 1: QuickGraph’s annotation interface - A) main interface: i. hierarchical label palette, ii. annotation mode toggle, iii. cluster navigation, vi. document status tray, and v. cluster tray, B) *accepted* entity tooltip: buttons left-right: *apply all*, *delete one*, *delete all*, *quick search*, C) *suggested* entity tooltip: buttons left-right: *accept one*, *accept all*, *reject one*, *reject all*, *quick search*, and D) relation annotation: i. relation popover, and ii. applied relation.

worth a thousand words. Entity propagation is performed through case-insensitive sub-string matching, and relation propagation is implemented through a deterministic token/phrase offset matching algorithm. For unambiguous and consistently offset tokens and phrases, propagation can massively speed up annotation without compromising precision (Figure 1B-D). For example, users of QuickGraph can apply thousands of entities or hundreds or relations in a single click. As a result, these features enhance productivity and contribute to quickly capturing diverse contextual annotations to support deep learning applications.

The process of propagation involves cascading *suggested* (weak) annotations across the entire corpus, emulating weak supervision, but without the need for external resources. Throughout the annotation process, *suggested* annotations can be viewed and individually or bulk accepted, converting them into *accepted* (silver) annotations. At any point throughout the project, all created annotations can be downloaded. The presence of weakly labelled documents can be used in a similar fashion to their

treatment in unsupervised learning methods (Ratner et al., 2017). Gold annotations are automatically generated by aggregating entity mentions and/or triples with respect to a desired inter-annotator agreement threshold.

### 3.2.3 Pre-annotation

Like other tools, our tool permits pre-annotation of corpora at project creation. Pre-annotation reduces annotation effort by pre-applying labels based on external resources such as gazetteers. A novel feature of QuickGraph is its ability to pre-annotate both entities and relations through sets of pre-labelled artefacts<sup>5</sup>.

### 3.2.4 Built-in corpus pre-processing

QuickGraph supports corpus pre-processing as part of the project creation process rather than requiring external solutions. Consequently, corpora can be annotated end-to-end without external steps or dependencies, simplifying

<sup>5</sup>In the format of  $\langle s_{span}, s_{type}, r_{type}, t_{span}, t_{type}, t_{offset} \rangle$  where s - source, r - relation, and t - target.

and speeding up the annotation process. Pre-processing stages currently consists of: i) character casing, ii) character removal, and iii) document deduplication.

### 3.3 Why is QuickGraph powerful?

#### 3.3.1 Thousands of documents - at once

Unlike other tools, QuickGraph prospers with large-scale corpora. We have loaded and simultaneously annotated corpora consisting of 100,000 short user-generated texts with the tool whilst maintaining performance. While other tools limit annotators to a view small group of documents (Nghiem and Ananiadou, 2018; Kummerfeld, 2019; Islamaj et al., 2020; Li et al., 2021), QuickGraph users can view up to 100 documents simultaneously. Support for large document groups promotes quick identification of cross-document information, aiding annotators by viewing concepts in different contexts.

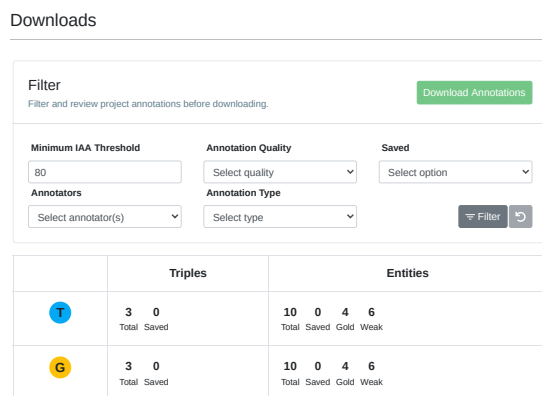


Figure 2: QuickGraph’s flexible downloads component.

#### 3.3.2 Annotation export flexibility

Exporting annotations to support deep learning applications is easy in QuickGraph (Figure 2). At any time, downloads can be filtered and exported. Our tool allows users to filter annotations based on their: i) inter-annotator agreement score, ii) quality (e.g. *gold*, *silver* or *weak*), iii) saved state, iv) annotator(s), and v) annotation type (e.g. entity mentions or triples).

### 3.4 How does QuickGraph help consistency?

#### 3.4.1 Semantic clustering

An overlooked feature of current tools is document clustering to promote annotator productivity. Clustering is a core feature of QuickGraph,

and has two primary benefits. First, annotators maintain a consistent mental model whilst annotating as clustered documents are likely to share semantic and express similar phenomena. Second, user actions are simplified as similar documents likely share similar concepts reducing the need to repetitively navigate through large hierarchical entity label spaces. Our tool implements agglomerative clustering of documents embedded with SBERT (Reimers and Gurevych, 2019) sentence embeddings.

#### 3.4.2 Powerful ontology editor and relation constraints

Applying relation annotations consistently is difficult and time-consuming (Mintz et al., 2009). Besides preset ontologies available within QuickGraph, custom entity and relation ontologies can easily be created. Unlike other IE tools, our tool supports hierarchical entities and is the first to permit relation constraints through entity domain and ranges.

Relation constraints are made possible as QuickGraph applies relations between entities (*associated with token spans*) rather than on token spans directly (see Figure 1D). Consequently, this feature can enhance annotator productivity and consistency through restrained relation selection. For annotators using large formal ontologies, such as those found in technical domains, this can significantly reduce the search space of relations by using pre-defined domains and ranges.

### 3.5 Why is QuickGraph intuitive?

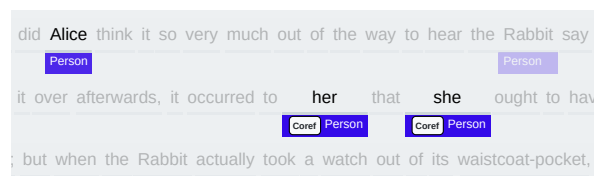


Figure 3: Example of relation annotation mode. *Alice* is the source entity.

#### 3.5.1 Minimalistic interface

Instead of providing annotators with *everything but the kitchen sink*, akin to the current generation of annotation tools (Abrami et al., 2019, 2020), our tool has been designed with the tenet of minimalism. Significantly, this has been applied to the presentation of relations. Instead of rendering dependencies between entities and relations as free-flowing arrows, QuickGraph renders only what annotators choose to see, as

means of promoting focused annotation (Figure 3). Additionally, toggling between entity and relation annotation is seamless, requiring only a single click or key press (Figure 1A.ii).

### 3.5.2 Cluster annotation and navigation

At any time<sup>6</sup>, QuickGraph users can drill in and out of document clusters with a single click (Figure 1A.v). Navigation between clusters is also trivial owing to interpretable cluster descriptions, each derived from their document sets top-n terms (Figure 1A.iii)

## 3.6 Why is QuickGraph insightful?

### 3.6.1 Real-time knowledge graphs

Novel to QuickGraph is its real-time knowledge graph construction from annotations<sup>7</sup>. This feature enables annotators to gain insight into, and improve understanding of, their annotations. Two graph types are available for annotated documents: i) aggregated; documents are aggregated together, and ii) separated; documents are represented as sub-graphs.

Adjudicator

Document  
replace burst hydraulic hose 21% 0% 5%  
entities relations overall

	Source	Relation	Target	Annotation
1	replace MaintenanceActivity	hasParticipant	hose GainingObject	replace burst hydraulic hose MaintenanceActivity MaintenanceActivity
2	burst MaintenanceActivity	hasParticipant	hose GainingObject	replace burst hydraulic hose MaintenanceActivity MaintenanceActivity
3	hose GainingObject	contains	hydraulic ManufacturingObject	burst hydraulic hose Component GainingObject GainingObject
4	replace MaintenanceActivity	hasParticipant	hydraulic hose GainingObject	replace burst hydraulic hose MaintenanceActivity MaintenanceActivity
5	hydraulic hose GainingObject	hasDisposition	burst Disposition	replace burst hydraulic hose MaintenanceActivity GainingObject GainingObject

Showing 1 to 5 of 5 annotations

Figure 4: Example of three dimension adjudication for a terse document with two annotators.

### 3.6.2 Multi-dimensional adjudication

Adjudication in our tool (Figure 4) is supported by three dimensions of inter-annotator agreement (IAA) : i) triples (referred to as *overall*), ii) entities, and iii) relations. Inspired by SemBLEU (Song and Gildea, 2019), pair-wise IAA is calculated through a modified-BLEU score (Papineni et al., 2002). Currently, IAA is

<sup>6</sup>If clustering was selected upon project creation.

<sup>7</sup>If entity and relation annotation was selected upon project creation.

strictly enforced on directionality and hierarchical entities and relation types. Adding relaxed IAA will be the focus of future development.

## 4 System architecture

QuickGraph is a multi-user tool built using the modern full-stack framework MERN<sup>8</sup>, Python and Docker. Our tool consists of four containerised components (Figure 5): i) web client, ii) NoSQL database, iii) server, and iv) cluster server. Using Docker, our tool can be built and ready to annotate in minutes<sup>9</sup>.

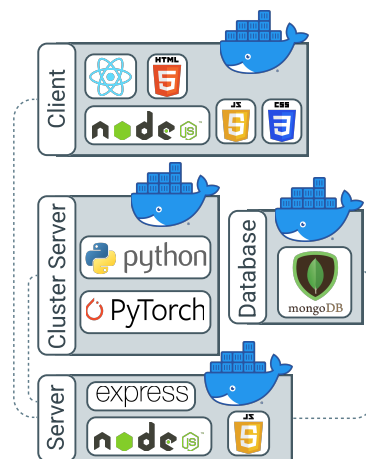


Figure 5: QuickGraph's system architecture and technology stack.

QuickGraph's NoSQL database consists of the three collections: *Projects*, *Texts* and *Users*. *Projects* contain information pertinent to the project's: manager, name, description, assigned annotators, settings, details of tasks, pre-processing operations, uploaded texts, clustering details, and entity and relation ontology information. *Texts* consist of all texts including details of their: original value, tokens, entity and relation markup, annotator saved states, weight, rank, and cluster designation. Lastly, the *Users* collection contains information such as the users: username, hashed and salted password, email, personalisation settings, and assigned and invited projects.

## 5 Comparison with existing tools

A qualitative comparison between QuickGraph and existing open-source annotation tools that support entity and relation annotation, or have

<sup>8</sup>MongoDB-Express-React-Node. <https://www.mongodb.com/mern-stack>

<sup>9</sup>QuickGraph GitHub. <https://github.com/nlp-tlp/quickgraph>

Tool	Annotation Type	Multi-label	Pre-annotation	Auto-annotation	Relation Constraints	Annotation Propagation	Document Clustering
<b>QuickGraph</b>	E/HE/R	✓	✓	-	✓	✓	✓
ActiveAnno	D	✓	✓	✓	-	-	-
APLenty	E	-	-	✓	-	-	-
brat	E/R	✓	-	-	-	-	-
FITAnnotator	E/R	-	-	✓	-	-	-
INCEpTION	E/R	✓	-	✓	✓	-	-
Paladin	D	✓	-	✓	-	-	-
RedCoat	E/HE	✓	✓	-	-	-	-
SALKG	T	-	✓	-	-	-	-
SLATE	E/R	✓	-	-	-	-	-
TeamTat	E/R	✓	✓	-	-	✓	-
TextAnnotator	E/R	✓	✓	✓	-	-	-
WebAnno	E/R	✓	✓	-	-	-	-

Table 1: Comparison of QuickGraph’s features to 12 popular, existing open-source annotation tools. Annotation type abbreviations: E - entity, HE - hierarchical entity, R - relation, T - triple, D - document.

design features to enhance annotator productivity, is provided in Table 1.

*Annotation type:* 75% of the reviewed tools support entity annotation, with most also allowing relation annotation. Only RedCoat (and QuickGraph) permit hierarchical entity annotation.

*Multi-label:* 75% of the reviewed tools support multi-label annotation. Of these, ActiveAnno and Paladin permit multi-labels, but are restricted to document classification tasks.

*Pre-annotation:* 50% of the reviewed tools allow pre-annotation of corpora prior to manual labelling. These tools are limited to entities, while QuickGraph also supports relations through triples.

*Automatic annotation:* Less than 50% of the reviewed tools support automatic annotation. This feature is implemented through AI-assistance, typically using active learning, and is limited to entity annotation. QuickGraph purposely does not have this feature, as we believe uncontrolled automatic annotation for complex MT-IE can be unproductive.

*Relation constraints:* Of the reviewed tools, only INCEpTION allows for relation constraints. However, INCEpTION’s constraints need to be expressed in a bespoke constraint language. In contrast, this feature of QuickGraph requires users to simply specify entity domain and ranges on relations.

*Annotation propagation:* Of the reviewed tools, only TextAnnotator provides annotation propagation via ‘entity cascading’. However, this feature is restricted to entities, and the tool’s interface is cumbersome and challenging to use. In contrast, QuickGraph allows easy and intuitive entity and relation propagation.

*Document clustering:* No reviewed tool offer

document clustering. Only QuickGraph enables document clustering to improve annotator productivity and consistency.

## 6 Conclusion and future work

We introduced QuickGraph, a collaborative annotation tool for multi-task information extraction that accelerates annotator productivity. Distinguishing features of QuickGraph are its support for diverse information extraction tasks through hierarchical entity and open/closed relation annotation, annotation propagation, and semantic clustering.

Whilst QuickGraph is ready to use, there are features still under development, including: i) expanding available semantic embedding and clustering options, ii) improving annotation propagation processes, iii) relaxing inter-annotator agreement metrics, and iv) adding support for cross-document annotation.

## Acknowledgements

This research is supported by the Australian Research Council through the Centre for Transforming Maintenance through Data Science (grant number IC180100030), funded by the Australian Government. Additionally, Bikaun acknowledges funding from the Mineral Research Institute of Western Australia. Bikaun and Liu acknowledge the support from ARC Discovery Grant DP150102405.

The authors are grateful to Mrs. Shu Wong for her assistance and helpful feedback on the user-interaction and user-experience elements of QuickGraph.

## References

- Giuseppe Abrami, Alexander Mehler, Andy Lücking, Elias Rieb, and Philipp Helfrich. 2019. [Textannotator: A flexible framework for semantic annotations](#). In *Proceedings of the 15th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-15)*. Association for Computational Linguistics, London, UK, pages 1–12.
- Giuseppe Abrami, Manuel Stoeckel, and Alexander Mehler. 2020. [TextAnnotator: A UIMA based tool for the simultaneous and collaborative annotation of texts](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 891–900, Marseille, France. European Language Resources Association.
- Josh Attenberg and Foster Provost. 2011. [Inactive learning? difficulties employing active learning in practice](#). *SIGKDD Explor. Newsl.*, 12(2):36–41.
- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. [Freebase: A collaboratively created graph database for structuring human knowledge](#). In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, page 1247–1250, New York, NY, USA. Association for Computing Machinery.
- M Craven and J Kumlien. 1999. [Constructing biological knowledge bases by extracting information from text sources](#). In *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, page 77–86.
- Alden Dima, Sarah Lukens, Melinda Hodkiewicz, Thurston Sexton, and Michael P. Brundage. 2021. [Adapting natural language processing for technical text](#). *Applied AI Letters*, 2(3):e33.
- Roxana Girju, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney, and Deniz Yuret. 2007. [SemEval-2007 task 04: Classification of semantic relations between nominals](#). In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 13–18, Prague, Czech Republic. Association for Computational Linguistics.
- Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. [SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals](#). In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 33–38, Uppsala, Sweden. Association for Computational Linguistics.
- Rezarta Islamaj, Dongseop Kwon, Sun Kim, and Zhiyong Lu. 2020. [Teamtat: a collaborative text annotation tool](#). *Nucleic acids research*, 48(W1):W5–W11.
- Jan-Christoph Klie, Michael Bugert, Beto Boulos, Richard Eckart de Castilho, and Iryna Gurevych. 2018. [The inception platform: Machine-assisted and knowledge-oriented interactive annotation](#). In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics. Event Title: The 27th International Conference on Computational Linguistics (COLING 2018).
- Jonathan K. Kummerfeld. 2019. [SLATE: A super-lightweight annotation tool for experts](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 7–12, Florence, Italy. Association for Computational Linguistics.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. [Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia](#). *Semantic web*, 6(2):167–195.
- Yanzeng Li, Bowen Yu, Li Quangang, and Tingwen Liu. 2021. [FITAnnotator: A flexible and intelligent text annotation system](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 35–41, Online. Association for Computational Linguistics.
- Xiao Ling and Daniel Weld. 2021. [Fine-grained entity recognition](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 26(1):94–100.
- David Lowell, Zachary C. Lipton, and Byron C. Wallace. 2019. [Practical obstacles to deploying active learning](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 21–30, Hong Kong, China. Association for Computational Linguistics.
- Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. [Distant supervision for relation extraction without labeled data](#). In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011, Suntec, Singapore. Association for Computational Linguistics.
- Mariana Neves and Ulf Leser. 2012. [A survey on annotation tools for the biomedical literature](#). *Briefings in Bioinformatics*, 15(2):327–340.
- Mariana Neves and Jurica Ševa. 2019. [An extensive review of tools for manual annotation of documents](#). *Briefings in Bioinformatics*, 22(1):146–163.

- Minh-Quoc Nghiem and Sophia Ananiadou. 2018. [APLenty: annotation tool for creating high-quality datasets using active and proactive learning](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 108–113, Brussels, Belgium. Association for Computational Linguistics.
- Minh-Quoc Nghiem, Paul Baylis, and Sophia Ananiadou. 2021. [Paladin: an annotation tool based on active and proactive learning](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 238–243, Online. Association for Computational Linguistics.
- Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2018. [A survey on open information extraction](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3866–3878, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: A method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, page 311–318, USA. Association for Computational Linguistics.
- James Pustejovsky and Amber Stubbs. 2012. *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. "O'Reilly Media, Inc."
- Alexander J. Ratner, Stephen H. Bach, Henry R. Ehrenberg, and Chris Ré. 2017. [Snorkel: Fast training set generation for information extraction](#). In *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD '17*, page 1683–1686, New York, NY, USA. Association for Computing Machinery.
- Nils Reimers and Iryna Gurevych. 2019. [Sentence-BERT: Sentence embeddings using Siamese BERT-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Linfeng Song and Daniel Gildea. 2019. [SemBleu: A robust metric for AMR parsing evaluation](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4547–4552, Florence, Italy. Association for Computational Linguistics.
- Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. [Conceptnet 5.5: An open multilingual graph of general knowledge](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).
- Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. [Supervised open information extraction](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 885–895.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. [brat: a web-based tool for NLP-assisted text annotation](#). In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.
- Michael Stewart and Wei Liu. 2020. [Seq2KG: An End-to-End Neural Model for Domain Agnostic Knowledge Graph \(not Text Graph\) Construction from Text](#). In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning*, pages 748–757.
- Michael Stewart, Wei Liu, and Rachel Cardell-Oliver. 2019. [Redcoat: A collaborative annotation tool for hierarchical entity typing](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 193–198, Hong Kong, China. Association for Computational Linguistics.
- Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. [Yago: A core of semantic knowledge](#). In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, page 697–706, New York, NY, USA. Association for Computing Machinery.
- Mingwei Tang, Cui Su, Haihua Chen, Jingye Qu, and Junhua Ding. 2020. [Salkg: A semantic annotation system for building a high-quality legal knowledge graph](#). In *2020 IEEE International Conference on Big Data (Big Data)*, pages 2153–2159.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003. [Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition](#). pages 142–147.
- Denny Vrandečić and Markus Krötzsch. 2014. [Wikidata: a free collaborative knowledgebase](#). *Communications of the ACM*, 57(10):78–85.
- Max Wiechmann, Seid Muhie Yimam, and Chris Biemann. 2021. [ActiveAnno: General-purpose document-level annotation tool with active learning integration](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 99–105, Online. Association for Computational Linguistics.
- Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. [WebAnno: A flexible, web-based and visually](#)



supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia, Bulgaria. Association for Computational Linguistics.