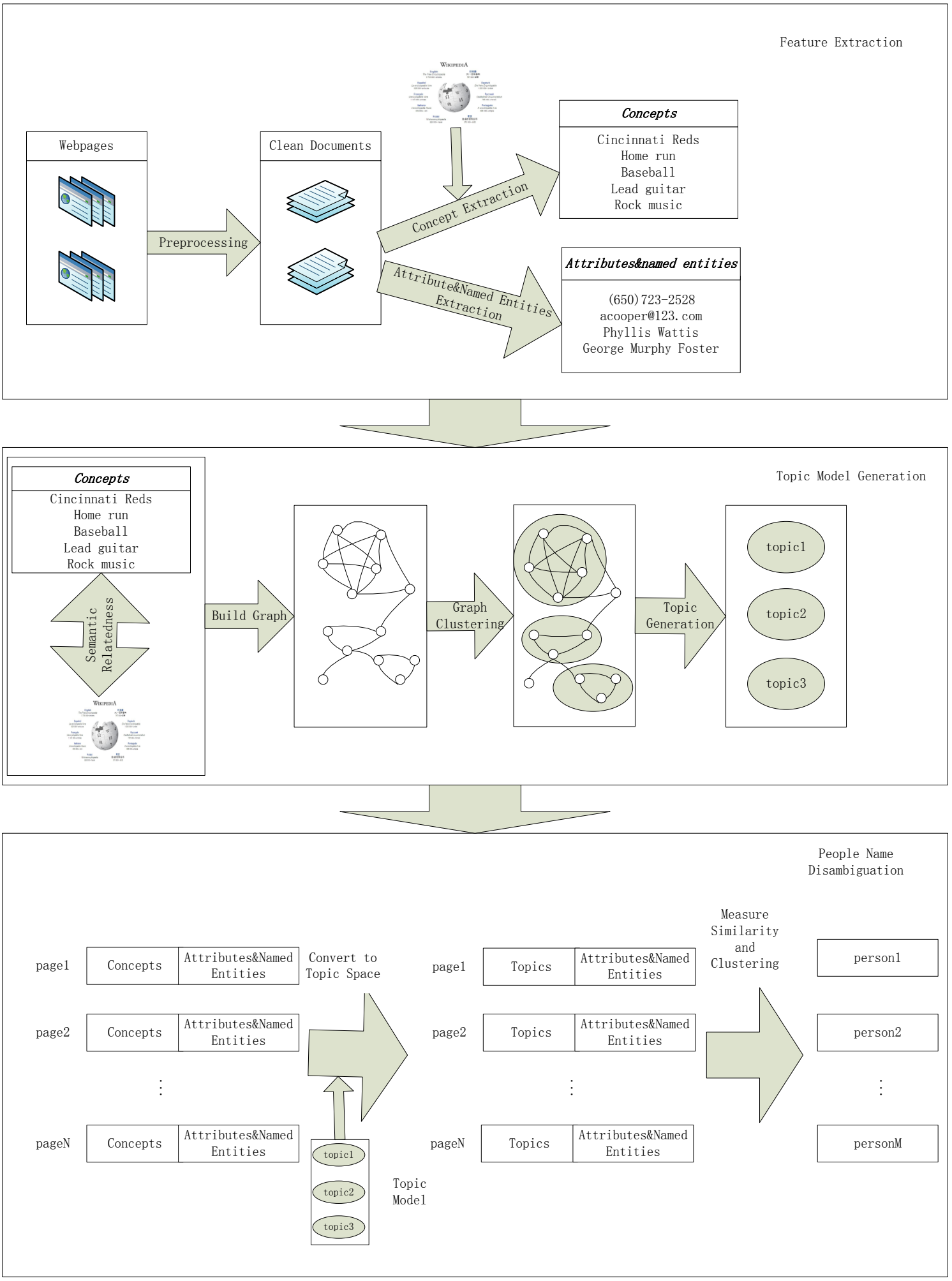
Feature Extraction
Webpages
Preprocessing
Clean Documents
Concept Extraction
Concepts
Cincinnati Reds
Home run
Baseball
Lead guitar
Rock music
Attribute&Named Entities Extraction
Attributes&named entities
(650)723-2528
acooper@123.com
Phyllis Wattis
George Murphy Foster
Topic Model Generation
Concepts
Cincinnati Reds
Home run
Baseball
Lead guitar
Rock music
Semantic Relatedness
Build Graph
Graph Clustering
Topic Generation
topic1
topic2
topic3
People Name Disambiguation
page1
Concepts
Attributes&Named Entities
page2
Concepts
Attributes&Named Entities
pageN
Concepts
Attributes&Named Entities
Convert to Topic Space
topic1
topic2
topic3
Topic Model
page1
Topics
Attributes&Named Entities
page2
Topics
Attributes&Named Entities
pageN
Topics
Attributes&Named Entities
Measure Similarity and Clustering
person1
person2
personM