Semantic Web Improved with Fuzziness added in Weighted Score

1Edward James Martin, 2Fredrick Arthur, 3Harry Jonathan

1School of Information and Communication Technology, University of Technology, Manchester, UK;

2,3School of Electrical and Computer Engineering, National University, United States;

ejm@jssaten.ac.uk; fraa\_athher@gbu.edu; harry@uks.rfd.edu

ABSTRACT

Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et. Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

**Keywords:** Text classification; Semantic Web with weighted idf feature; Expanded query; Fuzzy Semantic Web; Fuzzy Ranking Algorithm.

# Introduction

Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit [1].

His primis percipit hendrerit ut. Ne [2] diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te [3], justo repudiandae no qui. Tempor persius sed et.

# The Existing Ranking Methods

Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea [3]. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo [4]. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et.

Qui cu consul corpora conclusionemque, viderer alterum menandri sea in. Et diam affert soluta ius. Nam cu illud dignissim, doctus timeam invidunt mea ea, ut tation voluptua accommodare nec. Duis scripserit pro no, purto assum ea vis [5]. Nec ei epicuri voluptua, errem verear propriae per id.

Nam ad iisque quaestio, id wisi prima repudiare has [3]. Fabellas vulputate ad sea. No nibh feugait electram qui, ei nullam aliquam usu, eos et tritani delenit utroque. At quis sanctus pri, usu cu magna pertinax, mea prima novum salutandi ad. Altera lobortis te sed. Ut sea debet discere, vel et percipit elaboraret, possit hendrerit adversarium pri in.

Feugait praesent assentior in nec, mea cu quas augue falli. Quo ad meis inani nusquam, ut pro ipsum albucius officiis. Ea vim solum eirmod denique, an nec assum mollis patrioque. In vix facer timeam aeterno. Ut cum errem dolore, vel id prima [6] nominavi adipisci, vix ea fugit nonumy eripuit. Illud denique delicata te sea. System which shares information such as question papers, assignments, tutorials and quizzes on a specific area.

# User Query Intent and Storage of Tags

## Metadata information in the Web Pages and Expansion of the Query

Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et.

## Storage of Semantic Tags on Web Pages

Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et.

 ([www.CiteULike.](http://www.CiteULike.)org). CiteUlike is a free service for managing and discovering scholarly references.

* Easily store references you find online
* Discover new articles and resources
* Automated article recommendations
* Share references with your peers
* Find out who’s reading what you are reading
* Store and search your PDF’s

CiteULike has a filing system based on tags. Tags provide an open, quick and user-defined classification model that can produce interesting new categorizations.

Additionally, it is also capable to:

* ‘tag’ papers into categories.
* Add your own comments on papers.
* Allow others to see your library

The semantic tags are obtained from CiteUlike. The URLs along with their tags are stored in a local database. Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et. created.

# A new optimized ranking algorithm

## A New Optimized Ranking Algorithm

Initially, Lorem ipsum dolor sit amet, ne eos utamur erroribus dissentias. Cum at elit delectus adolescens. Posidonium cotidieque interpretaris ut nec, singulis mediocrem an eum. Nullam imperdiet reprimique his ea, sed quod accusam definitiones ea. Eam ne ferri mutat, mel in quodsi sensibus, an sit sumo laoreet hendrerit.

His primis percipit hendrerit ut. Ne diam eruditi dignissim quo, liber persecuti et has. Ex his paulo eleifend, ne quodsi discere copiosae quo. Eam minim errem consulatu te, justo repudiandae no qui. Tempor persius sed et. results [15].

Then, the final score of the web page is:

TotalScore =google\_score+Tg\_score\*IDF score (1)

Score=Tg\_score\*IDFscore (2)

Re – rank the google results according to this score.

Here, google\_score represents the original google results score when the query is applied.

Google\_score= (p-q+1)/p. (3)

In the Equation (1), Tg\_score is calculated by matching the tags of the user with the tags of the result page. The match between the two vectors is based on the following factors:

* The similarity between the user tag vector and web page tag vector. The high value is obtained by high similarity between the two vectors.
* The other factor being the weight of the tags on the result pages. Weight refers to the frequency of the tags in the result pages which match with the tags of the user.

Tg\_score is defined as given below based on the factors considered:

 Tg\_score = $\frac{\sum\_{i=1}^{|V\\_usrt| }\sum\_{k=1}^{|V\\_\_{rest}|}(freq(V\\_rest[i])\*sim(V\\_usrt[i]),V\\_rest[k]))}{\sum\_{k=1}^{|V\_{\\_rest}|}freq(V\\_rest[k])} $ (4)

In the above equation, freq (tag) represents the frequency or weight of the particular tag on the result page. $sim(V\\_usrt[i]),V\\_rest[k])$ Represents the similarity between the user tag vector $V\\_usrt\left[i\right]$and the result page tag vector $V\\_rest[k]$ and similarity is defined as given below:

Given a document collection D, a word w, and an individual document d Є D, we calculate

wd=fw,d\*log(|D|/fw,D) (6)

Where fw,d  equals the number of times w appears in d, |D| is the size of the corpus, and fw,D  equals the number of documents in which w appears in D. Words with high wd imply that w is an important word in d but not common in D.

# Experiments and Analysis

The experiments are performed as follows:

* Initially, submit the query to Google, and obtain the original Google search results.
* Now, submit the Google search results to CiteUlike to obtain the relevant tags.
* Re-rank the search results according to our algorithm.
* Compare the Google results with our algorithm.

##  Data Set

### Query Set

Initially, we determine the queries which we input to the search engine. We determine a total of ten queries. The queries are a combination of keywords and tags. These queries are submitted to Google. We have chosen academic domain as CiteUlike provides tags for the academic database only

### Result Set

 Now, submit each query to Google and record the first 50 results. This way, the result set of 10 queries become 500 results.

### Results Tag Set

Now, we submit the 500 results to CiteUlike and the resulting tag vector is recorded. We obtain lots of tag values for a result.

For these queries, we compute the values of normalized DCG gains for Google as well as for our algorithm (Fuzzy JEKS algorithm) in Table1.

Table 1. Normalized DCG gains of Google and our fuzzy JEKS algorithm.

|  |  |  |
| --- | --- | --- |
| **Query** | **Google Ranking** | **Fuzzy JEKS Algorithm Ranking** |
| Q1 | 0.980211 | 0.959274 |
| Q2 | 0.896716 | 0.92342 |
| Q3 | 0.937156 | 0.926431 |
| Q4 | 0.979388 | 0.978542 |
| Q5 | 0.987652 | 0.987472 |
| Q6 | 0.94898 | 0.948706 |
| Q7 | 0.98502 | 0.98638 |
| Q8 | 0.91282 | 0.877635 |
| Q9 | 0.900639 | 0.929049 |

We obtained normalized DCG values for the 10 queries for our algorithm as well as for Google results. It can be seen that our algorithm acquires higher values of normalized DCG for 3 queries out of 10 queries when compared to Google.

# Conclusion

In a professional context it often happens that private or corporate clients corder a publication to be made and presented with the actual content still not being ready. Think of a news blog that's filled with con

Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live the blind texts. Separated they live in Bookmarksgrove right at the coast of the Semantics, a large language ocean. A small river named Duden flows by their place and supplies it with the necessary regelialia. It is a paradisematic country, in which roasted parts of sentences fly into your mouth. Even the all-powerful Pointing has no control about the blind texts it is an almost unorthographic life One day however a small line of blind text by the name of Lorem Ipsum decided to leave for the far World of Grammar. The Big Oxmox advised her not to do so, because there were thousands of bad Commas, wild Question Marks and devious Semikoli, but the Little Blind Text didn’t listen. She packed her seven versalia, put her initial into the belt and made herself on the way. When she reached the first hills of the Italic Mountains, she had a last view back on the skyline of her hometown Bookmarksgrove, the headline of Alphabet Village and the subline of her own road, the Line Lane. Pityful a rethoric question ran over her cheek, then

In the future work, we will further improve the algorithm. Blind texts it is an almost unorthographic life One day however a small line of blind text by the name of Lorem Ipsum decided to leave for the far World of Grammar by adding these effects.

References

1. Kanski, J.J., *Clinical ophthalmology*. 6th edition ed2007, London: Elsevier Health Sciences (United Kingdom). 952.
2. Liang, Z., et al., *The detection and quantification of retinopathy using digital angiograms.* Medical Imaging, IEEE Transactions on, 1994. 13(4): p. 619-626.
3. Teng, T., M. Lefley, and D. Claremont, *Progress towards automated diabetic ocular screening: A review of image analysis and intelligent systems for diabetic retinopathy.* Medical and Biological Engineering and Computing, 2002. 40(1): p. 2-13.
4. Winder, R.J., et al., *Algorithms for digital image processing in diabetic retinopathy.* Computerized Medical Imaging and Graphics, 2009. 33(8): p. 608-622.
5. Heneghan, C., et al., *Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis.* Medical image analysis, 2002. 6(4): p. 407-429.
6. Haddouche, A., et al., *Detection of the foveal avascular zone on retinal angiograms using Markov random fields.* Digital Signal Processing. 20(1): p. 149-154.
7. Narasimha-Iyer, H., et al., *Automatic Identification of Retinal Arteries and Veins From Dual-Wavelength Images Using Structural and Functional Features.* Biomedical Engineering, IEEE Transactions on, 2007. 54(8): p. 1427-1435.
8. Hatanaka, Y., et al. *Automated analysis of the distributions and geometries of blood vessels on retinal fundus images*. SPIE.
9. Foracchia, M., *Extraction and quantitative description of vessel features in hypertensive retinopathy fundus images*, in *Book Abstracts 2nd Int. Workshop on Computer Assisted Fundus Image Analysis*, E. Grisan, Editor 2001. p. 6-6.
10. Xiaohong, G., et al. *A method of vessel tracking for vessel diameter measurement on retinal images*. in *Image Processing, 2001. Proceedings. 2001 International Conference on*.
11. Lowell, J., et al., *Measurement of retinal vessel widths from fundus images based on 2-D modeling.* Medical Imaging, IEEE Transactions on, 2004. 23(10): p. 1196-1204.
12. Hong, S., et al., *Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images.* Information Technology in Biomedicine, IEEE Transactions on, 2001. 5(1): p. 77-91.
13. Pinz, A., et al., *Mapping the human retina.* Medical Imaging, IEEE Transactions on, 1998. 17(4): p. 606-619.
14. Zana, F. and J.C. Klein, *A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform.* Medical Imaging, IEEE Transactions on, 1999. 18(5): p. 419-428.
15. Can, A., et al., *A Feature-Based, Robust, Hierarchical Algorithm for Registering Pairs of Images of the Curved Human Retina.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. 24: p. 347-364.
16. Fritzsche, K., et al., *Automated Model Based Segmentation, Tracing and Analysis of Retinal Vasculature from Digital Fundus Images*, in *State-of-The-Art Angiography, Applications and Plaque Imaging Using MR, CT, Ultrasound and X-rays*2003, Academic Press. p. 225-298.
17. Hoover, A. and M. Goldbaum, *Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels.* Medical Imaging, IEEE Transactions on, 2003. 22(8): p. 951-958.
18. Foracchia, M., E. Grisan, and A. Ruggeri, *Detection of optic disc in retinal images by means of a geometrical model of vessel structure.* Medical Imaging, IEEE Transactions on, 2004. 23(10): p. 1189-1195.
19. Huiqi, L. and O. Chutatape, *Automated feature extraction in color retinal images by a model based approach.* Biomedical Engineering, IEEE Transactions on, 2004. 51(2): p. 246-254.
20. Fritzsche, K.H., *Computer vision algorithms for retinal vessel detection and width change detection*, 2004, Rensselaer Polytechnic Institute: Troy, NY, USA.
21. Houben, A.J.H.M., et al., *Quantitative analysis of retinal vascular changes in essential and renovascular hypertension.* Journal of hypertension, 1995. 13(12).
22. Wasan, B., et al., *Vascular network changes in the retina with age and hypertension.* Journal of hypertension, 1995. 13(12).
23. Koozekanani, D., et al., *Tracking the Optic Nerve Head in OCT Video Using Dual Eigenspaces and an Adaptive Vascular Distribution Model.* Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, 2001. 1: p. 934.