UQ IElab at TREC 2019 Decision Track

Jimmy
University of Queensland, Brisbane, Australia
University of Surabaya (UBAYA), Surabaya, Indonesia
jimmy@uqconnect.edu.au

Guido Zuccon
University of Queensland
Brisbane, Australia
g.zuccon@uq.edu.au

ABSTRACT

We describe our participation to the TREC 2019 Decision Track. The first year of this track challenges participants to devise search technologies that retrieve correct health advice from web resources, with the intent to better support people's health decision making.

Our solution addressed this challenge by extending the Entity Query Feature Expansion model (EQFE), a knowledge base (KB) query expansion method. In previous work we showed that Wikipedia and the Consumer Health Vocabulary resource can be effective as basis for consumer health search query expansion, within the EQFE method. To obtain query expansion terms, first, we mapped entity mentions to KB entities by performing exact matching. After mapping, we used the Title of the mapped KB entities as the source for expansion terms. Despite previous evaluation demonstrating the effectiveness of this method, EQFE did not provide the expected gains over not using query expansion, on both relevant-based and credibility-based evaluation measures.

1 INTRODUCTION

The TREC Decision Track is a multi-year track that aims to: (1) foster the development of search engine technologies that help people making better (health) decisions, and (2) develop evaluation methods that account for the quality of the (health) decisions made based on the retrieved search results. In 2019 (the first year of the track), the track challenges participants to devise search technologies that retrieve correct health information for a user information need.

As part of the TREC Decision Track, 50 consumer health search topics are provided. Each search topic contains four fields: a query string, a Cochrane identifier, a description and a verbose narrative for the topic – all could be used to form retrieval runs. The document corpus used in this challenge is the ClueWeb12-B13.

To address this challenge we applied and extended the Entity Query Feature Expansion model (EQFE), a knowledge base (KB) query expansion method [2], which we have recently found performing competitively on the previous CLEF e-Health IR challenges [4–6]. By producing query expansions using EQFE, we seek to overcome the issue of poor query formulation, common in consumer health search [11, 12]; EQFE does so by augmenting the query with related terms terms.

2 OUR APPROACH

The Entity Query Feature Expansion model is formally defined as:

$$\hat{\vartheta}_q = \sum_M \sum_f \lambda_f \vartheta_{f(EM,SE)} \qquad (1)$$

where M are the entity mentions and contain uni-, bi-, and tri-grams generated from the query; f is a function used to extract the expansion terms. $\lambda_f \epsilon (0, 1)$ is a weighting factor. $\vartheta_{f(EM,SE)}$ is a function to map entity mention M to the KB features EM (e.g., "Title", "Aliases", "Links", "Body", etc.) and extract expansion terms from source of expansion SE (e.g.,"Title", "Aliases", etc.).

Knowledge-bases

One source of high quality health expansion terms is the Unified Medical Language System (UMLS) [1]. However, UMLS concepts are rarely mentioned in consumer health queries: Keselman et al. [7] showed that only 8.1% of the possible n-grams constructed from consumer queries can be mapped (i.e., exact match) to UMLS concepts. This is unlike the queries contained in this year's TREC Decision Track task, for which our method could map all queries to at least one UMLS concept – and in fact it mapped 85.8% of the query terms to UMLS concept (exact match, no UMLS semantic type restriction). This may be due to the growth of the UMLS terminology over the years (Keselman et al.'s study is 15 years old at the time of writing this notebook paper). Another reason for this difference may be that the queries considered in this TREC track are representative of only a (very specific) subset of the tasks, information needs and associated queries that manifest in consumer health search.

Unlike the UMLS, Wikipedia is a crowdsourced, general purpose KB allowing people to promote and describe new concepts or augment existing concepts. While general purpose, Wikipedia contains considerable and detailed health information that has been effectively used in health related information retrieval [5, 6, 9].

Other than the UMLS and Wikipedia, we have also considered the Consumer Health Vocabulary (CHV) [7, 13] as basis for query expansion. The CHV was built to provide a mapping between consumer health terms and UMLS concepts. This mapping was constructed by extracting n-grams from MedlinePlus queries and various health-focused bulletin boards; then, automatically mapping these n-grams to UMLS via exact match comparison. Any un-mapped n-grams are then manually mapped to the UMLS [7]. From 2007, the CHV is available as part of the UMLS entries with "CHV" as source (i.e., tuples in table MRCONSO with attribute "SAB" equal to "CHV").

For this challenge, we considered the Wikipedia and CHV as the knowledge-bases to source our expansion terms from as expanded queries based on these two KBs outperformed expanded queries based on UMLS KB in our previous work [6]. In that work, we found that using UMLS KB as the basis for query expansion led to the largest number of expansion terms, often drifting the query away from its original intent. Figure 1 shows the features we used for mapping the queries to entities in the KB and features that we considered as the source of expansion terms. For the Wikipedia KB, a single entity is represented by a single Wikipedia page (the page title identifies the entity). Beyond titles, Wikipedia also contains many page features useful in a retrieval scenario: entity title (E), categories (C), links (L), aliases (A), and body (B). As for the CHV

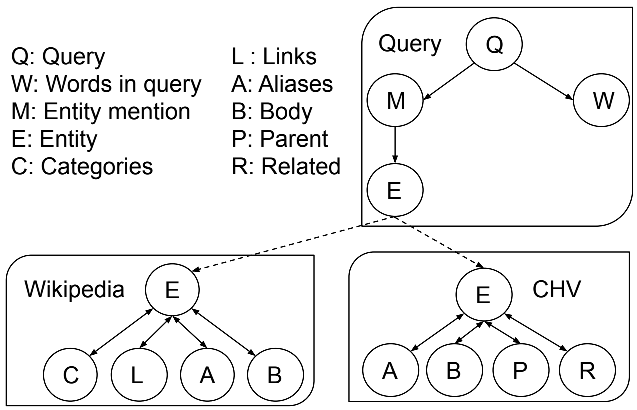

Figure 1: Summary of expansion sources for our extension of the EQFE model.

Id	Field	Query String
01	Query	Query field
02	Description	Description field
03	Description	UMLS entities in the description field
04	Narrative	UMLS entities in the narrative field
05	Query	Query + title of the mapped CHV entities
06	Description	Description + title of the mapped CHV entities
07	Query	Query + title of the mapped Wikipedia entities
08	Description	Description + title of the mapped Wikipedia entities
09	Query	Query + title of the mapped CHV and Wikipedia entities
10	Description	Description + title of the mapped CHV and Wikipedia entities

Table 1: Summary of the runs submitted to TREC 2019 Decision Track.

KB, a single entity is represented by the most frequently used terms for a single concept unique identifier (CUI). Features of a CHV KB entity are aliases (A), body (B), parent concepts (P), and related concepts (R).

Description of Runs

To produce all runs, we indexed the ClueWeb12-B13 corpus using Elasticsearch 5.1.1, with stopping and Porter stemming. As underlying retrieval model we used BM25F. We used the title field and the body field, with weighting factors 1 and 2, respectively. These were found to be the optimal weights for BM25F for the CLEF 2016 eHealth collection [3], which used the same corpus for a similar task.

We submitted 10 runs as described in Table 1: four runs based on the query field and six "other" runs based on either the description or the narrative field. Next, we describe each run in details.

Run 01 and **Run 02** are baseline runs that used the content of either query or description fields as query, respectively.

To obtain **Run 03** and **Run 04**, we extracted health entity mentions from the description and narrative fields, respectively. We used QuickUMLS [10] to extract the health entity mentions.

To obtain **Run 05** and **Run 06**, we:

(1) indexed English and non-obsolete CHV concepts that are associated to the four key aspects of medical decision criteria (i.e., symptoms, diagnostic test, diagnoses, and treatments), as used by Limsopatham et al. [8].
(2) extracted uni-, bi-, tri-grams of the query field (Run 05) and the description field (Run 06) that matched entities in the CHV.
(3) exact matched the extracted n-grams to the CHV's aliases.
(4) used the title of the matched entities as expansion terms.

To obtain **Run 07** and **Run 08**, we:

(1) indexed Wikipedia pages with Medicine infobox type and pages with infobox containing links to medical terminologies such as Mesh, UMLS, SNOMED CT, etc.
(2) extracted uni-, bi-, tri-grams of the query field (Run 07) and the description field (Run 08) that matched CHV entities.
(3) exact matched the extracted n-grams to the Wikipedia's aliases.
(4) used the title of the matched entities as expansion terms.

To obtain **Run 09**, we combined expansion terms obtained for Run 05 and Run 07. Finally, to obtain **Run 10**, we combined expansion terms obtained for Run 06 and Run 08. This combination performed the best when compared to other possible combinations in our previous study [6].

3 RESULTS

Table 2 reports the average effectiveness of our four runs based on query field, on the four measures used by this TREC 2019 track. We measured statistical significant differences between runs using paired t-test adjusted with Bonferroni correction. Statistical significant differences ($\alpha \leq 0.05$) are indicated with superscripts. For comparison, we also report the average of best and median performance across runs from all participants (i.e., 32 runs in total).

The results in Table 2 suggest that the expanded queries (Run 05, 07, and 09) failed to outperform the original title queries (Run 01), with the exception of Run 05 and Run 09 which performed better on NLRE (but no statistical significant improvements). This is despite the titles of the top ten results from each of these runs appearing to be relevant at first sight (Table 4 – further analysis below). Further investigation is needed to understand why these methods fail to provide the expected improvements: note that improvements are generally not seen for both traditional ad-hoc measures (MAP, nDCG) and those centred on credibility (NLRE, CAM).

Table 3 reports the average score of our "other" runs. Results show that verbose descriptions (i.e., description and narrative fields) performed worst than the query field. This may be because the query field accurately contains the targeted concepts. Therefore, additional terms in description and narrative fields may have introduced noise that lead to less relevant results being retrieved.

Along with the evaluation measures provided by the organisers, we also manually evaluated the title of the top ten documents from each of our runs. For brevity, Table 4 shows the top ten titles only for runs based on the query field for two topics. We found that the titles

Id	NLRE	CAM	MAP	nDCG@10
01	0.9960^{b}	$\mathbf{0.5208}^{5,9,b,m}$	$\mathbf{0.3334}^{5,9,b,m}$	$\mathbf{0.4828}^{5,9,b,m}$
05	$\mathbf{0.9963}^{b}$	$0.4665^{1,b}$	$0.2642^{1,b}$	$0.4125^{1,b}$
07	0.9962^{b}	$0.5084^{9,b}$	$0.3129^{9,b,m}$	0.4651^{b}
09	$\mathbf{0.9963}^{b}$	$0.4621^{1,7,b}$	$0.2568^{1,7,b}$	$0.4065^{1,b}$
best	$1.0000^{1,5,7,9,m}$	$0.6257^{1,5,7,9,m}$	0.4845^{m}	0.6761^{m}
median	0.9967^{b}	$0.4265^{1,b}$	$0.2117^{1,7,b}$	$0.3379^{1,b}$

Table 2: Mean performance across queries from runs using query field. Superscripts indicate statistical significance difference ($\alpha \leq 0.05$). Best performing scores for each measure are highlighted in bold. The last two rows show the average of best and median performance across 32 runs from all participants

Id	NLRE	CAM	MAP	nDCG@10
02	0.9959^{b}	0.3872^{b}	0.2082^{b}	0.3525^{b}
03	$\mathbf{0.9966}^{b}$	$\mathbf{0.4547}^{b}$	$\mathbf{0.2723}^{b}$	$\mathbf{0.3948}^{b}$
04	0.9957^{b}	0.4234^{b}	0.2387^{b}	0.3565^{b}
06	0.9959^{b}	0.4493^{b}	0.2613^{b}	0.3906^{b}
08	0.9963^{b}	$\mathbf{0.4547}^{b}$	0.2718^{b}	0.3936^{b}
10	0.9961^{b}	0.4491^{b}	0.2611^{b}	0.3898^{b}
best	$1.0000^{2,3,4,6,8,10,m}$	$0.6257^{2,3,4,6,8,10,m}$	$0.4845^{2,3,4,6,8,10,m}$	$0.6761^{2,3,4,6,8,10,m}$
median	0.9967^{b}	0.4265^{b}	0.2117^{b}	0.3379^{b}

Table 3: Mean performance across "other" queries from runs using other than the query field. Superscripts indicate statistical significance difference ($\alpha \leq 0.05$). Best performing scores for each measure are highlighted in bold. The last two rows show the average of best and median performance across 32 runs from all participants

of the top ten search results are at first sight indicative of relevant document. Nevertheless, we also found that results from Run 07 (i.e., query field expanded with title of the mapped Wikipedia entity) are similar to the Run 01 (query field), which indicated that expansions from Wikipedia failed to impact the search results. In fact, using the Wikipedia KB we were only able to expand 13 out of the 50 queries; this resulted in adding on average 1.15 expansion terms to each original query, when expansion was possible. On the other had, using the CHV KB we were able to expand 42 out of the 50 queries, resulting to the addition of 3.33 expansion terms on average. This may be because Wikipedia contains a limited number of health entities compared to CHV, as noted in our previous study [6].

4 CONCLUSIONS

In this paper we reported the methods used in the IElab research team's participation to the TREC 2019 Decision Track, to which we submitted a total of ten runs. We applied query expansion methods that we found effective in our previous studies [5, 6]. However, the empirical results obtained on the TREC 2019 Decision Track collection show that the obtained expanded queries failed to perform better than the original queries.

The type of queries used in this challenge relate to the intent of obtaining additional information about a known health concept. In other words, the queries explicitly contain the target health concepts, for example "cranberries urinary tract infections", unlike the more ambiguous and underspecified queries used in our previous evaluation of EQFE (self-diagnosis oriented queries). In fact, all queries in this challenge contained at least one UMLS concept mention, as extracted by QuickUMLS. Furthermore 85.8% of the query terms are matched by QuickUMLS to UMLS concepts (this analysis did not restrict matches to specific UMLS semantic types, e.g., to disease concepts only). Our attempts in clarifying such precise queries may have lead to query drift, which therefore lead to less accurate results.

ACKNOWLEDGMENTS

Jimmy is sponsored by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan / LPDP) (20151022014644). Guido Zuccon is the recipient of an Australian Research Council DECRA Research Fellowship (DE180101579) and a Google Faculty Award; both these grants funded this study.

REFERENCES

[1] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32, suppl_1 (2004), D267–D270.

[2] Jeffrey Dalton, Laura Dietz, and James Allan. 2014. Entity Query Feature Expansion Using Knowledge Base Links. In *SIGIR'14*. 365–374.

[3] Jimmy, Guido Zuccon, and Bevan Koopman. 2016. Boosting Titles Does Not Generally Improve Retrieval Effectiveness. In *ADCS'16*. 25–32.

[4] Jimmy, Guido Zuccon, and Bevan Koopman. 2017. QUT ielab at CLEF 2017 e-Health IR Task: Knowledge Base Retrieval for Consumer Health Search.. In *CLEF (Working Notes)*.

[5] Jimmy, Guido Zuccon, and Bevan Koopman. 2018. Choices in knowledge-base retrieval for consumer health search. In *European Conference on Information Retrieval*. Springer, 72–85.

Topic 1, Query: cranberries urinary tract infections				
Rank	**Run 01**	**Run 05**	**Run 07**	**Run 09**
1	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	Urinary Tract Infection Cranberry | Details	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	urinary tract infection natural treatment Urinary Infection Remedies Treating Urinary Tract Infections with Natural Health
2	The Cranberry and Urinary Tract Infections : FurtherHealth com	urinary tract infection natural treatment Urinary Infection Remedies Treating Urinary Tract Infections with Natural Health	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	Natural Remedy Urinary Tract Infection
3	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	The Cranberry and Urinary Tract Infections : FurtherHealth com	Urinary Tract Infection Cranberry | Details
4	urinary tract infections pics Urinary Tract Cure Drinking Cranberry Juice Doesn???t Always Work to Cure U t i	Natural Remedy Urinary Tract Infection	women and urinary tract infections | 5 Sure Fire Tips For Treating a Urinary Tract Infection	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?
5	Cranberry Juice Is One Of The Finest Urinary Tract Infect	The Cranberry and Urinary Tract Infections : FurtherHealth com	infection urinary tract | Urinary Tract Infection Did You Know You Can Treat It Naturally?	urinary tract infection untreated Home Remedies for Urinary Tract Cure UTI with Holistic Health Tips
6	Urinary Tract Infections In Men | Natural Cures For Urinary Tract Infections	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?	chronic urinary tract infection photos | How to Get Rid of a Urinary Tract Infection Simple Secrets Your Doctor "Forgot" to Tell You	Urinary Tract Infection Alternative Treatment are Cranberries the Only U t i Natural Cure?
7	infection urinary tract | Urinary Tract Infection Did You Know You Can Treat It Naturally?	urinary tract infection untreated Home Remedies for Urinary Tract Cure UTI with Holistic Health Tips	Urinary Tract Infection Natural Remedies	The Cranberry and Urinary Tract Infections : FurtherHealth com
8	urinary tract infection natural treatment Urinary Infection Remedies Treating Urinary Tract Infections with Natural Health	Urinary Tract Infection	Natural Remedy Urinary Tract Infection	urinary tract infection pain | You Have Asked About Turmeric And Urinary Tract Infection
9	cranberry pills help uti Prevalence of Urinary Tract Infection in Females Why They are More Prone to UTI's	urinary tract infections pics Urinary Tract Cure Drinking Cranberry Juice Doesn???t Always Work to Cure U t i	cranberry juice and urinary tract infections UTI Treatment 7 Things You Need to Know About Your Urine Diet and Infection	Urinary Tract Infection Natural Remedies
10	women and urinary tract infections | 5 Sure Fire Tips For Treating a Urinary Tract Infection	Cranberry Juice Is One Of The Finest Urinary Tract Infect	cranberry pills help uti Prevalence of Urinary Tract Infection in Females Why They are More Prone to UTI's	uti pictures | Five Secrets to Begin Your Urinary Infection Cure And Keep it from Coming Back
Topic 2, Query: acupuncture insomnia				
Rank	**Run 01**	**Run 05**	**Run 07**	**Run 09**
1	Acupuncture for Insomnia Insomniacs	Insomnia Treatment	Acupuncture for Insomnia Insomniacs	Insomnia Treatment
2	Learn How Acupuncture Can Relieve Insomnia and Improve Your Energy	Insomnia Solutions	Learn How Acupuncture Can Relieve Insomnia and Improve Your Energy	Insomnia Solutions
3	Archive ELECTRIC ACUPUNCTURE FOR INSOMNIA	Cures for Insomnia | Health and Life	Archive ELECTRIC ACUPUNCTURE FOR INSOMNIA	Cures for Insomnia | Health and Life
4	Acupuncture Helps Insomnia	Archive ELECTRIC ACUPUNCTURE FOR INSOMNIA	Acupuncture Helps Insomnia	Archive ELECTRIC ACUPUNCTURE FOR INSOMNIA
5	Insomnia How to Fall Asleep | Balanced Being Acupuncture	Trouble Sleeping?: Acupuncture and Chinese Medicine	Insomnia How to Fall Asleep | Balanced Being Acupuncture	Trouble Sleeping?: Acupuncture and Chinese Medicine
6	The Pulse Acupuncture & Massage Newsletter	Treating Insomnia with Chinese Medicine | ACOS Traditional Chinese Medicine & Acupuncture School | TCM Acupuncturist & Chinese Herbal Medicine Programs	The Pulse Acupuncture & Massage Newsletter	Treating Insomnia with Chinese Medicine | ACOS Traditional Chinese Medicine & Acupuncture School | TCM Acupuncturist & Chinese Herbal Medicine Programs
7	Treat insomnia using Tradition Chinese Medicine acupuncture Sanjiu 39 Medicine Center	Acupuncture for Insomnia Insomniacs	Treat insomnia using Tradition Chinese Medicine acupuncture Sanjiu 39 Medicine Center	Acupuncture for Insomnia Insomniacs
8	Acupuncture | The Wellness Center	Sound And Music In Finding An Insomnia Cure	Acupuncture | The Wellness Center	Sound And Music In Finding An Insomnia Cure
9	Trouble Sleeping?: Acupuncture and Chinese Medicine	Acupuncture | The Wellness Center	Trouble Sleeping?: Acupuncture and Chinese Medicine	Acupuncture | The Wellness Center
10	Acupuncture Com Recent Research August 2005	Homeopathic Cures for Insomnia & Sleeplessness	Acupuncture Com Recent Research August 2005	Homeopathic Cures for Insomnia & Sleeplessness

Table 4: Top ten results from runs based on the query field for topic 1 and 2.

[6] Jimmy, Guido Zuccon, and Bevan Koopman. 2018. Payoffs and pitfalls in using knowledge-bases for consumer health search. *Information Retrieval Journal* (2018), 1–45.

[7] Alla Keselman, Catherine Arnott Smith, Guy Divita, Hyeoneui Kim, Allen C Browne, Gondy Leroy, and Qing Zeng-Treitler. 2008. Consumer health concepts that do not map to the UMLS: where do they fit? *Journal of the American Medical Informatics Association* 15, 4 (2008), 496–505.

[8] Nut Limsopatham, Craig Macdonald, and Iadh Ounis. 2013. Inferring conceptual relationships to improve medical records search. In *Proceedings of the Tenth Conference on Open Research Areas in Information Retrieval*. 1–8.

[9] Luca Soldaini, Arman Cohan, Andrew Yates, Nazli Goharian, and Ophir Frieder. 2015. Retrieving medical literature for clinical decision support. In *European Conference on Information Retrieval*. Springer, 538–549.

[10] Luca Soldaini and Nazli Goharian. 2016. Quickumls: a fast, unsupervised approach for medical concept extraction. In *MedIR workshop, sigir*.

[11] Isabelle Stanton, Samuel Ieong, and Nina Mishra. 2014. Circumlocution in Diagnostic Medical Queries. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*.

[12] Qing Zeng, Sandra Kogan, Nachman Ash, Robert A Greenes, and Aziz A Boxwala. 2002. Characteristics of consumer terminology for health information retrieval. *Journal of Methods of Information in Medicine* 41, 4 (2002), 289–298.

[13] Qing T Zeng and Tony Tse. 2006. Exploring and developing consumer health vocabularies. *Journal of the American Medical Informatics Association* 13, 1 (2006), 24–29.