

ETBTRank: Ranking Biterms in Paper Titles for Emerging Topic Discovery



Junfeng Wu, Guangyan Huang, Roozbeh Zarei
School of Information Technology, Deakin University, VIC 3125, AU

Motivation: AI assisted *early discovery of emerging scientific topics* for researchers.

Research problem: Automatic Biterm Extraction (ABE) to rank the importance of biterms on how well they represent emerging topics.

Potential applications:

- optimizing resource allocations for promising research areas
- predicting future scientific topics and technology trends,
- finding knowledge gaps that require new research,
- identifying new concepts in the scientific literature,
- recommending personalized research directions, etc.

Important findings: we have found an effective way to rank the importance of biterms on how well they represent emerging scientific topics.

The score of ETBTRank is defined by the following equation:

$$S_{\text{predict}}(b) = \frac{S(t_1)S(t_2)}{\text{sim}(t_1, t_2)}, \quad (1)$$

where:

- $S(t_1)$ and $S(t_2)$ are the ATE scores of the two terms t_1 and t_2 in the biterm b , computed using any chosen ATE model;
- $\text{sim}(t_1, t_2)$ is the similarity of the two terms, which can be computed easily using, for example, Jaccard Similarity that captures the overlapping ratio of their neighbor-term sets.

Example: how to analyze a paper's critical biterm "unet" + "image segmentation" (we highlight the the synonyms of the two terms in this biterm):

Title: **Nas-unet**: Neural architecture search for **medical image segmentation**

Some references:

U-Net Based Architecture for an Improved **Multiresolution Segmentation** in Medical Images

BUSU-Net: An Ensemble **U-Net** Framework for **Medical Image Segmentation**

CNL-U-Net: A novel lightweight deep learning architecture for multimodal **biomedical image segmentation** with false output suppression

ENAS U-Net: Evolutionary Neural Architecture Search for **Retinal Vessel Segmentation**

Transfer Learning **U-Net** Deep Learning for **Lung Ultrasound Segmentation**

TMD-U-Net: **Triple-U-Net** with Multi-Scale Input Features and Dense Skip Connection for **Medical Image Segmentation**

Each term *has lots of synonyms* in the references, but the co-occurrence of the two terms is *rare*, implying accelerating cooperation of two research fields, thus the topic is probably an emerging topic!

Experiments: comparison of our ETBTRank model to the counterparts in two metrics. ETBTRank outperforms the counterparts in both.

