Motivation
- Discover open-ended contextual information from user reviews. Context doesn’t have to be predefined.
- Incorporate the open-ended contextual information together with ratings into a context sensitive recommender system.
- Make recommendations that better satisfy user goals.

Proposed Solution
- Our system is called Rich Context (RC) and its has two main components: RCMiner and RCR recommender.

Assumptions
- Specific reviews describe experiences.
- Generic reviews give general reviews of products.
- Specific reviews contain more contextual information than generic ones.

Classifying Reviews
- We separate specific and generic reviews using a Logistic Regression classifier.
- Feature subset selection showed that the number of words and the number of verbs in past tense in a review are the most relevant features.

Topic modeling
- We use Latent Dirichlet Allocation (LDA) to find the topics that compose the reviews.

Example
Specific review
"During the summer, we like to take a mini [staycation]. This year it was extra special as we also got engaged. Our stay at the Biltmore was just fantastic. The service exceptional, the food amazing- it was great at the pool. Wrights and also at Frank and Alberts. The only reason I am not giving it a full 5 stars is the 'upgraded' room was just a nice basic room. Though it was certainly nice, it wasn’t what I expected for being the Biltmore. However, everything else certainly lived up to that expectation."

Generic review
"Nice hotel, all the amenities you need, great complex of pools. Just make sure your room is as far from the Vista Lounge as possible; otherwise you’ll be bombarded with crappy live music, fully audible from the lounge to all the surrounding rooms above it, for four hours a day. Horrible. The Internet access is not free, which is lame. The room service is good but overpriced by at least 40 percent. So it goes with resorts. Otherwise a very nice hotel."

RCMiner
- The filtering process looks at which topics appear more frequently in specific reviews than in generic ones. Those topics are then marked as contextual topics.

RCRecommender
- The context of a review is represented as a vector with the probability distribution of contextual topics.
- The user queries the system giving text as an input, which is transformed into context vector.
- The neighbourhood is composed of users who have rated items in contexts similar to the user.
- Finally, the predicted rating is a weighted aggregate:

\[ P_{u,i,c} = \bar{\rho}(u, \epsilon_3) + \frac{\sum_{\epsilon \in N_u} \left( r_{ui} - \bar{r}_{ui} \right) \times usim(u, n, \epsilon_3)}{\sum_{\epsilon \in N_u} usim(u, n, \epsilon_3)} \]

Experiments
- Yelp Data Challenge dataset of hotels (5034 reviews) and spas (5579 reviews).
- 5-fold cross validation.
- Compared against a k-NN user-based baseline recommender.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Top N</th>
<th>RMSE</th>
<th>Coverage</th>
<th>Change over UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>Recall</td>
<td>RMSE</td>
<td>Coverage</td>
<td></td>
</tr>
<tr>
<td>Hotels</td>
<td>0.56</td>
<td>1.30</td>
<td>8%</td>
<td>+49.5% +13.7% +74.6%</td>
</tr>
<tr>
<td>Spas</td>
<td>0.51</td>
<td>1.32</td>
<td>3%</td>
<td>+59.6% +18.2% +34.8%</td>
</tr>
</tbody>
</table>

Future Work
- Improve RCMiner by designing new features and using different classifiers.
- Improve the coverage of RCR recommender on sparse datasets.