That's Interesting, Tell Me More!

Finding Descriptive Support Passages for Explaining Knowledge Graph Relationships

Sumit Bhatia[1], Purusharth Dwivedi[2], and Avneet Kaur[2]

[1]IBM Research AI, New Delhi, India

[2]Indraprastha Institute of Technology (IIIT), Delhi, India

We are used to Search Results with Some Supporting Description

Google

Janaki Ammal

All Images News Maps Videos More Settings Tools

About 279,000 results (0.34 seconds)

Janaki Ammal - Wikipedia
https://en.wikipedia.org/wiki/Janaki_Ammal
Janaki Ammal Edavalath Kakkat (4 November 1897 – 7 February 1984) was an Indian botanist who conducted scientific research in cytogenetics and ...
Fields: Botany, Cytology **Nationality**: Indian
Institutions: University Botany Laboratory, Madras
Early life · Career · Research · Janaki Ammal National ...

Meet Janaki Ammal, the Legendary Botanist who Sweetened India's ...
https://www.thebetterindia.com/75174/janaki-ammal-botanist-sugarcane-magnolia/
Nov 16, 2016 - One of the first women scientists to receive the Padma Shri way back in 1977, Edavaleth Kakkat **Janaki Ammal** lived a life only a handful of ...

Janaki Ammal Biography - Childhood, Life Achievements & Timeline
https://www.thefamouspeople.com/profiles/janaki-ammal-6543.php
Nov 1, 2017 - **Janaki Ammal** Edavaleth Kakkat was an Indian botanist best remembered for her work on sugarcane and eggplant. ... Born in late 19th century in Kerala, India, **Ammal** was encouraged to pursue her intellectual interests from a young age. ... Her father Dewan Bahadur Edavalath Kakkat Krishnan ...

Edavaleth Kakkat Janaki Ammal - An Introduction
www.iiim.res.in/herbarium/edavaleth-kakkat-janaki-ammal.htm

Why not for facts stored in Knowledge Graphs?

p21 is a potent cyclin-dependent kinase inhibitor (CKI). The p21 (CIP1/WAF1) protein binds to and inhibits the activity of cyclin-CDK2, -CDK1, and -CDK4/6 complexes, and thus functions as a regulator of cell cycle progression at G_1 and S phase.

It helps...

- Exploratory search
- Increases users' trust and confidence in the system
- Crucial in domains like healthcare, finance, security and intelligence
- When working with uncertain information, natural language descriptions help in decision making (Gkatzia et al., 2016)
- Knowledge Graphs constructed automatically using machine learned methods

Related Work

In context of Web Search – explaining relevance for an ad hoc text query (Blanco and Zaragoza, 2010)

Specific to web search, relies on external resources such as web logs

Using templates for relationships (Saldanha et al., 2016, Voskarides et al., 2017)

Needs to construct different templates for all possible relationship types in the graph

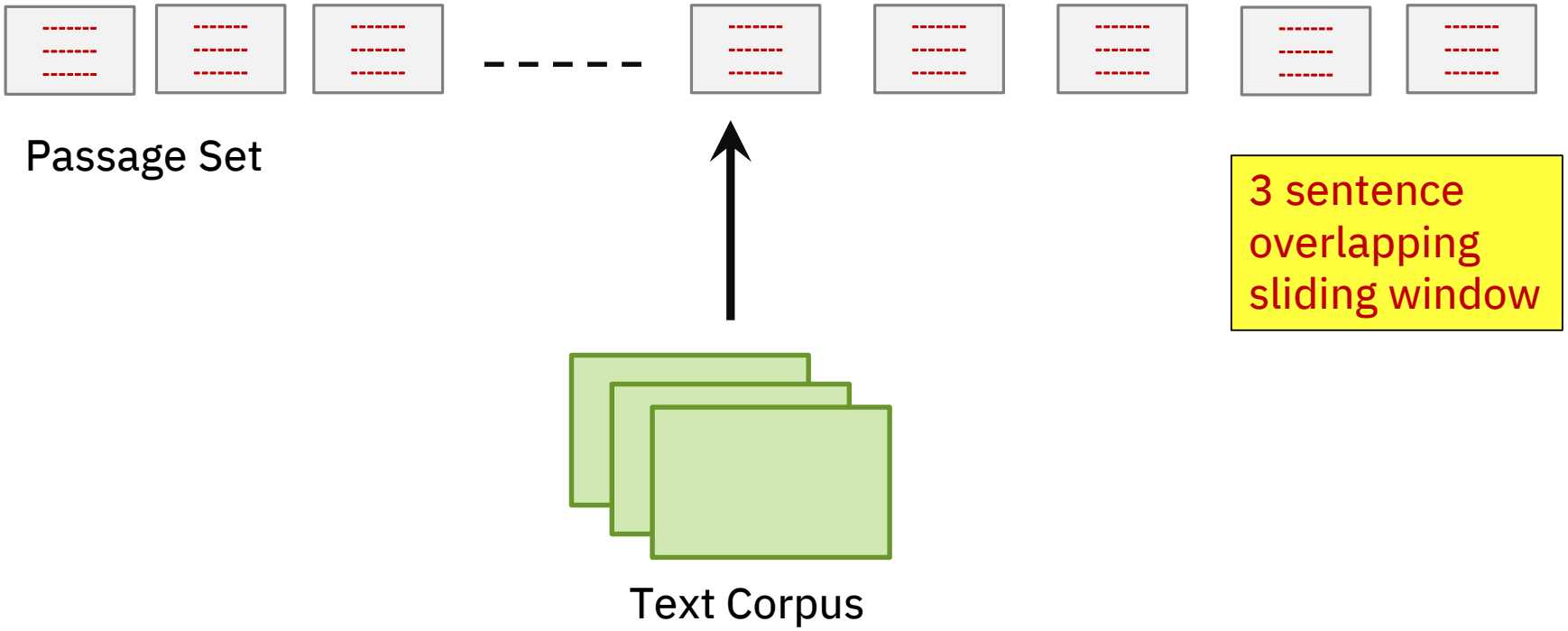
Essentially, a *textual representation* of relationship, not description

Learning to rank sentences (Voskarides et al., 2015)

Supervised, often training data not easy to get

How we do it

Key Assumption: Availability of a background text corpus



How we do it

Given the relationship *R = {s,t,r}*, what is the probability of generating the passage *p?*

$$P(p|\mathcal{R}) \propto \underbrace{P(s|p) \times P(t|p)}_{\text{entity probability}} \times \underbrace{P(r|p)}_{\text{relationship probability}} = \prod_{w \in S \cup T \cup R} \boxed{P(w|p)}$$

Passage Set

How we do it

Estimating $P(w|p)$

Challenge: Passages are small, lack sufficient evidence to compute probabilities

- Terms that are present get high probabilities
- Terms that are absent get zero probabilities

How we do it

Estimating $P(w|p)$

Solution: A smoothed mixture model for probability estimation

$$P(w|p) = P(w|\Theta_{MM})$$
$$= \lambda_1 \underbrace{P(w|\Theta_p)}_{\text{passage-level evidence}} + \lambda_2 \underbrace{P(w|\Theta_d)}_{\text{document-level evidence}} + \lambda_3 \underbrace{P(w|\Theta_c)}_{\text{collection-level evidence}}$$

Advantage 1: Overcome data sparsity

Advantage 2: Passages from documents mainly about source/target entities get a boost; indirect coreference resolution

How Well Does the Proposed Solution Work

Source Corpus: Wikipedia

Relationship Triples: Wikipedia Top-25 most viewed pages used to create a set of seed entities; their relationships selected from WikiData

Baselines:

1. Passage retrieval model based on inference networks (Inf. N/w)
2. Passage Retrieval using query expansion for relationship synonyms (Inf. N/w + Rel. Exp.)

Relevance Judgments: User study with 5 evaluators

Data, relevance judgments, and implementation available from:

https://github.com/sumit-research/kg-support-passages

How Well Does the Proposed Solution Work

Effectiveness Evaluation

50 queries, top 5 passages by each method

	P@1	Precision	MRR
Inf. N/w	0.251	0.156	0.272
Inf. N/w + Rel. Exp.	0.165	0.088	0.144
Proposed Solution	**0.860**	**0.727**	**0.805**

How Well Does the Proposed Solution Work

Preference Evaluation

150 queries, top passage by proposed method and baseline shown side by side

	Both not Useful	Both equally useful	Baseline Preferred	Proposed Solution Preferred
Evaluator 1	14	13	3	20
Evaluator 2	6	17	2	25
Evaluator 3	10	6	6	28
Total	30	36	11	73

Some Examples

<John Cena, nickname, The Prototype>

Overestimation

Baseline: A prototype is something that is representative of a category of things, or an early engineering version of something to be tested. Prototype may also refer to: Automobiles. Citro¨en Prototype C, a range of vehicles created by Citro¨en from 1955 to 1956 Citro¨en Prototype Y, a project of replacement of the Citro¨en Ami studied by Citro¨en in the early seventies Daytona Prototype, a sports car

Scores boosted by document level evidence

Proposed: In 2001, Cena signed a developmental contract with the WWF and was assigned to its developmental territory Ohio Valley Wrestling (OVW). During his time there, Cena wrestled under the ring name The Prototype and held the OVW Heavyweight Championship for three months and the OVW Southern Tag Team Championship (with Rico Constantino) for two months. Throughout 2001, Cena would receive four tryouts for the WWF main roster, as he wrestled multiple enhancement talent wrestlers on both WWF house shows and in dark matches before WWF television events

Some Examples

<Alan Colmes, employer, Fox News>

Goldlines television advertising includes cable networks such as CNN, CNBC, Fox News, History International and Fox Business. Goldline has also been the sponsor of the shows of a number of conservative radio and television hosts, including The American Advisor, and The Glenn Beck Program, The Laura Ingraham Show, The Fred Thompson Show, The Huckabee Report, The Lars Larson Show, The Monica Crowley Show, The Mark Levin Show, and The Alan Colmes Show. In 2009, Goldline incorrectly labeled Glenn Beck as a paid spokesman on its website which raised concerns with his employer, Fox News, which prohibit such a relationship; they later corrected it to radio sponsor.

References to Colmes, Fox News; originating document also talks about fox news.

But a poor result as does not describe the relationship

To Conclude

A probabilistic method to find supporting passages describing knowledge graph relationships

Unsupervised, no dependence on external sources

For various applications, quick easy first-cut solution

Further Directions

Handling numeric relations *< Burj Khalifa, height, 828 metres >*

Handling negations *< X,spouseOf, Y> -- X is not spouse of Y*

Integrating with Fact checking systems such as DeFacto (Lehmann et al. 2012)

References

Gkatzia, D., Lemon, O., Rieser, V.: Natural language generation enhances human decision making with uncertain information. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016

Blanco, R., Zaragoza, H.: Finding support sentences for entities. In: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. pp. 339–346. ACM (2010)

Saldanha, G., Biran, O., McKeown, K., Gliozzo, A.: An entity-focused approach to generating company descriptions. In: The 54th Annual Meeting of the Association for Computational Linguistics. p. 243 (2016)

Voskarides, N., Meij, E., de Rijke, M.: Generating descriptions of entity relationships. In: Advances in Information Retrieval: 39th European Conference on IR Research, ECIR 2017

Voskarides, N., Meij, E., Tsagkias, M., de Rijke, M., Weerkamp, W.: Learning to explain entity relationships in knowledge graphs. In: ACL (1). pp. 564–574 (2015)

Lehmann, J., Gerber, D., Morsey, M., Ngomo, A.C.N.: Defacto-deep fact validation. In: International Semantic Web Conference. pp. 312–327. Springer (2012)