Survey of Query Intent Detection

Pankaj Singh, Pushpak Bhattacharyya
Indian Institute of Technology Bombay
pankaj.min1247@gmail.com

Abstract

With the growing popularity of the World Wide Web, English is no longer the dominant language of the documents present on the internet. The documents present in Indian Languages and their user base have also grown significantly. It is crucial to build effective multilingual Information Retrieval (IR) systems with competitive performance as their English counterparts. To achieve this, a robust Query Understanding pipeline plays a key role. It involves extracting key pieces of information from the query string before propagating it through IR systems. This information includes the domain, intent, and key entities of the user query. This paper discusses the latest advancements and past work in the field of query intent detection.

1 Introduction

The aim of our query understanding system is, given a multilingual search query, identify the domain, intent, and entities in it. We first develop our entire QU system for Hindi and English and later extend the intent detection system for four more languages. The first system should be capable of handling queries in English, Hindi, and Code/Script Mixed Hindi-English. This system comprises three modules to identify the required information. These three modules are explained with a sample example as follows:

Domain detection: Given a users search query, detect the domain to which that query belongs. This will be a binary classifier that has two labels as in-domain and out-of-domain. A separate classifier will be required for each domain which will filter the in-domain queries from out-of-domain queries. In Figure 1, for the search query- “Buy latest nike shoes online”, the system will detect this query as in-domain given that the classifier is built for the e-commerce domain.

Intent detection: This module will try to capture the goal of the user which he/she wants to achieve via the query. Depending upon the domain, we have to develop the corresponding intent taxonomy. In the above example, the intent can be categorized as “New arrival" as the user is trying to look for the latest product in the e-commerce domain.

Entity Extraction: This module will extract key entities present in the search query. In the above example, “nike” and “shoes” are key entities present in the query.

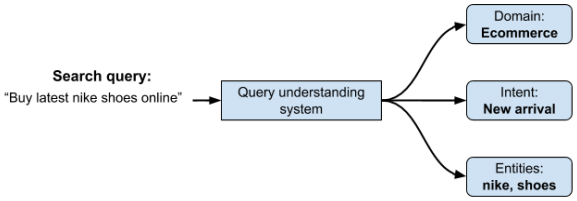

Figure 1: Overview of an sample QU system

Query intent detection is the task of determining the underlying goal user wants to achieve from a text query. Intent detection plays an important role in providing a satisfying user experience. This is an intricate task as users can ask the same thing in many different ways. Poor query formulation by a user will also makes the intent detection task challenging. Along with search engines, queries are the main mode of interaction for various applications like chat-bots, smart agents, etc. We detect the intent of the user from the text query and incorporate this information in our rank scoring function to get the more relevant search results. In Figure 3, we can how intent detection can be used to obtain better-ranked results.

2 Related Work

Query understanding has been an important topic of research for the last several years. Query intent

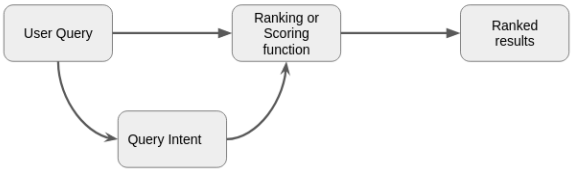

Figure 2: Incorporating query intent detection to get better search results.

detection is a crucial component of query understanding which has been researched extensively. Broder (2002) proposed a query intent taxonomy for web users which has been widely used by many researchers. According to Broder, search queries can have three types of user intents: (1) navigational (the user wants to reach a particular website), (2) informational (the user wants to find a piece of information on the Web), and (3) transactional (the user wants to perform a web-mediated task). There are many approaches that have been proposed for the detection of query intent. One of these approaches involves using various query log information such as users click-through data, anchor-text, anchor-link distribution, etc. (Lee et al., 2005; Brenes et al., 2009) . Another approach involves using query text to predict the intent of the query using text classification systems. ATIS (Hemphill et al., 1990) and SNIPS (Coucke et al., 2018) are two very popular open domain datasets for benchmarking the query intent detection systems. Wang et al. (2018) proposed a bimodal-based semantic phase parsing system which is SOTA for intent detection in the ATIS dataset. Qin et al. (2019) proposed stack propagation with BERT framework and achieved state of art in SNIP dataset. Having a well-defined taxonomy is a challenging part of query intent detection. As mentioned, the most widely used taxonomy is by Broader et al. but it has only three categories of intent. Many other detailed taxonomies have been proposed for domain-specific intent detection. Kumar and Politi (2019) proposed intent taxonomy for the legal domain having 4 classes. Hashemi (2016) proposed a detailed 2 level intent taxonomy having 125 unique intent classes. We propose a much finer taxonomy with 3 levels of intents and 138 unique intent classes.

3 Multilevel Intent Taxonomy

Developing an exhaustive intent taxonomy for a given domain is an important part of any intent detection system. We developed intent taxonomy for the entertainment domain after extensive research of newspapers, magazines, online media, etc. Our taxonomy was finalized after getting elaborated feedback from domain and industry experts. The intent taxonomy has three levels, level 1 intent is broad level or major intent consisting of 8 intent categories. The deeper we go into taxonomy level (from level 1 to level 3) the specificity of the intent increases. Table 1 shows the level 1 intent categories in our Entertainment domain intent taxonomy with one example query.

Level 1 Intent	Example Search Query
Movie	bahubali film watch online
Music	old songs list
TV/web-series	latest episode of suits
Social Media	baba ka dhaba viral
Celebrity	salman khan home
Books/Literature	premchand novels
Fashion	london fashion week dates
Others	PS5 india launch

Table 1: Level 1 intent categories with example queries

Level 2 is further categorization of level 1 intent classes. Some of the level 2 intents are further granularized and level 3 intent is defined for them. So overall a query can have one, two, or three levels of intents depending upon its nature. In level 2 we have 54 intent categories and level 3 has 71 intent categories. A total of 138 unique intent categories all possible on combining all three levels of the taxonomy. This taxonomy will form the basis of our intent detection system and also assist us in collecting more exhaustive search queries to create a quality dataset. In appendix A we have added all three levels of the taxonomy tree for one of the nodes of level 1 with example queries.

Having a detailed multilevel taxonomy helps to pinpoint the user query intent. Our proposed three-level taxonomy has a very detailed intent classification for queries. In Table 2, we show details of taxonomies for popular intent classification datasets

4 QID using Deep Bi-directional LSTM Network (Sreelakshmi et al., 2018)

Various machine learning algorithms have been applied to get intent from text queries provided by the user. Sreelakshmi et al. (2018) proposed a deep learning-based framework using Bi-Directional

Intent Dataset	Taxonomy Levels	Unique Intents
ATIS	1	17
SNIPS	1	7
Hashemi et al.	2	125
Ours	**3**	**138**

Table 2: Intent taxnomy details of various datasets

Long Short-Term Memory (BLSTM) for intent identification. They use semantically enrich Glove word embeddings to ensure semantic correctness of the embeddings. Glove word embeddings place words with similar context closer in embedding space, but this also puts words like antonyms closer in embedding space, which generally appear in the same context but are not semantically close. The enrichment process deals with this problem by using **Synonyms, Antonyms, Related words, Hypernyms, and Hyponyms** present in vocabulary to enrich the Glove word embeddings.

After obtaining semantically rich word embeddings, they performed a sequence of experiments and explored various setups of deep learning models, and concluded that the Bi-directional LSTM with Enriched word embedding performed best. Figure 3 shows the basic architecture proposed method.

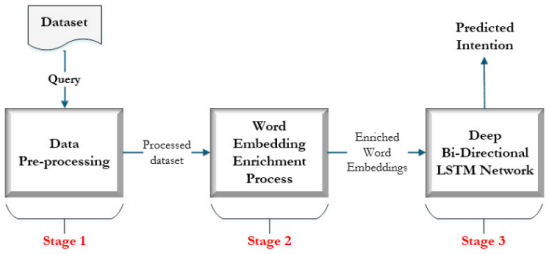

Figure 3: Basic Architecture proposed by Sreelakshmi et al., 2018

These experiments were performed on Airline Travel Information System (ATIS) dataset, which is very popular for intent detection system evaluation. This dataset has 4978 training queries, 899 testing queries, and 22 intents. Table 3 summarizes the results obtained by Sreelakshmi et al., 2018.

The proposed approach is able to capture higher-order non-linear features and can handle non-consecutive dependencies. Due to the Bi-directional model, more context is included while generating the final features representation of a query. The word embedding enrichment process improves the semantic correctness of the word vectors. In future work, authors propose to include external knowledge bases and use of more sophisticated deep learning models like GRUs and LSTMs with memory networks to further improve the performance of the intent detection system.

Model	Accuracy
LSTM + GloVe	93.461
BLSTM + GloVe	94.368
Deep LSTM + GloVe	95.873
Deep BLSTM + GloVe	97.562
LSTM + Enriched Vectors	95.081
BLSTM + Enriched Vectors	96.487
Deep LSTM + Enriched Vectors	96.918
Deep BLSTM + Enriched Vectors	**98.221**

Table 3: Performance comparison in terms of accuracy

5 Shareable Representation for Search Query Understanding (Kumar et al., 2020)

BERT model has shown that transformer-based architectures trained on language modeling tasks have outstanding performance in various NLP tasks. Kumar et al. (2020) proposed adapting transformer encoder architectures with language model pre-training to learn the intent of search queries. By training the model on a domain-specific task, they build a model with a shareable internal representation for query understanding tasks. This sharing will allow for fewer models needed at inference time for multiple tasks, and new task-specific models can be easily built by re-utilizing existing model representations.

They first initialize the BERT model with pre-trained weights obtained from language modeling tasks and then perform domain-specific training to fine-tune the BERT model to adjust to the new input-output format. Their ultimate task was to build an intent classifier for *Amazon* e-commerce search engine queries. The domain-specific training is done via a product classification task for over 1 Billion search queries. This is an extreme multi-label classification on tens of thousands of product categories. This domain-specific training also helps to boost the performance of the downstream task i.e., intent classification. After this

domain-specific training, they build a binary classifier for each of the intent. They considered three types of intents which are:

- **Help Intent:** 2,511,997 queries
- **Adult Intent:** 2,903,319 queries
- **Low Average Selling Price Intent:** 3,006,440 queries

The binary classification was used instead of joint classification for this task due to its scalability from an operation standpoint. Furthermore, if the new intent category is to be added in class labels in the future, then the whole model will need retraining and error due to this will also regress to the performance of the existing classifier. However, in the case of binary classifiers, we can easily scale the model by training a new classifier, which will have most of the layers shared with the existing classifier, and we will need to train on few last layers.

The right training strategy is very important to squeeze the best performance out of a deep learning model. Different layers of the model can learn different aspects of a particular domain. Shallow layers learn general features, and a deeper layer learns more complex domain-specific representation. Instead of fine-tuning the entire model parameters, tuning only the last few layers gives a good performance. This idea of transfer learning is very popular in the Computer Vision community and has been exploited very effectively. BERT was shown to have similar properties and some layers could be shown to learn language features, and some layers learned task-specific information. Figure 4 shows the architecture of $\text{BERT}_{\text{BASE}}$ and training strategy to make the last few layers trainable and freeze the rest layers to make them shareable across tasks.

It was observed that the performance of the model was better in intent classification tasks when the model was initialized with domain-specific trained weight rather than BERT pre-trained weights. The number of layers that need to unfreeze depends upon the task and how much it is different from the domain-specific task. In general, freezing the initial 10 layers gave competitive performance across all three intent classifiers, so these 10 layers are suitable candidates for being shareable across multiple tasks. In Table 4 we can see that sharing the layers significantly drops

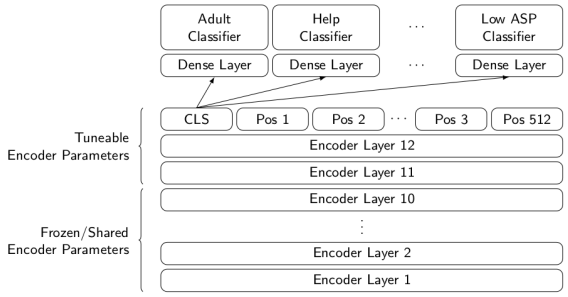

Figure 4: $\text{BERT}_{\text{BASE}}$ model architecture with frozen and trainable layers

the number of parameters to be trained for each classifier and will be crucial for full filling memory requirement if the model is being used in an online service as having multiple models with a large number of parameters to be stored online will bloat the system.

Model	Pre-Train	Frozen L.	Train L.	Train Params	Help	Adult	Low ASP
DNN + n-grams	Random	0	1	11,167,234	0.905	0.905	0.751
Trans Enc (1)	Random	0	1	30,926,594	0.926	0.928	0.759
Trans Enc (12)[a]	BERT	0	12	108,893,186	0.940	0.940	0.771
Trans Enc (12)	DST[b]	0	12	108,893,186	**0.945**	0.943	0.783
Trans Enc (12)	BERT	10	2	14,177,282	0.938	0.933	0.771
Trans Enc (12)	DST	10	2	14,177,282	0.938	**0.944**	0.786
Trans Enc (12)	BERT	11	1	7,089,410	0.937	0.925	0.763
Trans Enc (12)	DST	11	1	7,089,410	0.939	0.938	**0.788**
Trans Enc (12)	BERT	12	0	1,538	0.894	0.843	0.675
Trans Enc (12)	DST	12	0	1,538	0.928	0.924	0.756

Table 4: Performance comparison in terms of accuracy

6 Conclusion and Future Work

We described the past and recent approaches for query intent detection in detail. Literature shows us the need for detailed taxonomy for finer information extraction. We created a language-agnostic, multi-level, exhaustive entertainment domain intent taxonomy with validation from domain experts. We devised strategies for efficient data collection and annotation to build and evaluate our multilingual query understanding pipeline. Our proposed modules have shown remarkable performance on query understanding tasks such as domain detection, intent detection, and entity extraction. We extended our intent detection system for Marathi, Bengali, Tamil, and Telugu languages.

References

David J. Brenes, Daniel Gayo-Avello, and Kilian Pérez-González. 2009. Survey and evaluation

of query intent detection methods. In *Proceedings of the 2009 Workshop on Web Search Click Data*, WSCD '09, page 17, New York, NY, USA. Association for Computing Machinery.

Andrei Broder. 2002. A taxonomy of web search. *SIGIR Forum*, 36:3–10.

Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, Maël Primet, and Joseph Dureau. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *CoRR*, abs/1805.10190.

J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*.

Dan Geiger and Christopher Meek. 2005. Structured variational inference procedures and their realizations (as incol). In *Proceedings of Tenth International Workshop on Artificial Intelligence and Statistics,* The Barbados. The Society for Artificial Intelligence and Statistics.

Joshua Goodman. 2001a. A bit of progress in language modeling. *CoRR*, cs.CL/0108005v1.

Joshua T. Goodman. 2001b. A bit of progress in language modeling. *Computer Speech & Language*, 15(4):403–434.

H. B. Hashemi. 2016. Query intent detection using convolutional neural networks.

Charles T. Hemphill, John J. Godfrey, and George R. Doddington. 1990. The ATIS spoken language systems pilot corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27,1990*.

Rebecca Hwa. 1999a. Supervised grammar induction using training data with limited constituent information. *CoRR*, cs.CL/9905001. Version 1.

Rebecca Hwa. 1999b. Supervised grammar induction using training data with limited constituent information. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*.

Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, second edition. Pearson Prentice Hall.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukdar. 2021. Muril: Multilingual representations for indian languages. *CoRR*, abs/2103.10730.

Mukul Kumar, Youna Hu, Will Headden, Rahul Goutam, Heran Lin, and Bing Yin. 2020. Shareable representations for search query understanding. *CoRR*, abs/2001.04345.

Sachin Kumar and Regina Politi. 2019. Understanding user query intent and target terms in legal domain. In *Natural Language Processing and Information Systems*, pages 41–53, Cham. Springer International Publishing.

Uichin Lee, Zhenyu Liu, and Junghoo Cho. 2005. Automatic identification of user goals in web search. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, page 391400, New York, NY, USA. Association for Computing Machinery.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? *CoRR*, abs/1906.01502.

Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. 2019. A stack-propagation framework with token-level intent detection for spoken language understanding. 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 2078–2087, Hong Kong, China. Association for Computational Linguistics.

K Sreelakshmi, P C Rafeeque, S Sreetha, and E S Gayathri. 2018. Deep bi-directional lstm network for query intent detection. *Procedia Computer Science*, 143:939–946. 8th International Conference on Advances in Computing Communications (ICACC-2018).

Yu Wang, Yilin Shen, and Hongxia Jin. 2018. A bi-model based RNN semantic frame parsing model for intent detection and slot filling. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 309–314, New Orleans, Louisiana. Association for Computational Linguistics.

A Intent taxonomy Tree

In Table 5 we show a part of our proposed taxonomy of one of the level 1 intent. Similar to this we have seven more level 1 intents with their own level 2 and level 3 intent categories.

L1	L2	L3	Example
Movies	Movie title only		bachchan film
	Person	Actor	आलोक नाथ मूवी
		Director	rajkumar hirani tumbbad
		Producer	vikram bhatt movies list
		Character	अजय देवगन बाजीराव सिंघम
	Quantity	Run time	फिल्म एक घंटा की
		Show timings	bookmyshow pathankot
		Earnings	endgame box office
		Ratings	the body imdb
		Screen time	mohabbatein duration
		Budget	kanchana 3 budget
		Release Date or Year	tanhaji full movie hd 2020
	access via/platform	TV	war 2019 movie on tv
		OTT platforms	netflix पर गोलमाल फिल्म
		Download	नूरी फिल्म डाउनलोड
		In cinema (2D, 3D)	Robot 2.0 full movie in boxed 3D
	Publicity or Promotion		झील के उस पार ट्रेलर
	Events		बॉलीवुड events
	Location		नदिया के पार फिल्म की शूटिंग कहां हुई
	Production House		आरके स्टूडियो कुरावली
	Dialogues or Story		नदिया के पार holi hai (dialogues)
	Movie Scenes		दिलवाले दुल्हनिया ले जायेंगे मूवी भाग 1
	Award and Honors		सावित्रीबाई फुले पुरस्कार
	Language		बाहुबली 2 फुल मूवी हिंदी में
	Genre	Action	action box office movies 2020
		Horror	क्लासिक हॉरर मूवीज
		Comedy	हाउसफुल 4 जैसी फ़िल्में
		Science fiction	इंग्लिश फिल्म टर्मिनेटर हिंदी
		Romance	best romantic movies of all time
		Animation	आइस ऐज एनिमेटेडम film
		Adventure	डोरेमोन का एडवेंचर फिल्म
		Thriller	थ्रिलर एक्शन साउथ फिल्म
		Biopic	संजय दत्त की बायोपिक

Table 5: Intent taxonomy for Movies (one of the Level 1) intent