

# Micropinion Dataset

## *Micropinion Generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions*

Author: Kavita Ganesan ([kganes2@illinois.edu](mailto:kganes2@illinois.edu))

### Dataset Overview

This dataset is based on user reviews from [CNET](#). The reviews are on products from various categories like **tv**, **cell phones**, **gps** etc. You will find two versions of the dataset :- "**raw**" and "**pre-processed**". The "raw" folder has the original reviews from CNET without any pre-processing (each review is delimited by "\$\$;"). The "pre-processed" folder contains sentences from the [full review](#) section of the reviews. All the pros and cons from the original reviews are omitted in this version and this was the version used for summarization (See Section 5.1 of paper). In addition in the pre-processed version, a simple sentence splitter was used to split the review texts into different sentences.

### Citation Request

---

If you use this dataset for your own research, please cite the following paper:

Kavita Ganesan, ChengXiang Zhai, and Evelyne Viegas. 2012. [Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions](#). In *Proceedings of the 21st international conference on World Wide Web (WWW '12)*. ACM, New York, NY, USA, 869-878.

#### **Bibtex:**

```
@inproceedings{Ganesan:2012:WWW,
  author = {Ganesan, Kavita and Zhai, ChengXiang and Viegas, Evelyne},
  title = {Micropinion generation: an unsupervised approach to
generating ultra-concise summaries of opinions},
  booktitle = {Proceedings of the 21st international conference on
World Wide Web},
  series = {WWW '12},
  year = {2012},
  isbn = {978-1-4503-1229-5},
  location = {Lyon, France},
  pages = {869--878},
  numpages = {10},
  url = {http://doi.acm.org/10.1145/2187836.2187954},
  doi = {10.1145/2187836.2187954},
  acmid = {2187954},
  publisher = {ACM},
  address = {New York, NY, USA},
}
```