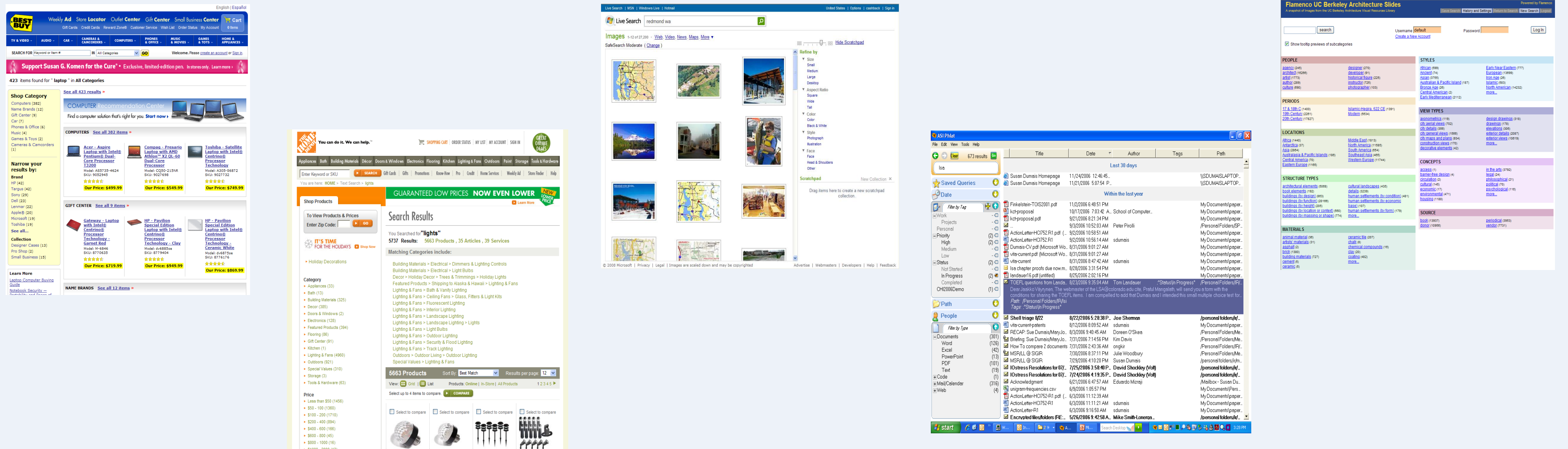# Challenges for Supporting Faceted Search in Large, Heterogeneous Corpora like the Web

Jaime Teevan, Susan Dumais, Zachary Gutt
Microsoft
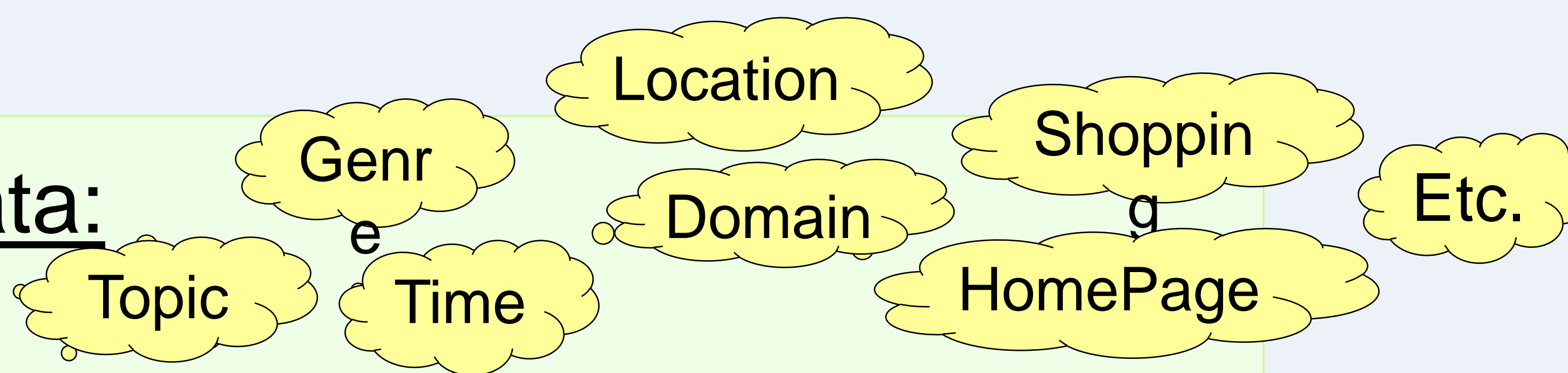
Faceted Search Today – Examples

Challenges with the Web – Scale, Heterogeneity

Generating metadata:

- Which facets?
- Scale (docs and facets)
- Accuracy vs. coverage
- Uncertainty – filtering vs. ranking vs. interacting

Topic · Genre · Time · Location · Domain · HomePage · Shopping · Etc.

Identifying which facets to surface:

- Predictability vs. relevance to query
- Discriminability vs. relevance to searcher

Computing previews:

- Accurately predicting counts, without examining all results

Microsoft