**UDC 81.38**

**A. Romaniuk, M. Romanyshyn**
Lviv Polytechnic National University
Computer-Aided Design Department

# NAMED-ENTITY RECOGNITION FOR SENTIMENT ANALYSIS OF UKRAINIAN REVIEWS

**This paper describes the necessity of named-entity recognition for the implementation of sentiment analysis and presents methods and tools for recognition of appropriate named entities in Ukrainian restaurant reviews.**

**Key words: sentiment analysis, named-entity recognition, regular expressions, part-of-speech tagging.**

**Йдеться про необхідність виділення іменованих сутностей для вирішення завдання емоційно-смислового аналізу. Також наведено методи та засоби для виділення потрібних іменованих сутностей в україномовних відгуках про заклади харчування.**

**Ключові слова: емоційно-смисловий аналіз, виділення іменованих сутностей, регулярні вирази, морфологічний аналіз.**

## 1. Problem

Sentiment analysis, or opinion mining, is a kind of text analysis, which aims to identify emotional attitudes or subjective judgments of the author concerning a particular object in the text message. The last part of the definition says that the sentiment itself should be defined with relation to a certain object. Very often the object that the sentiment is associated to is represented in the form of a named entity.

The recognition of named entities is very important for accurate sentiment detection and object-sentiment association. Unfortunately, there is no available named-entity recognition system for Ukrainian language yet. This research solves a purely domain-oriented part of this task specifically for sentiment analysis.

## 2. Recent Research Analysis

Sentiment analysis proved to be a really popular topic for research during the previous decade. This can be proven with a large number of projects, which appear every day: sentiment analysis of hotel reviews [3], bank reviews [2], restaurant reviews, comments on movies [7], products, messages about political events in blogs and social networks, etc. Text reviews in all of these domains contain certain named entities that can potentially play the role of the object of tonality: names of hotels, names of banks, names of restaurants, movie or book titles, names of products, people’s names, etc. Nevertheless, barely few sentiment analysis systems identify sentiments with relation to an entity and not sentiments for the whole review. The examples of the first can be Semantria [4], Skyttle [5] or Bitext [1]. Figure 1 shows the entities recognised by Bitext API demo and Figure 2 shows sentiment information assigned to them.

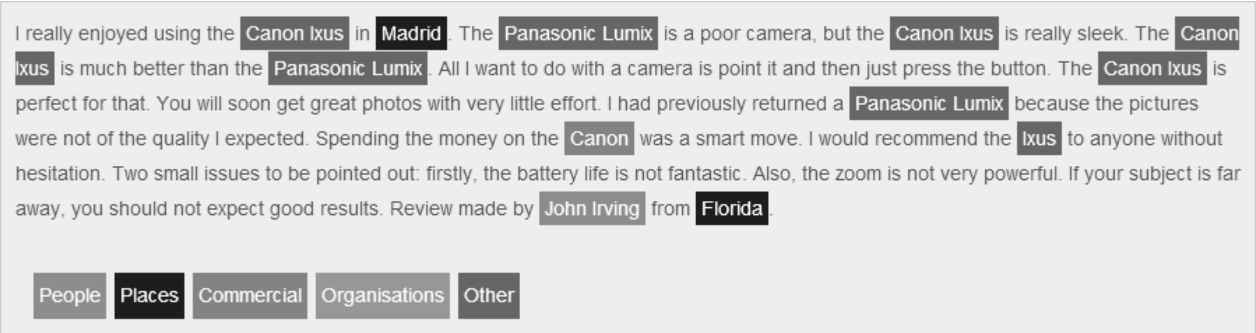
Although hard to implement, the resolution of object-sentiment association makes it possible to identify different sentiments within one sentence with reference to appropriate objects. However, this is impossible to implement without named-entity recognition.

## 3. Research Aims

The aim of this research is to implement named-entity recognition for restaurant reviews.

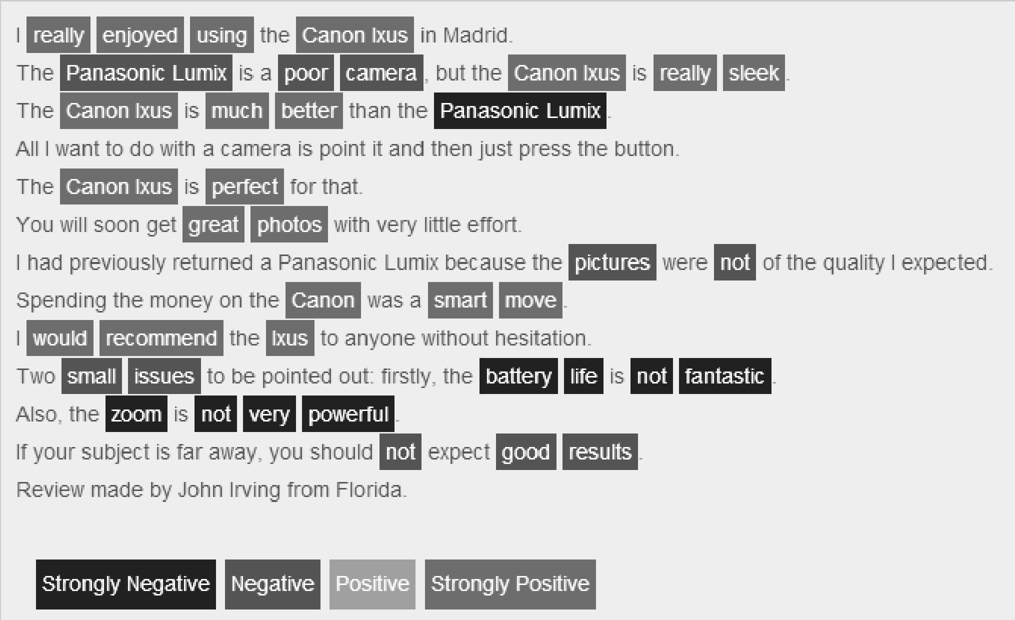
The objectives of the research are the following:

- justify the necessity of named-entity recognition for the task of sentiment analysis;
- define types of named entities for restaurant reviews;
- define appropriate tools for domain-based named-entity recognition.



I really enjoyed using the Canon Ixus in Madrid. The Panasonic Lumix is a poor camera, but the Canon Ixus is really sleek. The Canon Ixus is much better than the Panasonic Lumix. All I want to do with a camera is point it and then just press the button. The Canon Ixus is perfect for that. You will soon get great photos with very little effort. I had previously returned a Panasonic Lumix because the pictures were not of the quality I expected. Spending the money on the Canon was a smart move. I would recommend the Ixus to anyone without hesitation. Two small issues to be pointed out: firstly, the battery life is not fantastic. Also, the zoom is not very powerful. If your subject is far away, you should not expect good results. Review made by John Irving from Florida.

*Fig. 1. Entities recognised by Bitext API demo*



I really enjoyed using the Canon Ixus in Madrid.
The Panasonic Lumix is a poor camera, but the Canon Ixus is really sleek.
The Canon Ixus is much better than the Panasonic Lumix.
All I want to do with a camera is point it and then just press the button.
The Canon Ixus is perfect for that.
You will soon get great photos with very little effort.
I had previously returned a Panasonic Lumix because the pictures were not of the quality I expected.
Spending the money on the Canon was a smart move.
I would recommend the Ixus to anyone without hesitation.
Two small issues to be pointed out: firstly, the battery life is not fantastic.
Also, the zoom is not very powerful.
If your subject is far away, you should not expect good results.
Review made by John Irving from Florida.

*Fig. 2. Sentiments assigned to entities recognised by Bitext API demo*

## 4. Main Part

### 4.1. The Place of Named-Entity Recognition within Sentiment Analysis

Named-entity recognition is one of the stages of natural language processing (NLP), whose task is to identify and categorize specific set of named entities. The most common types of entities that are recognized by systems of named-entity recognition are company names, people's names, names of location, product names, etc. Named-entity recognition, which is now used by many natural language processing systems, is currently a part of such commonly used NLP libraries as nltk or OpenNLP (for English).

There are two main reasons for named-entity recognition within sentiment analysis. First, named entities may include common words that possess a certain polarity. Of course, we do not want the sentiment information, contained in the names, to influence the calculation of general or object-oriented sentiments. Consider the following examples of company names that contain sentiment information are: *«Веселий Роджер», «Крива Липа», «Старий рояль»*, etc. Secondly, the recognition of named entities helps identify the objects that the sentiments are associated to, which means that when we have the situation, like *«На відміну від Organization1, Organization2 не обманює клієнтів»*, we will be able to identify a positive sentiment for *Organization2* and a negative one for *Organization1*.

Unfortunately, the recognition of named entities has not been implemented for Ukrainian language yet, and full implementation of such a tool is beyond the scope of this study. Though, the recognition of domain-specific named entities, commonly used in restaurant reviews, has been implemented.

### 4.2. Types of Named Entities in Restaurant Reviews and Their Recognition

The material for our study consists of Ukrainian restaurant reviews. The most common named entity type used in these reviews is the names of institutions (restaurants, cafes, bars, etc.). Addresses and names of dishes also occur but less frequently.

The recognition of the above mentioned entity types has been implemented in two stages:

- named entity identification with the help of regular expressions;
- named entity categorization with the help of right and left context analysis.

The first stage uses regular expressions to identify entities using their distinct explicit features, like capital letters, different quotation marks, non-alphabetical symbols or letters of a different alphabet, etc.

The aim of the second stage is, first of all, to decide on the type of entity. This stage is also used to find whether one-word entities that come at the beginning of the sentence really are named entities, and to detect entities written in lowercase. To make it easier to analyse the context we run each review through a part-of-speech tagger. This provides each word's lemma and all morphological information that may be necessary. The part-of-speech tagger UGTag [6] used in this research was developed within the project of Ukrainian-Polish parallel corpus led by Natalia Kotsyba and is now available for download.

Sections 4.2.1, 4.2.2 and 4.2.3 provide examples of named entities for the domain of restaurant reviews.

#### 4.2.1. Recognition of Institution Names

Consider the following examples of institution names grouped according to the explicit features used to recognize them on the first stage.

1) the name of the institution wrapped into quotation marks:

*А я "Япону хату" люблю. Там все приємне - і іжа, і персонал...*

*Найпозитивніші враження в мене залишилися після "Під Золотою Розою".*

2) the name of the institution with a capital letter:

*Ще гарно в Чорному Коті, Амадеі, нещодавно відкрили для себе Цитадель.*

*Я люблю Деліс, Кумпель (хоч можливо і не зовсім корисно), Динамо-блюз (ми їхні клієнти вже сто років, святкувала там цього року велике ДН, було шикарно і знижка ого), Цукерню (100% не Вероніку), Євроготель (живу поруч, ресторан пристойний), Старгород (не смачно, на мою думку: перепробувала більшість страв, а голонка в Європі взагалі не зрівняна), але весело, цікавий інтер"єр, і дітей бавлять "до ночі".*

3) the name of the institution with all letters capitalized:

*В п"ятницю ходили з кумами в паб СТАРГОРОД,там була вечірка до дня захисника вітчизни...*

*Гуляла я по центру з подружкою, рішили купити суші та піти додому їх з"їсти. Зайшли в ЯПІ, заказали роли.*

4) the name of the institution containing letters of a different alphabet:

*Вчора приємно посиділи з чоловіком і друзями в Mons Pius.*

*були з сестрою кілька днів тому в піцерії «Белла Чао» ( «Bella ciao») на Вірменській - сподобалось все, піца кльова, смачнезна, велика, і ще багато ахів, дівчата-офіціантки працюють майже непомітно і дуже швидко приносять замовлення - нарешті я знайшла для себе ідеальну піцерію.*

5) the name of the institution containing non-alphabetical symbols:

*досить хороший заклад "Тиса +", у вашому районі, можете зайти подивитися, дуже хороші відгуки чула про нього!*

*сьогодні нарешті відвідала "Кафе №1" :) така тут реклям того закладу, шо-м не втояла, пішла*

#### 4.2.2. Recognition of Location Names

Consider the following examples of location names. The reviews under research tend to contain just the names of the streets and single names of cities, and not full addresses.

*А мені найприємнішою є "Штука"на Котлярській.*

*Якщо вам раптом заманеться перенести пізні домашні посиденьки з друзями/родичами в якусь кнайпу Львова і ваш вибір зупиниться на кафе "Цісар", що на вул.С.Бандери, то будьте обережні*

### 4.2.2. Recognition of the Names of Dishes

This type of entities is very domain-specific. Nevertheless, most of the time the names of the dishes are easy to recognize because of a general word from food vocabulary used right before the named entity itself. Consider the following examples of the names of dishes.

*Старший часто бере салат "Дари моря" з диким рисом, я пробувала, смачно!*

*Голодні були аки вовки, тож назамовляли всього багацько, а саме: пиво, коктейль "Сєкс на пляжі" (для мене), сало з грінками....*

## Conclusion

Named-entity recognition represents an important task in the field of natural language processing. It is also a necessary supplement for sentiment analysis systems as it prevents incorrect sentiment calculation and enables the researcher to detect objects that the sentiments are associated to, which leads to the possibility of deep sentiment analysis systems implementation.

This article described methods and tools used for domain-specific named-entity recognition. The types of entities commonly used in Ukrainian restaurant reviews have been identified. The stages of named-entity recognition have been defined: named entities identification and categorization. The distinct features of named entities have been singled out.

*1. Bitext API demo. – Режим доступу: http://www.bitext.com/api-demo.html 2. Deep sentiment analysis with attensity analyze optimises Lloyds' customer service. – Режим доступу: http://www.attensity.com/wp-content/uploads/2010/09/LloydsSuccessStory.pdf 3. Kasper W. Sentiment Analysis for Hotel Reviews / Walter Kasper, Mihaela Vela. – Proceedings of the Computational Linguistics-Applications Conference. – Jachranka, Poland: Polskie Towarzystwo Informatyczne, Katowice, 10/2011. – pp. 45–52. 4. Semantria. – Режим доступу: https://semantria.com/demo 5. Skyttle API. – Режим доступу: http://nlp.skyttle.com/api/nlp/ 6. UGTag – a morphological tagger for Ukrainian language. – Режим доступу: http://www.domeczek.pl/~polukr/parcor/ 7. Yessenov. Sentiment Analysis of Movie Review Comments / Yessenov, Kuat, Sasa Misailovic. – Massachusetts Institute of Technology, Spring 2009. – Режим доступу: http://people.csail.mit.edu/kuat/courses/6.863/report.pdf.*