Preface - SMART 2021

SMART 2021 [1] was the second edition of the SeMantic Answer Type and Relation Prediction Task (SMART), which part of the ISWC 2021 Semantic Web Challenge. It was co-located with the 19th International Semantic Web Conference (ISWC 2020)[1]. The first edition SMART2020 [2] was in ISWC 2020. Given a question in natural language, the task of SMART challenge is, to predict the answer type and relations using a target ontology. The challenge had 2 tracks (answer type prediction and relation prediction) with 2 KBs, one using the DBpedia ontology and the other using Wikidata ontology. There were six submissions for answer type prediction (DBpedia) and four systems for answer type prediction (Wikidata). Similarly, there were three systems for relation prediction (DBpedia) and three systems for relation prediction (Wikidata). This volume contains peer-reviewed system description papers of all the systems that participated in the challenge. More details about the challenge can be found at `https://smart-task.github.io/2021/`.

Challenge Description

This challenge is focused on answer type prediction and relation prediction, which play an important role in Question Answering systems.

Answer Type Prediction Given a natural language question, the task is to produce a ranked list of answer types of a given target ontology. Previous such answer type classifications in literature are performed as a short-text classification task using a set of coarse-grained types, for instance, either six types [3, 4, 5, 6] or 50 types [7] with TREC QA task[2]. We propose a more granular answer type classification using popular Semantic Web ontologies such as *DBpedia* and *Wikidata*.

Relation Prediction Given a natural language question, the task is to identify the relation and link to the relations in KG. Depending on the number of relations in the KG, the number of relation types to be linked varies.

Table 1 and Table 2 illustrates some examples. The participating systems can be either supervised (training data is provided) or unsupervised. The systems can utilise a wide range of approaches; from rule-based to neural approaches.

Presentations

Eight teams competed in SMART 2021 and presented their systems at the ISWC 2021 conference. Table 3 shows their presentation titles along with the authors.

[1] `https://iswc2021.semanticweb.org/`
[2] `https://trec.nist.gov/data/qamain.html`

Table 1: Example questions and answer types.

Question	Answer Type	
	DBpedia	Wikidata
Give me all actors starring in movies directed by and starring William Shatner.	dbo:Actor	wd:Q33999
Which programming languages were influenced by Perl?	dbo:ProgrammingLanguage	wd:Q9143
Who is the heaviest player of the Chicago Bulls?	dbo:BasketballPlayer	wd:Q3665646
How many employees does Google have?	xsd:integer	xsd:integer

Table 2: Example questions and relation types.

Question	Relation Type	
	DBpedia	Wikidata
Which languages were influenced by Perl?	dbo:influencedBy	wdt:P737
Give me all actors starring in movies directed by and starring William Shatner.	dbo:starring, dbo:director	wdt:P161, wdt:P57
How many employees does IBM have?	dbo:numberOfEmployees	wdt:P1128

Leaderboards

For each natural language question in the test set, the participating systems are expected to provide two predictions: answer category and answer type. Answer category can be either 'resource', 'literal' or 'boolean'. If the answer category is 'resource', the answer type should be an ontology class (DBpedia or Wikidata, depending on the dataset). The systems could predict a ranked list of classes from the corresponding ontology. If the answer category is 'literal', the answer type can be either 'number', 'date' or 'string'.

Answer Type Prediction

DBpedia Dataset

Category prediction will be considered as a multi-class classification problem and accuracy score will be used as the metric. As DBpedia follows DBpedia ontology for its classes, thus for type predication, we will use the metric lenient NDCG@k with a linear decay, adopted from Balog & Neumayer [8]. The results are shown in Table 3.

Slot	Title / Authors
Session 3C – SW Challenges: Thursday, 26th October, 2021	
13:26 – 13:33 EDT	**Reaching out for the Answer: Answer Type and Property Prediction** *Khaoula Benmaarouf, Kanchan Shivashankar,* and Nadine Steinmetz
13:33 – 13:40 EDT	**The Combination of BERT and Data Oversampling for Answer Type Prediction** *Thang Ta Hoang, Olumide Ebenezer Ojo, Olaronke Oluwayemisi Adebanji, Alexander Gelbukh* and Hiram Calvo.
	Q & A session for the first two talks
13:50 – 13:55 EDT	**CitySAT: A system for the semantic answer type prediction** *Chaeyoon Kim and Ernesto Jimenez-Ruiz*
13:55 – 14:00 EDT	**Semantic Answer Type Prediction** *G P Shrivatsa Bhargav, Dinesh Khandelwal,* Saswati Dana and Dinesh Garg
14:00 – 14:05 EDT	**Answer Type Prediction (SMART 2021 – AT)** *Xiao Ning, Ammar Ammar, Arif Yilmaz* Shervin Mehryar, Remzi Celebi
14:05 – 14:10 EDT	**Multilingual Hierarchical Expected Answer)** Type Classification over DBpedia and Wikidata *Aleksandr Perevalov and Andreas Both*
	Q & A session for the last four talks

Table 3: Presentation Schedule for the Participating Systems

Wikidata Dataset

Here again the category prediction will be considered as a multi-class classification problem and accuracy score will be used as the metric. Wikidata does not follow a strict ontology for the classes, it has a very large and rather flat set of classes and subclasses. Thus for type prediction, we use a mean reciprocal rank (MRR) based scoring system [9], where the expected type prediction is a list. The results are shown in Table 4.

Relation Prediction

Relation prediction is evaluated using the precision, recall and F1 metrics considering the gold standard list relations and predicted list of relations. Table 5 and Table 6 shows the results for the relation prediction task for DBpedia and Wikidata.

System	Accuracy	NDCG@5	NDCG@10
Kim et al.	0.984	0.842	0.854
Bhargav et al.	0.985	0.825	0.79
Celebi et al.	0.985	0.725	0.704
Hoang et al.	0.985	0.727	0.664
Steinmetz et al.	0.991	0.734	0.658
Perevalov et al.	0.991	0.643	0.577

Table 4: Leader-board Task1: Answer Type Prediction for DBpedia dataset

System	Accuracy	MRR
Hoang et al.	0.98	0.7
Celebi et al.	0.98	0.66
Steinmetz et al.	0.99	0.45
Perevalov et al.	0.98	0.43

Table 5: Leader-board Task1: Answer Type Prediction for Wikidata dataset

System	Precision	Recall	F1
Steinmetz et al.	0.86	0.88	0.86
Hoang et al.	0.83	0.82	0.83
Baselines Falcon	0.43	0.36	0.31

Table 6: Leader-board Task2: Relation Prediction for DBpedia dataset

System	Precision	Recall	F1
Steinmetz et al.	0.75	0.82	0.76
Hoang et al.	0.62	0.61	0.61
Baselines Falcon	0.43	0.36	0.31

Table 7: Leader-board Task2: Relation Prediction for Wikidata dataset

Organisation

In this section, we list the people who organised and contributed to the success of this event.

Challenge Chairs

- Nandana Mihindukulasooriya (IBM Research AI)
- Mohnish Dubey (InfAI Dresden)
- Alfio Gliozzo (IBM Research AI)

- Jens Lehmann (University of Bonn and Fraunhofer IAIS)
- Axel-Cyrille Ngonga Ngomo (Universität Paderborn)
- Ricardo Usbeck (Fraunhofer IAIS Dresden)
- Gaetano Rossiello (IBM Research AI)
- Uttam Kumar (University of Bonn)

Challenge Programme Committee Members

The challenge programme committee helped to peer-review the eight system papers. Each paper received 2 or 3 reviews from the programme committee members and authors took those feedback into account when preparing the camera-ready versions. The organisers would like to thank them for their valuable time.

- Ibrahim Abdelaziz (IBM Research AI)
- Carlos Badenes-Olmedo (Ontology Engineering Group, UPM)
- Pavan Kapanipathi (IBM Research AI)
- Debanjali Biswas (GESIS)
- Pablo Calleja (Ontology Engineering Group, UPM)
- Jennifer D'Souza (TIB, Leibniz University Hannover)
- Uttam Kumar (University of Bonn)
- Gaetano Rossiello (IBM Research AI)
- Sanju Tiwari (Universidad Autonoma de Tamaulipas)
- Ricardo Usbeck (University of Hamburg)
- Daniel Vollmers (Paderborn University)
- Mohnish Dubey (InfAI Dresden)
- Nandana Mihindukulasooriya (IBM Research AI)

Acknowledgements

We would like to thank the ISWC Semantic Web Challenge chairs, Ernesto Jimenez-Ruiz, Jiaoyan Chen and Despoina Magka, and the whole ISWC organising committee for their invaluable support to make this event a success. We would also like to thank the challenge participants for their interest, quality of work, and informative presentations during the event which made it attractive to the ISWC audience.

References

[1] Nandana Mihindukulasooriya, Mohnish Dubey, Alfio Gliozzo, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, Ricardo Usbeck, Gaetano Rossiello, and Uttam Kumar. Semantic Answer Type and Relation Prediction Task (SMART 2021). *CoRR/arXiv*, abs/2112.07606, 2022.

[2] Nandana Mihindukulasooriya, Mohnish Dubey, Alfio Gliozzo, Jens Lehmann, Axel-Cyrille Ngonga Ngomo, and Ricardo Usbeck. SeMantic AnsweR Type prediction task (SMART) at ISWC 2020 Semantic Web Challenge. *CoRR/arXiv*, abs/2012.00555, 2020.

[3] Han Zhao, Zhengdong Lu, and Pascal Poupart. Self-adaptive hierarchical sentence model. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[4] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630*, 2015.

[5] Yoon Kim. Convolutional neural networks for sentence classification. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014.

[6] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[7] Xin Li and Dan Roth. Learning Question Classifiers: the Role of Semantic Information. *Natural Language Engineering*, 12(3):229–249, 2006.

[8] Krisztian Balog and Robert Neumayer. Hierarchical target type identification for entity-oriented queries. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*, pages 2391–2394. ACM, 2012.

[9] Dragomir R. Radev, Hong Qi, Harris Wu, and Weiguo Fan. Evaluating web-based question answering systems. In *Proceedings of the Third International Conference on Language Resources and Evaluation, LREC 2002, May 29-31, 2002, Las Palmas, Canary Islands, Spain*. European Language Resources Association, 2002.