

# Opinions Corpora Acquisition Software - demo

Zygmunt Vetulani (1), Suleyman Menken (2) and Marta Witkowska (1)

(1) Adam Mickiewicz University in Poznań, Poland

(2) University of Kocaeli, Turkey

## Abstract

Opinion processing is, since recently, in focus of interest for computational linguists, public relation experts, marketing companies and politicians. Studies of natural language expression of opinions, desires, emotions and related phenomena require tools and methodologies. Within the empirical approach to these studies, we propose Opinions Corpora Acquisition Software (OCAS), a tool for collection of the empirical data in form of a corpus. The tool is used to collect corpora of the customers' written opinions concerning widely used on-line booking services in the area of hotel reservation (via Booking.com).

**Keywords:** text language resources, opinion processing, corpus-based methods, multilingual corpora, customers' opinions, OCAS



Figure 1. An example of the Booking.com opinion record completed by the guest

Below (Figure 2. a-f) we present the screenshots of Opinion Corpora Acquisition Software (OCAS)

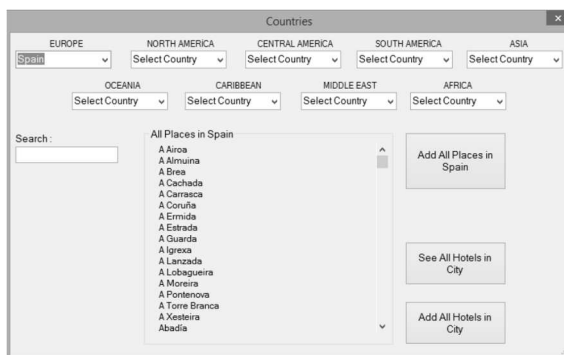
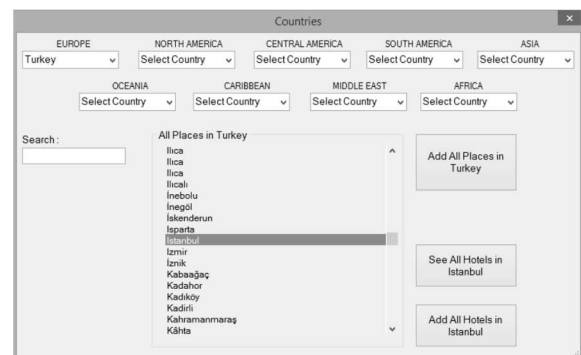
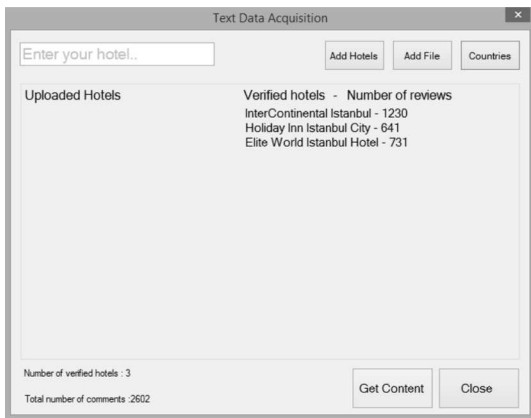


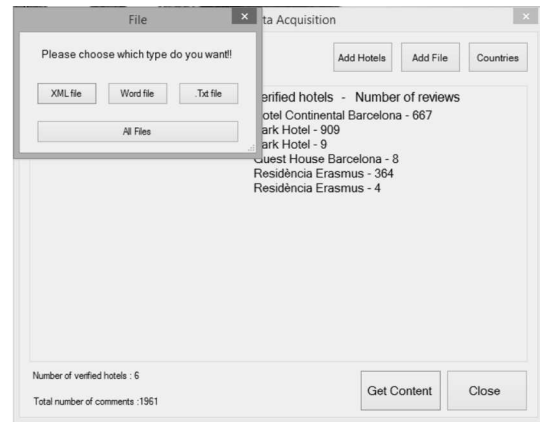
Figure 2 a).



b).



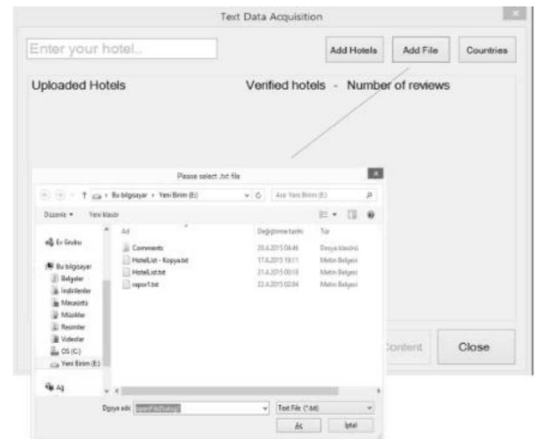
c).



d).



e).



f).

Format of final data present Figure 3 to 6.

```

<?xml version="1.0" encoding="utf-8"?>
<AllComments>
  <Hotel name="InterContinental Istanbul">
    <User name="Sharon">
      <Grade>8.3</Grade>
      <Assesment>Loved your hotel but restaurants too pricey for what was offered</Assesment>
      <Negative>. now that we are retired - found out too expensive to dine in your restaurants</Negative>
      <Positive>Loved your staff. especially concierge and the outside people with help with taxis - felt very safe</Positive>
    </User>
    ...
    <User name="Susan">
      <Grade>9.6</Grade>
      <Assesment>"Beautiful"</Assesment>
      <Negative>Confusion over booking as I included breakfast for 3 nights and this was confirmed... However they said I didn't pay for this
      and they tried to do on the morning of our arrival without my permission I was not happy about this at all</Negative>
      <Positive>Location and the city view incredible</Positive>
    </User>
    <User name="Sharon">
      ...
    </Hotel>
  </Hotel>
  <Hotel name="Holiday Inn Istanbul City">
    <User name="Isimsiz">
      <Grade>7</Grade>
      <Assesment>"Worth it"</Assesment>
      <Negative>Small room, slow wifi and no free drinking water bottles</Negative>
      <Positive>Location, Friendly staff, Clean and Quite</Positive>
    </User>
    ...
  </Hotel>
  <Hotel name="Elite World Istanbul Hotel">
    ...
  </Hotel>
</AllComments>

```

Figure 3. A fragment of the XML formatted corpus presenting three different opinions about hotels in Istanbul.



Figure 4.

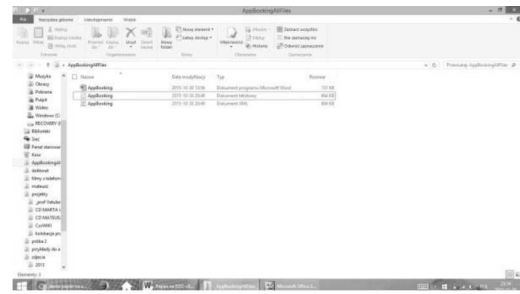


Figure 5.

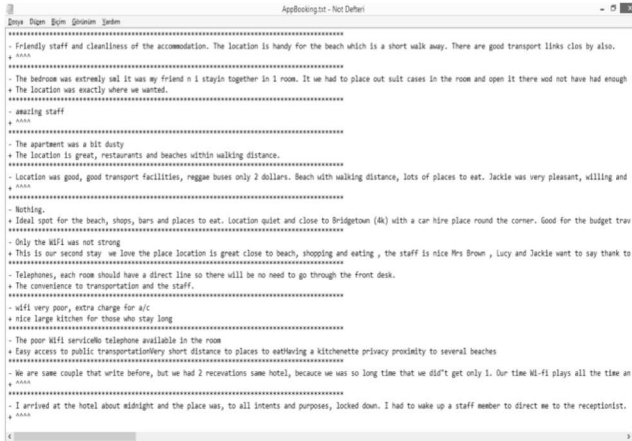


Figure 6.

**Bibliography:**

Vetulani, Z., Witkowska, M., Menken, S.(2015): Corpus Based Studies on Language Expression of Opinions, Proceedings of LTC 2015, Poznań.