

Inside ASCENT: Exploring a Deep Commonsense Knowledge Base and its Usage in Question Answering

Tuan-Phong Nguyen

Simon Razniewski

Gerhard Weikum

Max Planck Institute for Informatics

Saarbrücken, Germany

{tuanphong, srazniew, weikum}@mpi-inf.mpg.de

Abstract

ASCENT is a fully automated methodology for extracting and consolidating commonsense assertions from web contents (Nguyen et al., 2021). It advances traditional triple-based commonsense knowledge representation by capturing semantic facets like locations and purposes, and composite concepts, i.e., subgroups and related aspects of subjects. In this demo, we present a web portal that allows users to understand its construction process, explore its content, and observe its impact in the use case of question answering. The demo website¹ and an introductory video² are both available online.

1 Introduction

Commonsense knowledge (CSK) is an enduring theme of AI (McCarthy, 1960) that has been recently revived for the goal of building more robust and reliable applications (Monroe, 2020). Recent years have witnessed the emerging of large pre-trained language models (LMs), notably BERT (Devlin et al., 2018), GPT (Brown et al., 2020) and their variants which significantly boosted the performance of tasks requiring natural language understanding such as question answering and dialogue systems (Clark et al., 2020). Although it has been shown that such LMs implicitly store some commonsense knowledge (Talmor et al., 2019), this comes with various caveats, for example regarding degree of truth, or negation, and their commercial development is inherently hampered by their low interpretability and explainability.

Structured knowledge bases (KBs), in contrast, give a great possibility of explaining and interpreting outputs of systems leveraging the resources. There have been great efforts towards building large-scale commonsense knowledge bases

(CSKBs), including expert-annotated KBs (e.g., Cyc (Lenat, 1995)), crowdsourced KBs (e.g., ConceptNet (Speer and Havasi, 2012) and Atomic (Sap et al., 2019)) and KBs built by automatic acquisition methods such as WebChild (Tandon et al., 2014, 2017), TupleKB (Mishra et al., 2017), Quasimodo (Romero et al., 2019) and CSKG (Ilievski et al., 2020). Human-created KBs, although possessing high precision, usually suffer from low coverage. On the other hand, automatically-acquired KBs typically have better coverage, but also contain more noise. Nonetheless, despite different construction methods, these KBs are all based on a simple subject-predicate-object model, which has major limitations in validity and expressiveness.

We recently presented ASCENT (Nguyen et al., 2021), a methodology for automatically collecting and consolidating commonsense assertions from the general web. To overcome the limitations of prior works, ASCENT refines subjects with subgroups (e.g., *circus elephant* and *domesticated elephant*) and aspects (e.g., *elephant tusk* and *elephant habitat*), and captures semantic facets of assertions (e.g., *lawyer*, *represents*, *clients*, *LOCATION: in courts*) or *(elephant, uses, its trunk, PURPOSE: to suck up water)*.

For a given concept, ASCENT searches through the web with pattern-based search queries disambiguated using WordNet (Miller, 1995) hypernymy. Then, irrelevant documents are filtered out based on similarity comparison against the corresponding Wikipedia articles. We then use a series of judicious dependency-parse-based rules to collect faceted assertions from the retained texts. The semantic facets, which come from prepositional phrases and supporting adverbs are then labeled by a supervised classifier. Finally, assertions are clustered using similarity scores from word2vec (Mikolov et al., 2013) and a fine-tuned RoBERTa (Liu et al., 2019) model.

¹<https://ascent.mpi-inf.mpg.de>

²https://youtu.be/qMkJXqu_Yd4

We executed the ASCENT pipeline for 10,000 prominent concepts (selected based on their respective number of assertions in ConceptNet) as primary subjects. In (Nguyen et al., 2021), we showed that the content of the resulting CSKB (hereinafter referred to as ASCENT KB) is a milestone in both salience and recall. As extrinsic evaluation, we conducted a comprehensive evaluation of the contribution of CSK to zero-shot question answering (QA) with pre-trained language models (Petroni et al., 2020; Guu et al., 2020).

This paper presents a companion web portal of the ASCENT KB, which enables the following interactions:

1. Exploration of the construction process of ASCENT, by inspecting word sense and Wikipedia disambiguation, web search queries, clustered statements, and source sentences and documents.
2. Inspection of the resulting KB, starting from subjects, predicates, objects, or examining specific subgroups or aspects.
3. Observation of the impact of structured knowledge on question answering with pretrained language models, comparing generated answers across various CSKBs and QA settings.

The web portal is available at <https://ascent.mpi-inf.mpg.de>, and a screencast demonstrating the system can be found at https://youtu.be/qMkJXqu_Yd4.

2 ASCENT

Two major contributions of ASCENT are its expressive knowledge model, and its state-of-the-art extraction methodology. Details are in the technical paper (Nguyen et al., 2021). In this section, we revisit the most important points.

2.1 Knowledge model

ASCENT extends the traditional triple-based data model in existing CSKBs in two ways.

Expressive subjects. Subjects in existing CSKBs are usually single nouns, which implies two shortcomings: (i) different meanings for the same word are conflated, and (ii) refinements and variants of word senses are missed out. ASCENT has addressed this problem with the following means:

1. When searching for source texts, ASCENT combines the target subject with an informative hypernym from WordNet to distinguish different senses of the word (e.g., “bus public transport” and “bus network topology” for the subject *bus*).
2. ASCENT refines subjects with multi-word phrases into *subgroups* and *aspects*. For example, subgroups for the subject *bus* would be *tourist bus* and *school bus*, while one of its aspects would be *bus driver*.

Semantic facets. The validity of commonsense assertions is usually non-binary (Zhang et al., 2017; Chalier et al., 2020), and depends on specific temporal and spatial circumstances (e.g., lions live for 10-14 years in the wild but for more than 15 years in captivity). Moreover, CSK triples often benefit from further context regarding causes/effects and instruments (e.g., elephants communicate with each other *by creating sounds*, beer is served *in bars*). In ASCENT’s knowledge model, such information is added to SPO triples via semantic facets. ASCENT distinguished 8 types of facets: *cause*, *manner*, *purpose*, *transitive-object*, *degree*, *location*, *temporal* and *other-quality*.

2.2 Extraction pipeline

ASCENT is a pipeline operating in three phases: *source discovery*, *knowledge extraction* and *knowledge consolidation*. Fig. 1 illustrates the architecture of the pipeline.

Source discovery. We utilize the Bing Web Search API to obtain documents specific to each subject, with search queries refined by the subject’s hypernyms in WordNet. We manually designed query templates for 35 prominent hypernyms (e.g., if subject s_0 has hypernym *animal.n.01*, we produce the search query “ s_0 animal facts”, similarly for the hypernym *professional.n.01*, the search query will be “ s_0 job descriptions”). We then compute the cosine similarity between the bag-of-words representations of each obtained document and a respective Wikipedia article to determine the relevance of the documents. Low-ranked documents will be omitted in further steps.

Knowledge extraction. The extractors take in the relevant documents and their outputs include: open information extraction (OIE) tuples, list of subgroups and list of aspects. To obtain *OIE tuples*, we extend the STUFFIE approach (Prasoj et al.,

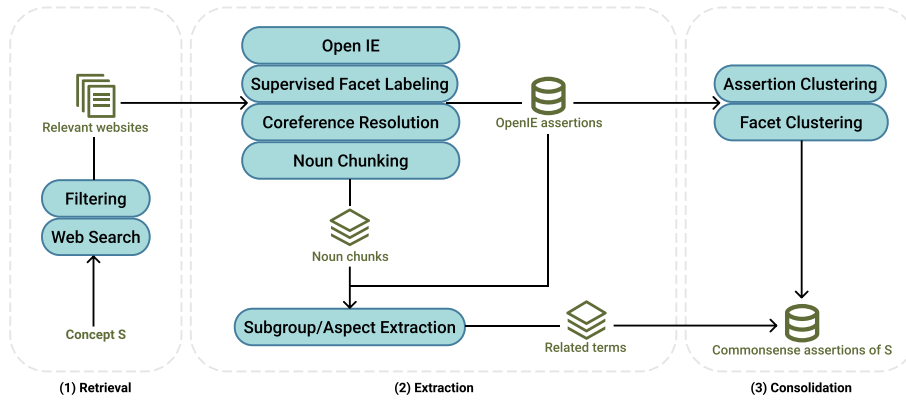


Figure 1: Architecture of the ASCENT extraction pipeline (Nguyen et al., 2021).

2018), a list of carefully crafted dependency-parse-based rules, to pull out faceted assertions from the texts. Then we classify each facet into one of the eight semantic labels using a fine-tuned RoBERTa model. For *subgroups*, noun phrases whose head word is the target subject are collected as candidates and then are clustered using the hierarchical agglomerative clustering (HAC) algorithm on average word2vec representations. Finally, we collect *aspects* from possessive noun chunks and SPO triples where P is either “have”, “contain”, “be assembled of” or “be composed of”.

Knowledge consolidation. We perform clustering on SPO triples and facet values. As *SPO triples*, we first filter triple-pair candidates with fast word2vec similarity. After that, advanced similarity of triple pairs computed by another fine-tuned RoBERTa model is fed to the HAC algorithm to group the triples into semantically similar clusters. For *facet values*, we group phrases with the same head words together (e.g., “during evening” and “in the evening”).

2.3 Web portal

The web portal (<https://ascent.mpi-inf.mpg.de>) is implemented in Python using Django, and hosted on an Nginx web server. The underlying structured CSK is stored in a PostgreSQL database, while for the QA part, statements of all CSKBs are indexed and queried via Apache Solr, for fast text-based querying. All components are deployed on a virtual machine with access to 4 virtual CPUs and 8 GB of RAM.

In the demonstration session, we show how users can interact with the portal for exploring the KB (Section 4.1), understanding the KB construction (Section 4.2), and observing its utility for question

answering (Section 4.3).

3 Commonsense QA setups

One common extrinsic use case of KBs is question answering. Recently, it was observed that priming language models (LMs) with relevant context can considerably benefit their performance in QA-like tasks (Petroni et al., 2020; Guu et al., 2020). In (Nguyen et al., 2021), to evaluate the contribution of structured CSK to QA, we conducted a comprehensive evaluation consisting of four different setups, all based on the above idea.

1. In *masked prediction* (MP), LMs are asked to predict single masked tokens in generic sentences.
2. In *free generation* (FG), LMs arbitrarily generate answer sentences to given questions.
3. *Guided generation* (GG) extends free generation by answer prefixes that prevent the LMs from evading answering.
4. *Span prediction* (SP) is the task of locating the answer of a question in provided context.

Examples of the QA setups can be seen in Table 1. Generally, given a question, our system will retrieve from CSKBs assertions relevant to it, and then use the assertions as additional context to guide the LMs. In the ASCENT demonstrator, we provide a web interface for experimenting with all of those QA setups with context retrieved from several popular CSKBs.

4 Demonstration experience

In the demonstration session, attendees will experience three main functionalities of our demonstration system.

Setup	Input	Sample output
MP	Elephants eat [MASK]. [SEP] Elephants eat roots, grasses, fruit, and bark, and they eat a lot of these things.	everything (15.52%), trees (15.32%), plants (11.26%)
FG	C: Elephants eat roots, grasses, fruit, and bark, and they eat... Q: What do elephants eat? A:	They eat a lot of grasses, fruits, and...
GG	C: Elephants eat roots, grasses, fruit, and bark, and they eat... Q: What do elephants eat? A: Elephants eat	Elephants eat a lot of things.
SP	question="What do elephants eat?" context="Elephants eat roots, grasses, fruit, and bark, and they eat..."	start=14, end=46, answer="roots, grasses, fruit, and bark"

Table 1: Examples of QA setups (Nguyen et al., 2021).

4.1 Exploring the ASCENT KB

Concept page. Suppose a user wants to know which knowledge ASCENT stores for *elephants*. They can enter the concept into the search field in the top right of the start page, and select the first result from the autocompletion list, or press enter, to arrive at the intended concept. The resulting website (see Fig. 2) is divided into three main areas.

At the top left, they can inspect an image from <https://pixabay.com>, the WordNet synset used for disambiguation, the Wikipedia page used for result filtering, and a list of alternative lemmas, if existing.

At the top right, users can see subgroups and related aspects, which in our knowledge representation model, can carry their own statements. This way, they can learn that the most salient aspects of *elephants* are their *trunks*, *tusks* and *ears*, or that *elephant trunks have more than 40,000 muscles*.

The body of the page, presents the assertions, organized into groups of same-predicate assertions. In each group, assertions are sorted by their frequency displayed beside their objects. For example, the most commonly mentioned *foods* of *elephants* are *grasses*, *fruits*, and *plants*. Many assertions come with a red asterisk. This indicates that the assertion comes with semantic facets. When clicking on an assertion, it will show a small box displaying an SVG-based visualisation of the assertion in which we illustrate all elements of the assertion: its subject, predicate, object, facet labels and values, frequency of the assertion as well as frequency of each facet. For example, one can see that the purpose of *elephants using their trunks* is *to suck up water*.

Searching and downloading assertions. Alternatively to exploring statements starting from a subject, users can start from a search functional-

ity under the *Browse* menu. This way, they can search, for instance, for all concepts that eat grass (*capybara*, *zebra*, *kangaroo*, ...).

The website also provides a JSON-formatted data dump (678MB) of all 8.9 million assertions extracted by the pipeline and their corresponding source sentences and documents. This dataset is also accessible via the HuggingFace Datasets package³.

4.2 Inspecting the construction of assertions

For many downstream use cases, it is important to know about the provenance of information.

Users can inspect general properties of the construction process by observing the WordNet lemma and the Wikipedia page used for filtering, as well as inspect specific statistics about the number of retained websites, sentences, and assertions, in a panel at the bottom of subject pages (e.g., 435 websites were retained for *elephant*, from which 50k OpenIE assertions could be extracted).

Furthermore, users can look deeply into the construction process of each assertion on its own dedicated page, which displays the following:

1. *Clustered triples*: These are triples that were grouped together in the knowledge consolidation phase (cf. Section 2.2), where the most frequent triple was selected as cluster representative. For example, for the assertion $\langle \textit{lion}, \textit{eat}, \textit{zebra}, \textit{DEGREE: mostly} \rangle$ (14), the cluster contains: $\langle \textit{lion}, \textit{eat}, \textit{zebra} \rangle$ (9), $\langle \textit{lion}, \textit{prey on}, \textit{zebra} \rangle$ (2), $\langle \textit{lion}, \textit{feed on}, \textit{zebra} \rangle$ (1), $\langle \textit{lion}, \textit{feed upon}, \textit{zebra} \rangle$ (1), $\langle \textit{lion}, \textit{prey upon}, \textit{zebra} \rangle$ (1). The numbers in parentheses indicate their corresponding frequency.
2. *Facets*: The assertion’s facets are presented in a table whose columns are facet value, facet type and clustered facets. The frequency of each clustered facet is also indicated.
3. *Source sentences and documents*: Finally, we exhibit the sentences from which the assertions were extracted and their parent documents (in the form of URLs). Furthermore, in the extraction phase, we also recorded the position of assertion elements (i.e., subject, predicate, object, facet) in the source sentences.

³https://huggingface.co/datasets/ascend_kb

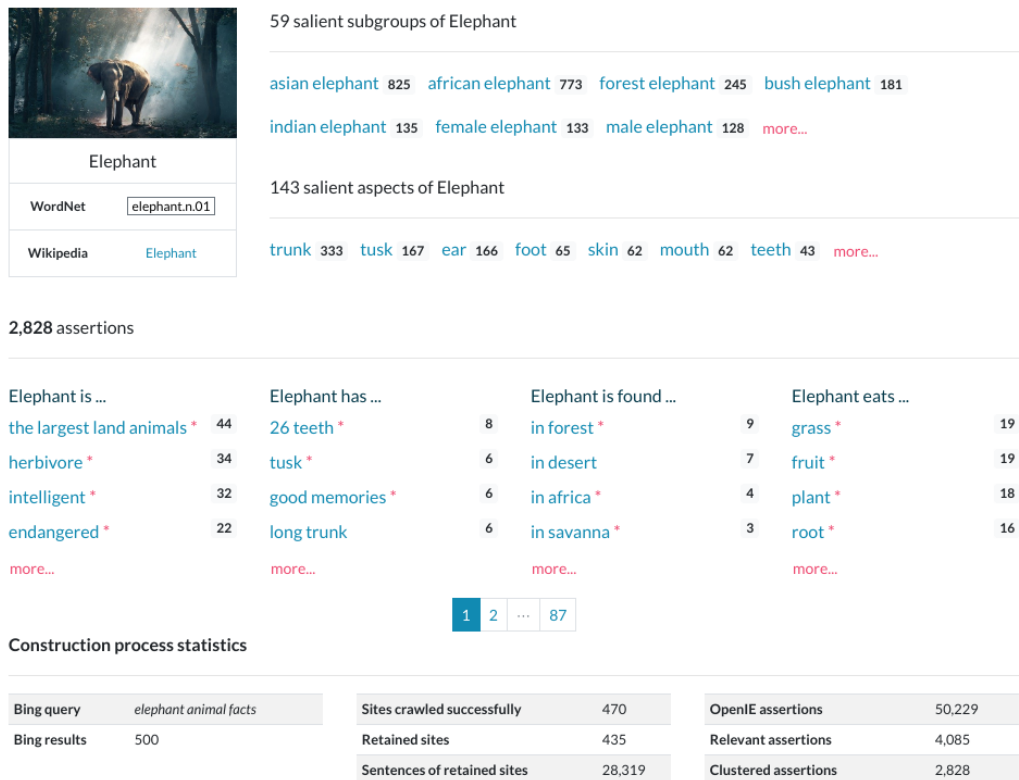


Figure 2: Example of ASCENT’s page for the concept *elephant*.

We show that information to users by highlighting each kind of element with a different color in the source sentences.

4.3 Experimenting with commonsense QA

The third functionality experienced in the demo session is the utilization of commonsense knowledge for question answering (QA).

Input. There are four main parts in the input interface for the QA experiment:

1. *QA setup*: The user chooses one QA setup they want to experiment with. Available are *Masked Prediction*, *Span Prediction* and *Free/Guided Generation*. If *Masked Prediction* is selected, the user can choose how many answers the LM should produce. For the *Generation* settings, users can provide an answer prefix to avoid overly evasive answers.
2. *Input query*: The user enters the text question as input. The question can be in the form of a masked sentence (in the case of *Masked Prediction*), or a standard natural-language question (in other setups).
3. *Retrieval options*: The user can select one supported retrieval method and the number of

assertions to be retrieved per CSKB for each question.

4. *Context sources*: The user selects the sources of context (i.e., “no context”, CSKBs and “custom context”). If a CSKB is selected, the system will retrieve from that KB assertions relevant to the given input question. If “custom context” is selected, user must then enter their own content. The “no context” option is available for all setups but *Span Prediction*.

Output. The QA system presents its output in the form of a table which has three columns: *Source*, *Answer(s)* and *Context*. For *Masked Prediction* and *Span Prediction*, answers are printed with their respective confidence scores, meanwhile for *Free/Guided Generation*, only answers are printed. For *Span Prediction* in which answers come directly from given contexts, we also highlight the answers in the contexts.

An example of the QA demo’s output for the question “*What do rabbits eat?*” under the *Free Generation* setting can be seen in Fig. 3. One can observe that language models’ predictions are heavily influenced by given contexts. Without context, GPT-2 is only able to generate an evasive answer.

When being given context, it tends to re-generate the first sentence in the context first, (e.g., see the answers aligning with ASCENT, TupleKB and ConceptNet in Fig. 3). For the context retrieved from Quasimodo, GPT-2 is able to overlook the erroneous first sentence, however its generated answer is rather elusive despite the fact that subsequent statements in the context all contain direct answers to the question.

The question “*Bartenders work in [MASK].*” under the *Masked Prediction* setting is another example for the influence of context on LMs’ output. Since *bartender* is a subject well covered by the ASCENT KB, the assertions pulled out are all relevant (i.e., *Bartenders work in bar. Bartenders work in restaurant. . .*) which help guide the LM to a good answer (*bar*). Meanwhile, because this subject is not present in TupleKB, its retrieved statements are rather unrelated (*Work capitals have firm. Work experiences include statement. . .*). Given that, the top-1 prediction for this KB was *tandem* which is obviously an evasive answer.

5 Related work

CSKB construction. Cyc (Lenat, 1995) is the first attempt to build a large-scale commonsense knowledge base. Since then, there have been a number of other CSKB construction projects, notably ConceptNet (Speer and Havasi, 2012), WebChild (Tandon et al., 2014, 2017), TupleKB (Mishra et al., 2017), and more recently Quasimodo (Romero et al., 2019), Dice (Chalier et al., 2020), Atomic (Sap et al., 2019), and CSKG (Ilievski et al., 2020). The early approach to building a CSKB is based on human annotation (e.g., Cyc with expert annotation and ConceptNet with crowdsourcing annotation). Later projects tend to use automated methods based on open information extraction to collect CSK from texts (e.g., WebChild, TupleKB and Quasimodo). Lately, CSKG is an attempt to combine various commonsense knowledge resources into a single KB. The common thread of these CSKB is that they are all based on SPO triples as knowledge representation, which has shortcomings (Nguyen et al., 2021). ASCENT is the first attempt to build a large-scale CSKB with assertions equipped with semantic facets built upon the ideas of semantic role labeling (Palmer et al., 2010).

KB visualization. Most CSKBs share their con-

tent via CSV files. Some, like ConceptNet⁴, WebChild⁵, Atomic⁶ and Quasimodo⁷, have a web portal to visualise their assertions. The most common way for CSKB visualisation is to use a single page for each subject and group assertions by predicate (e.g., in ConceptNet and WebChild). Quasimodo, on the other hand, implements a simple search interface to filter assertions and presents assertions in a tabular way (Romero and Razniewski, 2020). The ASCENT demo has both functionalities: exhibiting assertions of each concept in a separated page, and supporting assertion filtering. Our demo also uses an SVG-based visualisation of assertions with semantic facets, which are a distinctive feature of the ASCENT knowledge model.

Context in LM-based question answering.

Priming large pretrained LMs with context in QA-like tasks is a relatively new line of research (Petroni et al., 2020; Guu et al., 2020). In our original paper, we made the first attempt to evaluate the contribution of CSKB assertions to QA via four different setups based on that idea. While others use commonsense knowledge for (re-)training language models (Hwang et al., 2021; Ilievski et al., 2021; Ma et al., 2021; Mitra et al., 2020), to the best of our knowledge, our demo system is the first to visualize the effect of priming vanilla language models, i.e., without task-specific retraining.

6 Conclusion

We presented a web portal for a state-of-the-art commonsense knowledge base—the ASCENT KB. It allows users to fully explore and search the CSKB, inspect the construction process of each assertion, and observe the impact of structured CSKBs on different QA tasks. We hope that the portal enables interesting interactions with the ASCENT methodology, and that the QA demo allows researchers to explore the potentials of combining structured data with pre-trained language models.

References

Tom B Brown et al. 2020. Language models are few-shot learners. In *NeurIPS*.

⁴<http://conceptnet.io>

⁵<https://gate.d5.mpi-inf.mpg.de/webchild>

⁶<https://mosaickg.apps.allenai.org/kg-atomic2020>

⁷<https://quasimodo.r2.enst.fr>

Output

Retriever: BM25 Retriever

Predictor: Free Generator gpt2

Source	Answer(s)	Context
No context	They are very good at eating rabbits.	
ASCENT ^{triple}	Rabbits eat grass.	Rabbits eat grass. Rabbits eat plant. Rabbits eat vegetable. Rabbits eat food. Rabbits eat weed.
Quasimodo	Rabbits eat the same things as humans.	Eats can rabbit. Rabbits eat wires. Rabbits eat roses. Rabbits eat carpet. Rabbits eat mice.
TupleKB	Rabbits eat rabbit.	Rabbits eat rabbit. Rabbits eat food. Rabbits eat carrot. Rabbits eat leaf. Rabbits eat lettuce.
ConceptNet	Rabbits are good to eat.	Rabbits are good to eat. Eat vegetables need rabbit. Eating vegetables have pretend rabbit. Rabbits are cute. Rabbits are fast.

Figure 3: Free Generation output for question: “What do rabbits eat?”.

- Yohan Chalier, Simon Razniewski, and Gerhard Weikum. 2020. Joint reasoning for multi-faceted commonsense knowledge. In *AKBC*.
- Peter Clark et al. 2020. From ‘F’ to ‘A’ on the NY regents science exams: An overview of the Aristo project. *AI Magazine*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.
- Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: Retrieval-augmented language model pre-training. In *ICML*.
- Jena D Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. Comet-atomic 2020: On symbolic and neural commonsense knowledge graphs. In *AAAI*.
- Filip Ilievski, Alessandro Oltramari, Kaixin Ma, Bin Zhang, Deborah L McGuinness, and Pedro Szekely. 2021. Dimensions of commonsense knowledge. *arXiv preprint arXiv:2101.04640*.
- Filip Ilievski, Pedro Szekely, and Bin Zhang. 2020. Cskg: The commonsense knowledge graph. *arXiv preprint arXiv:2012.11490*.
- Douglas B Lenat. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Kaixin Ma, Filip Ilievski, Jonathan Francis, Yonatan Bisk, Eric Nyberg, and Alessandro Oltramari. 2021. Knowledge-driven data construction for zero-shot evaluation in commonsense question answering. In *AAAI*.
- John McCarthy. 1960. *Programs with common sense*. RLE and MIT computation center.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *ICLR*.
- George A Miller. 1995. Wordnet: a lexical database for English. *Communications of the ACM*.
- Bhavana Dalvi Mishra, Niket Tandon, and Peter Clark. 2017. Domain-targeted, high precision knowledge extraction. *TACL*.
- Arindam Mitra, Pratyay Banerjee, Kuntal Kumar Pal, Swaroop Mishra, and Chitta Baral. 2020. How additional knowledge can improve natural language commonsense question answering? *arXiv preprint arXiv:1909.08855*.
- Don Monroe. 2020. Seeking artificial common sense. *Communications of the ACM*.
- Tuan-Phong Nguyen, Simon Razniewski, and Gerhard Weikum. 2021. Advanced semantics for commonsense knowledge extraction. In *WWW*.
- Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*.
- Fabio Petroni, Patrick Lewis, Aleksandra Piktus, Tim Rocktäschel, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2020. How context affects language models’ factual predictions. In *AKBC*.
- Radityo Eko Prasajo, Mouna Kacimi, and Werner Nutt. 2018. Stuffie: Semantic tagging of unlabeled facets using fine-grained information extraction. In *CIKM*.
- Julien Romero and Simon Razniewski. 2020. Inside quasimodo: Exploring construction and usage of commonsense knowledge. In *CIKM*.

- Julien Romero, Simon Razniewski, Koninika Pal, Jeff Z. Pan, Archit Sakhadeo, and Gerhard Weikum. 2019. Commonsense properties from query logs and question answering forums. In *CIKM*.
- Maarten Sap et al. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *AAAI*.
- Robyn Speer and Catherine Havasi. 2012. Conceptnet 5: A large semantic network for relational knowledge. *Theory and Applications of Natural Language Processing*.
- Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. 2019. olympics - on what language model pre-training captures. *TACL*.
- Niket Tandon, Gerard de Melo, Fabian M. Suchanek, and Gerhard Weikum. 2014. Webchild: harvesting and organizing commonsense knowledge from the web. In *WSDM*.
- Niket Tandon, Gerard de Melo, and Gerhard Weikum. 2017. Webchild 2.0 : Fine-grained commonsense knowledge distillation. In *ACL*.
- Sheng Zhang, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme. 2017. Ordinal common-sense inference. *TACL*.