

Composite Items

LIG - SLIDE

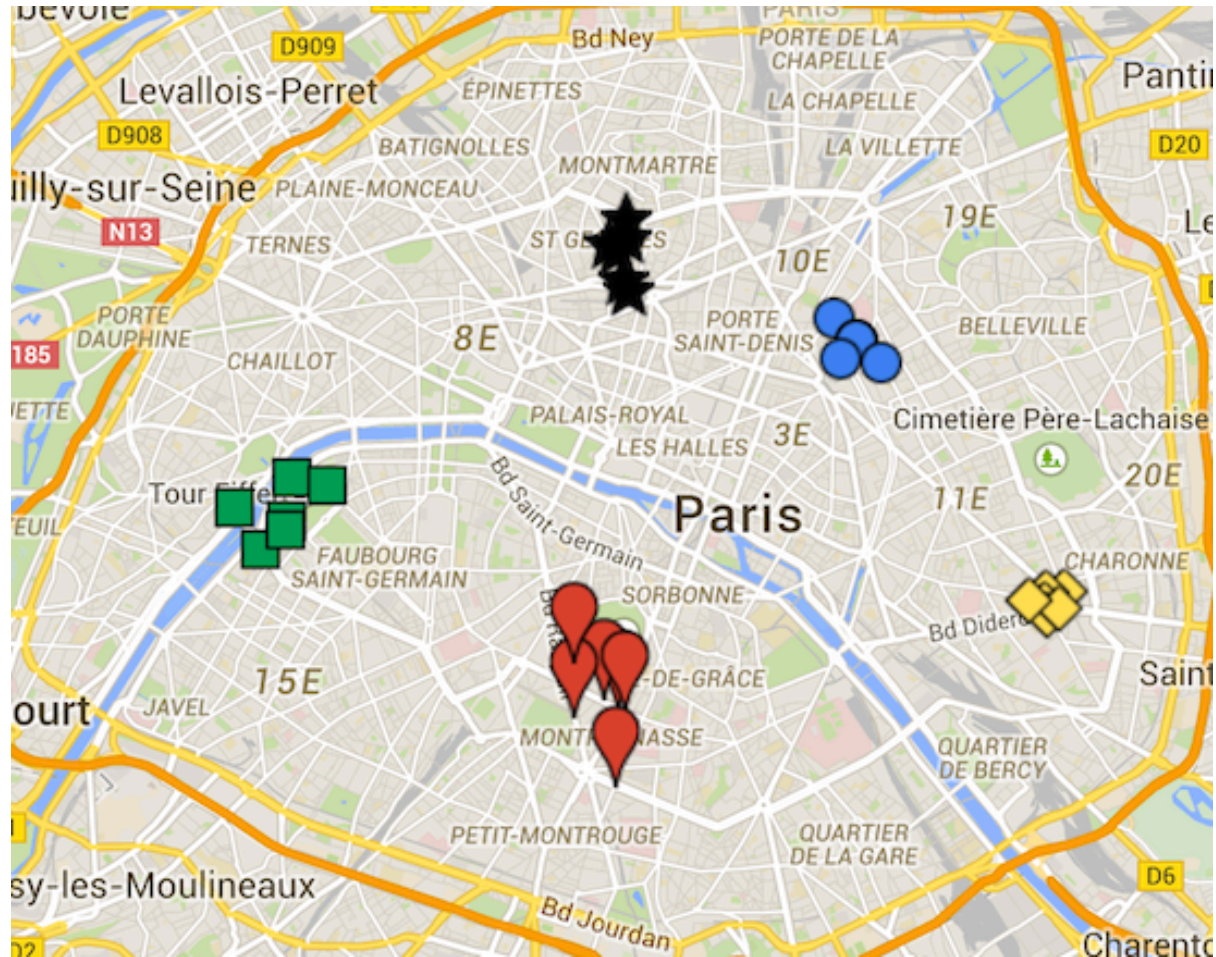
Motivation

- Summarize large collections of heterogeneous items
 - Representative
 - Cohesive
 - Valid
 - Personalized
 - ...

Representative Composite Items

- Budget vector
 - 2 public transportations
 - 2 restaurants
 - 2 health services

Composite POI over Paris



Composite Tasks for AMT workers

- Budget vector
 - 2 tasks from last week
 - 1 task from yesterday
 - 1 task from today
 - Minimum reward \$5

Composite Tasks on AMT

2 3 4 5 6 7 8 9

Please select which set contains tasks you would be more interested in completing.

Set A

Set B

Review media transcription for content accuracy. A2290075 (Audio length: 1 hour 18 minutes 56 seconds)
Requester: Speechpad
Description: Review Audio transcription for content accuracy.
Keywords: voicemail, transcription, transcribe, SpeechInk, review
Reward: 1421.0

Edit phone order audio transcript A2340605 (Audio length: 20 seconds)
Requester: Speechpad
Description: Review Audio transcription for content accuracy.
Keywords: pizza, voicemail, transcription, transcribe, SpeechInk, on-line orders, review
Reward: 9.0

Review media transcription for content accuracy. A2329057 (Video length: 1 minute 44 seconds)
Requester: Speechpad
Description: Review Video transcription for content accuracy.
Keywords: voicemail, transcription, transcribe, SpeechInk, review
Reward: 31.0

Review media transcription for content accuracy. A2286060 (Audio length: 49 minutes 27 seconds)
Requester: Speechpad
Description: Review Audio transcription for content accuracy.
Keywords: voicemail, transcription,

Review media transcription for content accuracy. A2290075 (Audio length: 1 hour 18 minutes 56 seconds)
Requester: Speechpad
Description: Review Audio transcription for content accuracy.
Keywords: voicemail, transcription, transcribe, SpeechInk, review
Reward: 1421.0

Edit phone order audio transcript A2340605 (Audio length: 20 seconds)
Requester: Speechpad
Description: Review Audio transcription for content accuracy.
Keywords: pizza, voicemail, transcription, transcribe, SpeechInk, on-line orders, review
Reward: 9.0

Review media transcription for content accuracy. A2329057 (Video length: 1 minute 44 seconds)
Requester: Speechpad
Description: Review Video transcription for content accuracy.
Keywords: voicemail, transcription, transcribe, SpeechInk, review
Reward: 31.0

Review media transcription for content accuracy. A2286060 (Audio length: 49 minutes 27 seconds)
Requester: Speechpad

General Framework

- Rely on fuzzy clustering to achieve representativity through centroids
- Select 1 CI around each centroid
 - Optimize secondary dimensions
 - Min/max price (aggregation)
 - Min/max per category
 - Diversity of attributes
 - ...

Using CIs on Vodkaster

- As a query
 - Budget vector
- More flexible
 - Encourage diversity of attributes without specifying categories explicitly
ex: different actors / directors
- Keep using ratings for distance computation using cosine similarity
 - Problem of sparsity