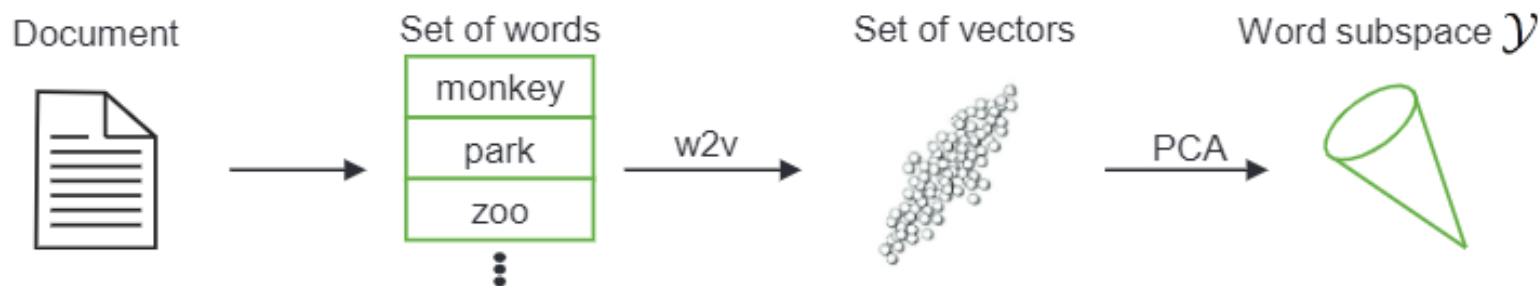


Subspace Representation for Natural Language Processing

Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

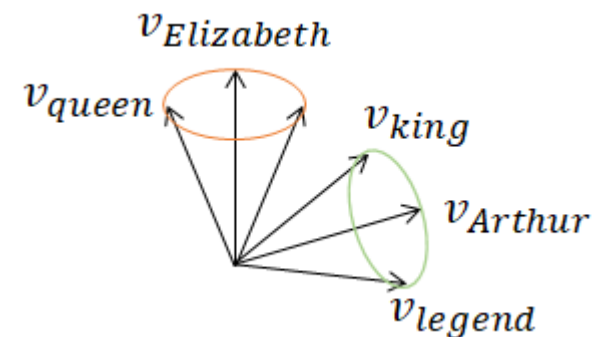
* University of Tsukuba, Japan

+ University of Grenoble Alpes, France



■ **Word Subspace**: Compact, scalable and meaningful representation of the context of a set of words.

- ▣ Low computational cost;
- ▣ No limitation of number of words;
- ▣ **Interpretable**: Basis vectors can be regarded as the **main hidden topics** of a text.



Subspace Representation for Natural Language Processing

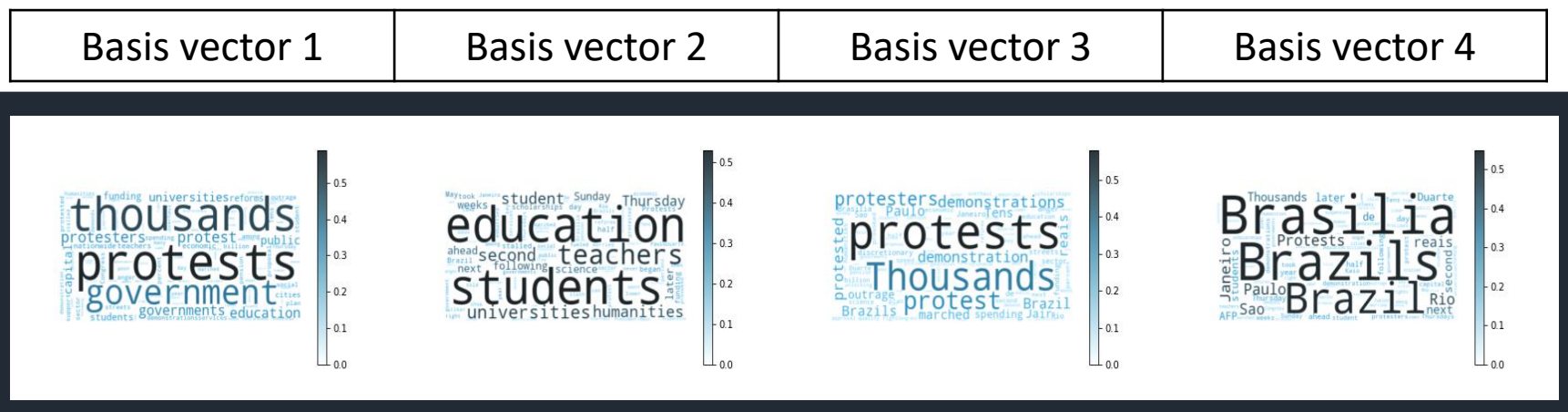
Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

* University of Tsukuba, Japan

+ University of Grenoble Alpes, France

- Text analysis based on the subspace representation.
- Word Importance Score: Measures how relevant a word w_k is with regards to a hidden topic ϕ_i .

$$I(x_k, \phi_i) = x_k^\top \phi_i$$



Most important words with respect to the main hidden topics of a text about protests against cuts in education funding in Brazil.

Subspace Representation for Natural Language Processing

Erica Kido Shimomoto*, François Portet+, Kazuhiro Fukui*

* University of Tsukuba, Japan

+ University of Grenoble Alpes, France

- Geometrical interpretation of the word vectors -> Propose the best subspace-based method to solve each task.
- Topic Classification: MSM.
- Sentiment Analysis: GOSM.

