

VISUALIZING TEXT USING t-SNE

Jasneet Sabharwal
SFU Natural Language Lab

Visualizing Text

Step 1: Analyze the Text & Extract Features

Mary sold the book to John

Bag of Words

noun
Mary verb the noun to noun
sold book John
det.

Part of Speech

seller
Mary sold the book to John
predicate theme recipient

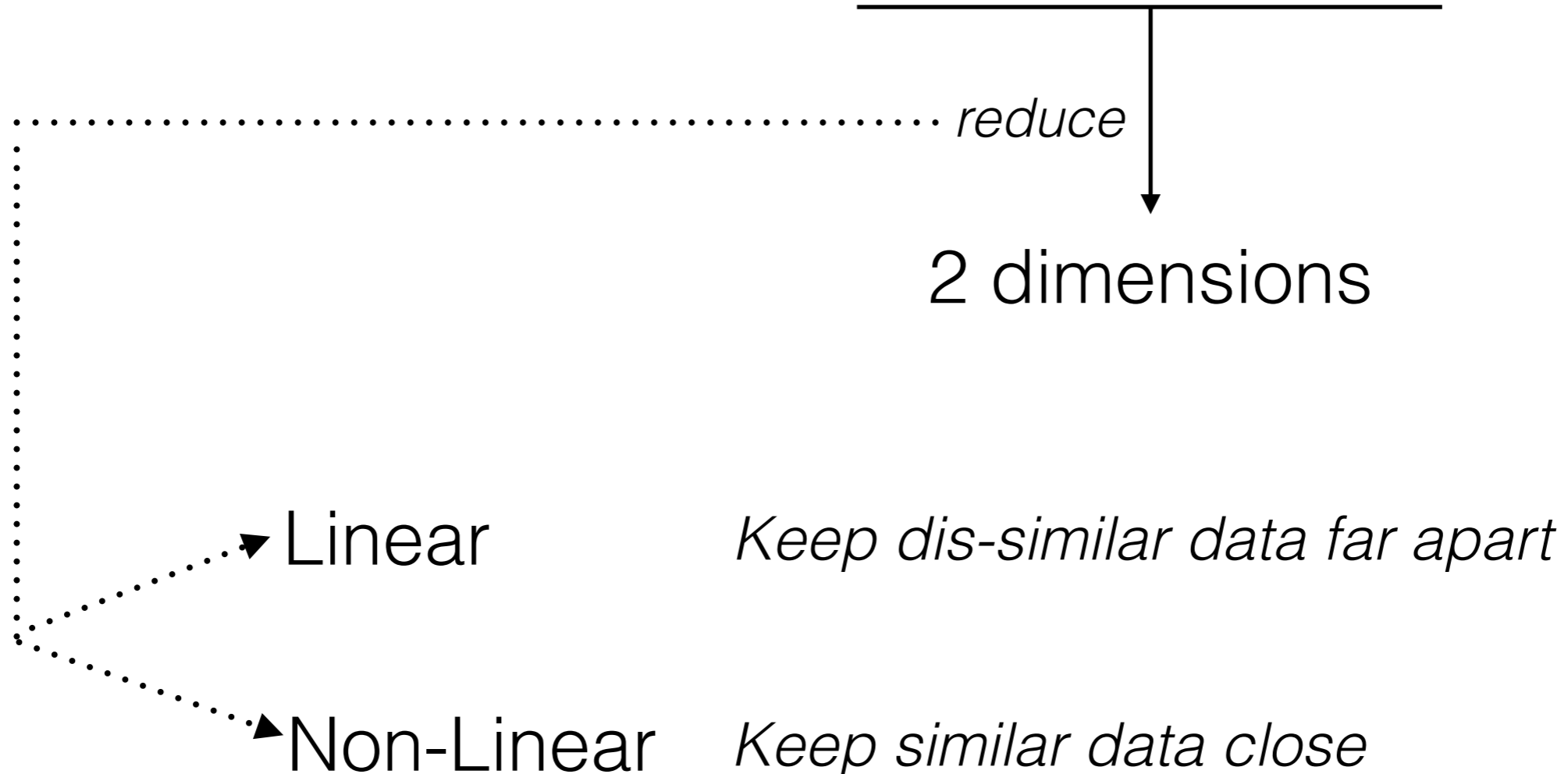
Semantic Role
Labels

1000 to 1,000,000 Features Extracted

Visualizing Text

Step 2: Visualize

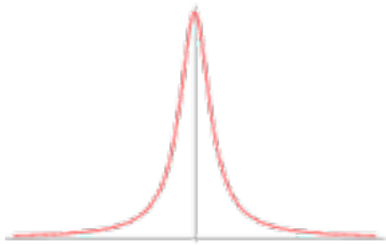
How do we visualize 1000+ dimensions?



t-Distributed Stochastic Neighbor Embedding*



Use Gaussian distribution to calculate probability of data points in higher dimension



Use Student-t distribution to represent points in lower dimension

Map points in higher dimension to lower dimension by minimizing Kullback-Leibler divergence using gradient descent

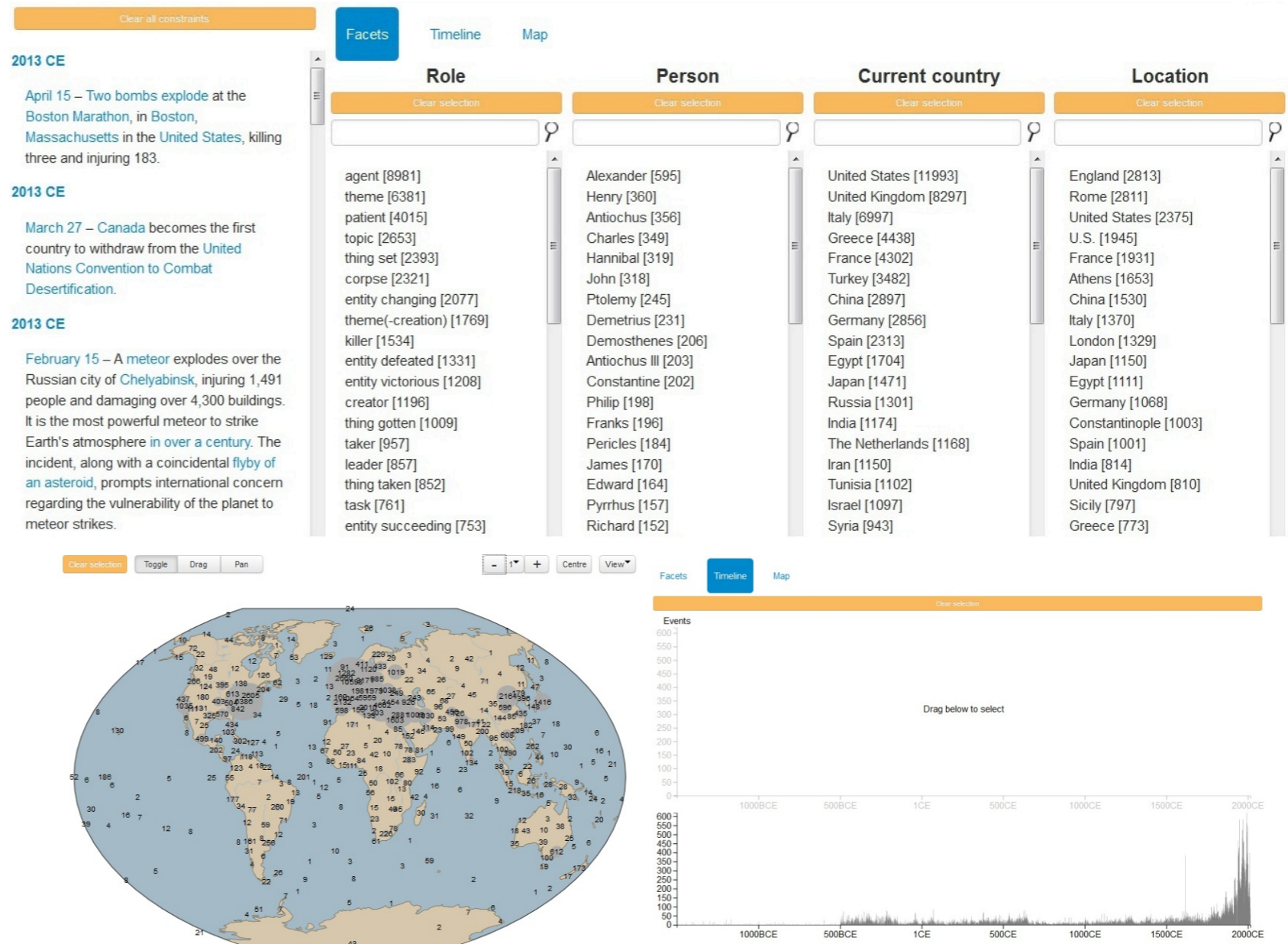
Uses Barnes-Hut approximation**

*Visualizing High-Dimensional Data using t-SNE: L.J.P. van der Maaten, Geoffrey E. Hinton, 2008

**Barnes-Hut-SNE: L.J.P. van der Maaten, 2013

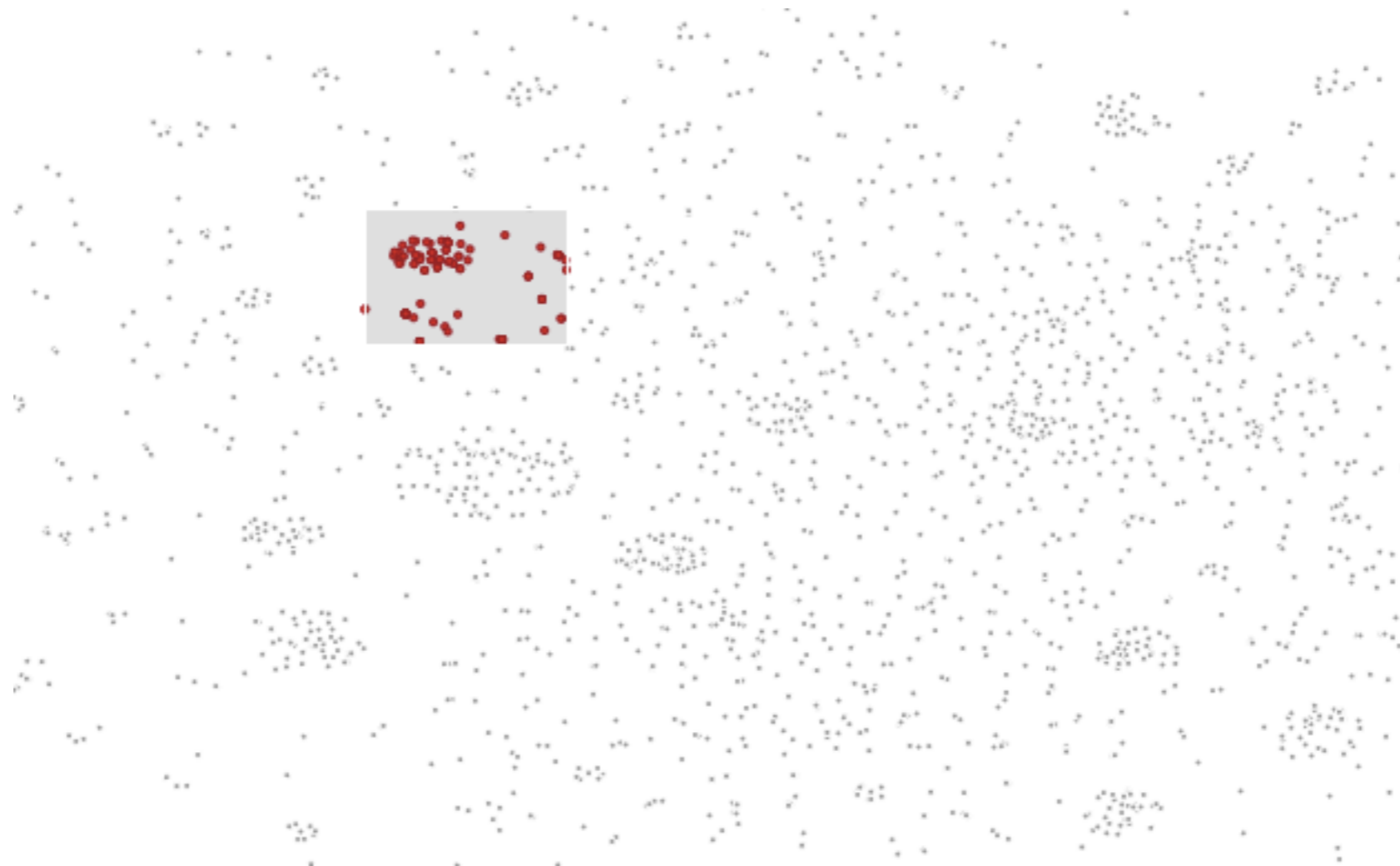
Lensing Wikipedia

Lensing Wikipedia



t-SNE in Lensing Wikipedia

- Predicates (Verbs) from Semantic Role Labelling on the text used as features for each document
- Use t-SNE to create 2-Dimensional Scatter Plot of historical human events from Wikipedia
- Well formed clusters of historical events based on the type of events



DEMO

THANK YOU