Feature based sentiment analysis

Conference Paper · November 2014

CITATION
1

READS
1,493

2 authors:

Katarzyna Wójcik
Cracow University of Economics
19 PUBLICATIONS **22** CITATIONS

SEE PROFILE

Janusz Tuchowski
Cracow University of Economics
28 PUBLICATIONS **13** CITATIONS

SEE PROFILE

FEATURE BASED SENTIMENT ANALYSIS

MSc Katarzyna Wójcik[1], MSc Janusz Tuchowski[2]
Computational Systems Department
Cracow University of Economics

[1] wojcikk@uek.krakow.pl
[2] tuchowsj@uek.krakow.pl

Abstract

Sentiment analysis or opinion mining is field of research that can have significant impact on today's business. Increasing number of consumers' reviews created the need of its automatic analysis. This issue is gaining popularity for both – researchers and entrepreneurs, for whom consumers' reviews are important source of business information.

There are three main areas of opinion mining:

- *opinions classification*
- *feature based opinion mining*
- *comparative sentences analysis*

This paper is focused on feature based sentiment analysis in which not the sentiment of the whole opinion is analyzed but how particular features of opinion's subject are seen. For the purpose of research ontology-based approach and pattern based approach to opinion mining are used. First of them allows to present products features in a form of hierarchy while second automatizes the process of features with its sentiments extraction. To every feature the sentiment is assigned.

Keywords: *sentiment analysis, opinion mining, text mining, opinions, reviews*

1. INTRODUCTION

Sentiment analysis or opinion mining is field of research that can have significant impact on today's business. Increasing number of consumers' reviews created the need of its automatic analysis. This issue is gaining popularity for both – researchers and entrepreneurs, for whom consumers' reviews are important source of business information.

There are three main areas of opinion mining:

- opinions classification
- feature based opinion mining
- comparative sentences analysis

This paper is focused on feature based sentiment analysis in which not the sentiment of the whole opinion is analyzed but how particular features of opinion's subject are seen. For the purpose of research ontology based approach and pattern based approach to opinion mining are used. First of them allows to present products features in a form of hierarchy while second automatizes the process of features with its sentiments extraction. To every feature the sentiment is assigned.

In the paper firstly sentiment analysis will be defined. Different areas of opinion mining and approaches do it will be mentioned. Then we will move on to description of feature based sentiment analysis. This part of paper will also contain explanation how different approaches are used in this area. Next part of paper will contain description and results of empirical studies. We will sum up with conclusions and further research plans.

2. SENTIMENT ANALYSIS

Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials.

Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication (Wikipedia EN - Sentiment_analysis, 2014).

In the field of sentiment analysis there are three main areas (Liu, 2007):

- Sentiment classification: assignment of sentiment to whole opinion or division of opinions into groups on the basis of its polarity; usually two or three groups are used – positive, negative and sometimes neutral opinions
- Featured-based opinion mining and summarization: discovering what aspects of product users like or dislike.
- Comparative sentence and relation mining: analysis of sentences comparing directly one object to another

2.1. OPINIONS

Reviews are a specific type of textual data. They have a subjective character - they express the attitude of opinions' authors to the objects of the opinions. In some services verbal opinions are supported by points or stars representing the value of opinion.

Opinions that can be found in the Internet can be divided into three groups according to their form (Liu, 2007):

Form1: advantages (pros), disadvantages (cons) and summary

Form2: advantages (pros) and disadvantages (cons)

Form3: no rules or restrictions, free form
The form of opinion determines in some approaches methods of its analysis.

2.2. TEXT MINING APPROACHES TO SENTIMENT ANALYSIS

There are few text mining approaches to sentiment analysis (Lula and Wójcik, 2011):

- Word-based approach – it is assumed that the meaning of the opinion (also its sentiment) is carried by separate words; so that the sentiment is assigned to every word in opinion
- Pattern-based approach – it is assumed that the sentiments are carried by phrases/expressions instead of separate words so the sentiment is assigned to identified phrases
- Ontology-based approach – in this approach ontology is used to present domain knowledge about the subject of opinion; it allows showing the structure of product or service which is rated in opinion
- Statistical learning approach – in this approach training set (containing opinions with sentiments given) is required; on this basis the model learns how to assign polarity to new opinions

There are some significant differences between those approaches. They can be used in different types of sentiment analysis. For each area the best approach can be identified (Wójcik and Tuchowski, 2012). Figure 1 below presents which types of sentiment analysis involves particular text mining approaches (Wójcik and Tuchowski, 2013).

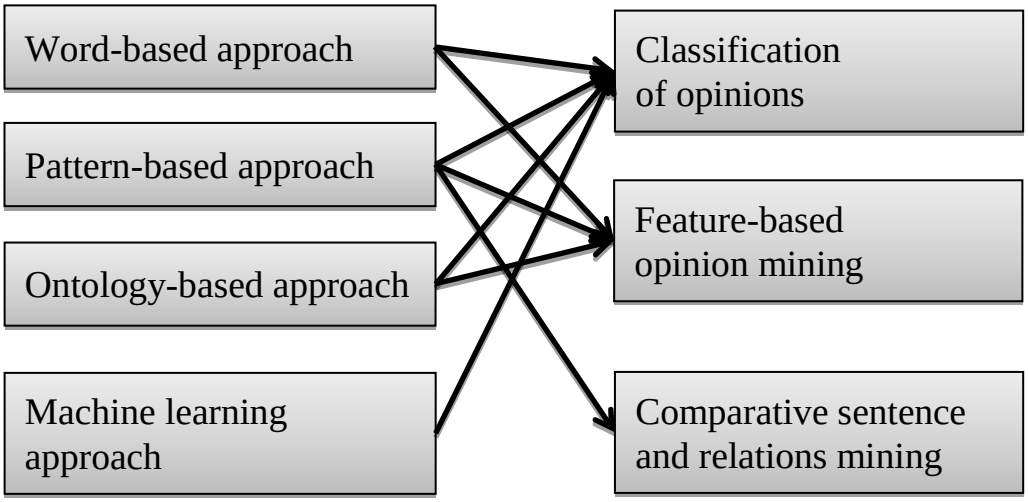

Figure 1. Different approaches used in particular opinion mining analysis

Source: own elaboration

Often it may be useful to combine different text-mining approaches to sentiment analysis to improve the results.

3. FEATURE BASED OPINION MINING

Featured-based opinion mining covers discovering what aspects of product users like or dislike. Not the whole opinion is examined but its parts that refer to listed features of product or service. As it can be seen on Figure 1 in feature based opinion mining three different text mining approaches to sentiment analysis can be used: word-based, pattern-based and ontology-based approach. In previous papers authors described them as addition to main topic of research. This paper is focused on featured-based opinion mining gathering all previously

mentioned materials and adding comparison of different approaches. It also contains attempt to identify possible mixtures of different approaches helping to improve the results reducing efforts needed to be taken.

3.1. WORD-BASED APPROACH

A tag cloud (or word cloud) is a visual representation for text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text (Wikipedia EN - Tag_cloud, 2014). Tags are usually single words, and the importance of each tag is shown with font size or color.

Cloud of tags can be used in feature-based sentiment analysis. We assume that feature which appears in cons cloud is its disadvantage while feature on pros cloud is product's advantage. Before cloud of tags can be prepared the removal of words from stoplist must be done. Those words can darken the results achieved while using this method (Wójcik and Tuchowski, 2013).

This method is easy to use. However it requires opinions with pros and cons distinguished in every examined opinion (opinions in 1st or 2nd form). On the tag cloud we can see how many users mention particular feature as advantage or disadvantage of product or service. We cannot observe what the strength of positive or negative emotion of particular user or group of users is.

3.2. ONTOLOGY-BASED APPROACH

Ontology is a formal and shared specification of a domain of interest. It formally represents knowledge as a set of concepts within a domain together with the relationships between those concepts. Classes (concepts) in ontology can have hierarchical structure.

Ontology also contains a set of objects (individuals, instances of concepts) which represent real items or beings from a given domain. The concepts may have attributes expressing their properties. Ontology can be used to describe the domain and to reason about the entities within that domain.

Single ontology concerns only one domain. To construct ontology the knowledge about particular field of interest is needed. Ontology designed for one domain cannot be applied to another one. Created ontology can be a starting point for many different analyses of texts representing particular domain including sentiment analysis.

The ontology as graph-like construction makes feature based opinion mining easier to conduct. Ontology can be used in sentiment analysis in different ways. In most popular approach single opinion can be presented as an instance of ontology (Kontopoulos *et al.* 2013). The main characteristics of the subject of opinion can be presented as its attributes in the ontology. Then the polarity of each feature must be determined either for single opinion or for the whole set of opinions. In a special case of this approach for every node two additional leafs representing positive and negative sentiment are added (Wei and Gulla, 2010).

The difficulty in this approach lies in process of assigning sentiment to each feature. It can be done manually by tagging certain parts of opinions and assigning them to proper features as attributes. To automate this process some rules must have been created. Second thing that have to be done is preparation of special dictionaries containing positive and negative words or expressions with sentiment value assigned.

3.3. PATTERN-BASED APPROACH

Pattern-based approach allows identifying certain phrases in opinions. To those phrases sentiments can be assigned. Advantage of this approach is possibility to detect phrases that

modify sentiment like negation, nullification, strengthening and others. In feature-based sentiment analysis patterns help to identify product's features and combine them with polarized words attached to them.

In this approach the key thing is to define rules. One must decide if the rules will be connected with particular product or service or if they will be more universal. In most solutions mechanisms based on regular expressions are used. Also in this approach sentiment dictionaries are required. The problem that must be taken into consideration is that some words have positive meaning in one context and negative in another.

3.4. MULTI-MODEL APPROACH TO FEATURE-BASED OPINION MINING

Comparing problems and solutions in different approaches it can be seen that they are complementing each other. Word-based approach can be used in features identification. All nouns used in opinions can be extracted. Then from most popular of them the selection of those describing product or service features can be done. Also dictionaries construction and complementation can be done using word based approach.

Ontology helps in creation of hierarchy of features. However it can be filled with polarized words using patterns. Hierarchy allows to group descendant features in one major. Then we can concentrate only on features from certain, higher level which are usually more important. In the same time it is possible how often and how very detailed features are rated in opinions. Patterns not only help in features identification but are also useful in interpretation of sentiment phrases assigned to particular features.

4. EMPIRICAL ANALYSIS

During research the simulation analysis was conducted. Its aim was to describe and compare exemplary usages of different approaches to feature based sentiment analysis. To reach this aim not only the accuracy of results was compared but also the amount of work involved in each analysis.

In the research process the following stages of analysis can be distinguished:

1. Extraction of opinions from the Internet
2. Word-based experiment
3. Ontology-based experiment
4. Pattern-based experiment
5. Results comparison

4.1. OPINIONS EXAMINED IN RESEARCH

The first step of research concerns extraction of opinions about Smartphones. In this case reviews about subsequent models from the same family were chosen. All 737 opinions were extracted from cokupić.pl via ceneo.pl. The analyzed opinions were all written in Polish language. Figure 2 presents exemplary opinion. It can be seen that the opinion is in the first form: it contains pros (1), cons (2) and summary (3). However in this case rather pros and cons summarize the opinion included in the plain text. Each review is also supported by overall mark expressed by stars (4).

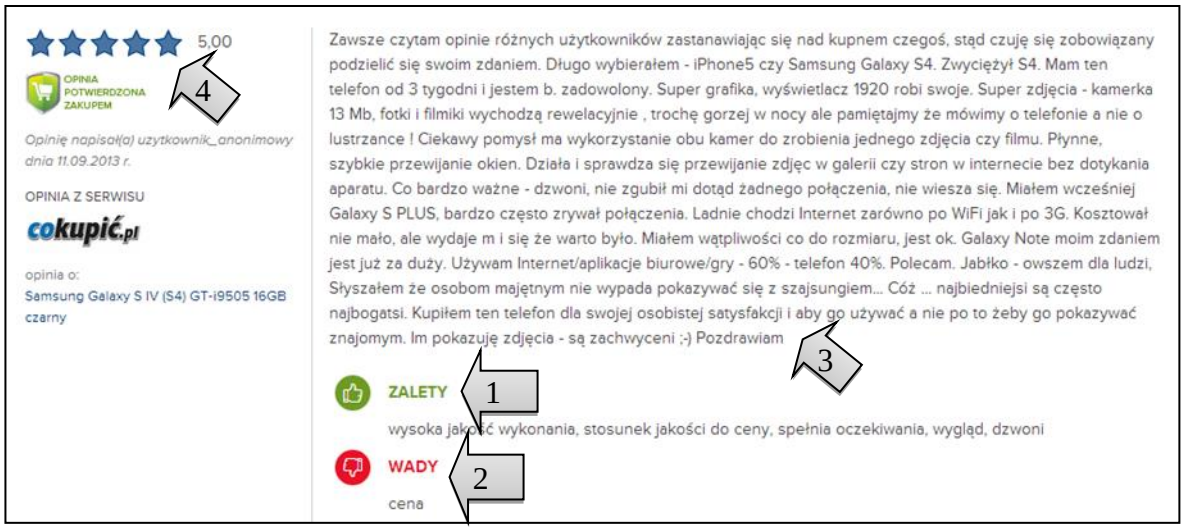

Figure 2. Exemplary opinion extracted for the purpose of research.

Source: http://www.ceneo.pl

Opinions from the source were extracted using DOM parser for PHP language. At first all necessary data were saved in database. Next each opinion was saved in single file containing pros, cons and summary. The name of each file came from model name, opinion index and number of stars.

4.2. WORD-BASED APPROACH EXPERIMENT

In word-based approach experiment opinions were divided by the phone model. For each model two files were created: one containing all pros from opinions about it and second containing all cons. Then stemming process was conducted. On this stage only nouns were left. From all nouns those describing product's features were taken into consideration. On certain level some words were replaced by their synonyms eg. screen and display (pl. *ekran* i *wyświetlacz*). Then for each file cloud of tags was prepared.

Figure 3. Clouds of tags for pros and cons for Samsung Galaxy S II

Source: own elaboration in http://www.tagxedo.com/

Figure 4. Clouds of tags for pros and cons for Samsung Galaxy S III

Source: own elaboration in http://www.tagxedo.com/

Figure 5. Clouds of tags for pros and cons for Samsung Galaxy S4

Source: own elaboration in http://www.tagxedo.com/

Figure 6. Clouds of tags for pros and cons for Samsung Galaxy S5

Source: own elaboration in http://www.tagxedo.com/

The bigger the word on tag cloud more often it appears in opinions. From figures 3-6 it can be inferred that in all models of Samsung Galaxy S family price (pl. *cena*) and battery (pl. *bateria*) was the disadvantages most often mentioned by users. In the same time for most users its advantage was how does it look like (pl. *wygląd*).

For different models some features appear on both clouds. It means that for some users they were advantages while for other disadvantages. For example:

- battery (pl. *bateria*) in Samsung Galaxy S II
- quality (pl. *jakość*) and screen (pl. *ekran*) in Samsung Galaxy S III
- look (pl. *wygląd*) and quality (pl. *jakość*) in Samsung Galaxy S 4
- battery (pl. *bateria*) in Samsung Galaxy S 5

4.3. ONTOLOGY-BASED APPROACH EXPERIMENT

For ontology-based approach experiment only 15 opinions about Samsung Galaxy S4 were chosen. This was caused by the time-consuming process of filling in the ontology. Without patterns it takes a lot of time and work. The aim was just to show the idea behind this approach.

The whole analysis begins with the design of ontology. As it was mentioned before separate ontology must be created for every domain. There are no strict rules of ontology design. The researcher must decide about its structure. The common problem appears when the decision if particular feature will be concept (class) or just the attribute of another class must be taken.

For the purpose of this research the ontology of Smartphone was created. It was based on the one described in (Haider, 2012) using also results of research published in (Jainarain, 2012). However features pointed in word-based approach could also be the basis for ontology creation.

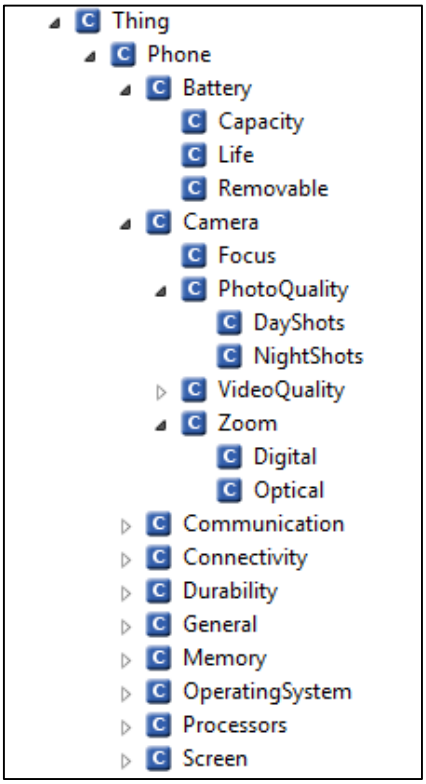

Figure 7. Ontology tree of Smartphone ontology.

Source: own elaboration

The root class is Thing and its direct descendant is Phone class. This class is the ancestor for classes representing major features of Smartphone. The next generations represents less important or more detailed characteristics of phones. In this case authors decided to represent all important features of Smartphone as separate classes in ontology.

Prepared ontology can be also called taxonomy as it contains hierarchy and classification of concepts representing features of Smartphone. Figure 7 presents a part of the ontology hierarchy. Battery and Camera subclasses are partially expanded. Figure 8 presents the same ontology as graph. Phone can be identified as the central node. Also Thing node as the root one can be pointed on that figure. Nodes are connected to each other with directed edges. The arrows in the middle of each edge represent the direction of relation. The relations "has a" or "is a" are used.

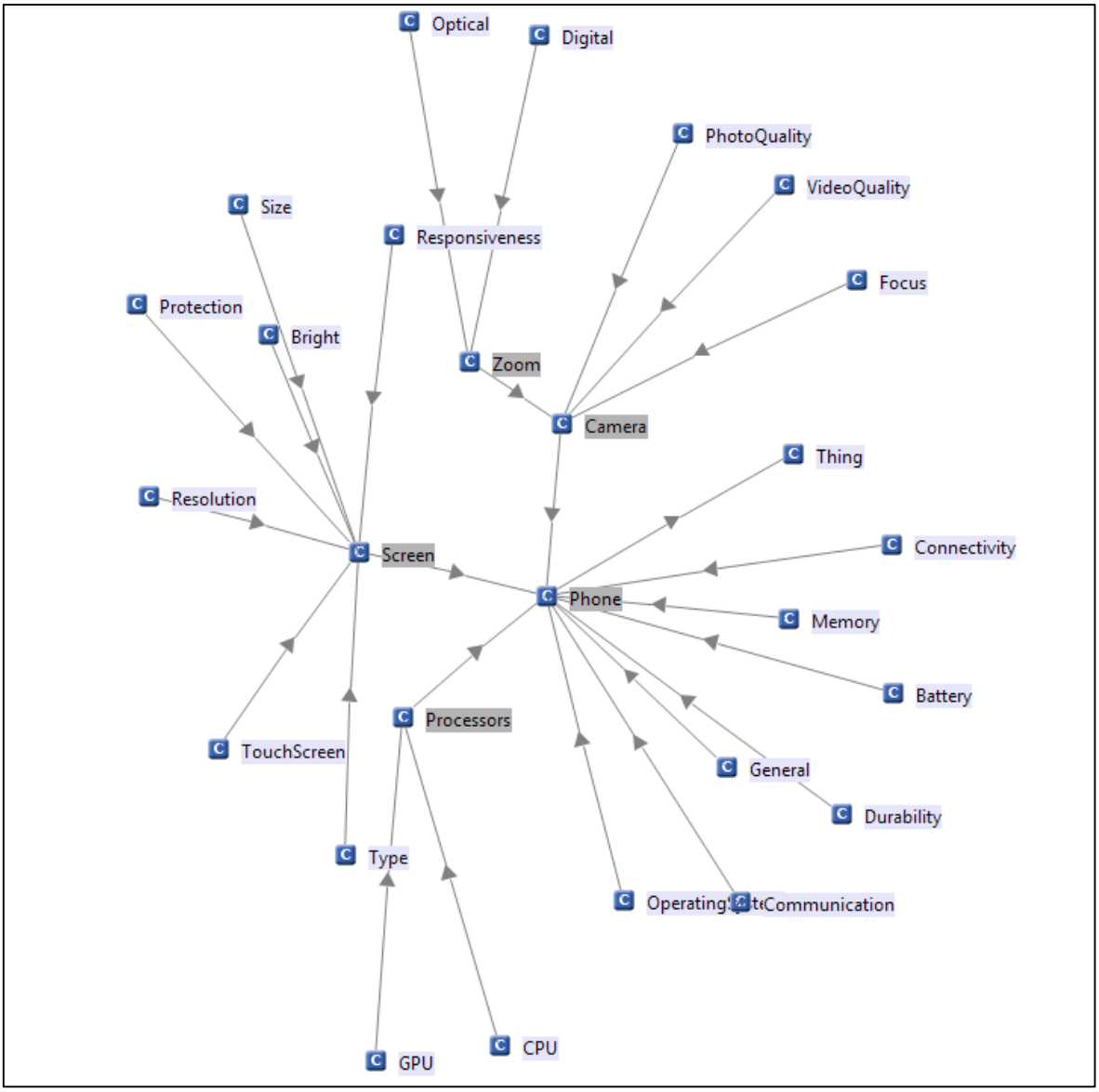

Figure 8. Smartphone ontology presented as graph.

Source: own elaboration

The extracted opinions must be in next step tagged with concepts from ontology. Only phrases that have polarity were chosen from each opinion. Then those phrases were combined with instances of classes representing the feature which sentiment was enclosed in particular

phrase. The instances were created in the tagging process and phrases became the labels of those instances. Figure 9 presents tagging process in GATE application.

To conduct tagging process in GATE all opinions must be loaded and combined into corpus. Then ontology is also loaded. In next phase phrases with polarity are marked (1) and class from ontology tree is chosen. This process includes creation of new instance of particular class (2) and assignment of phrase as the label of instance.

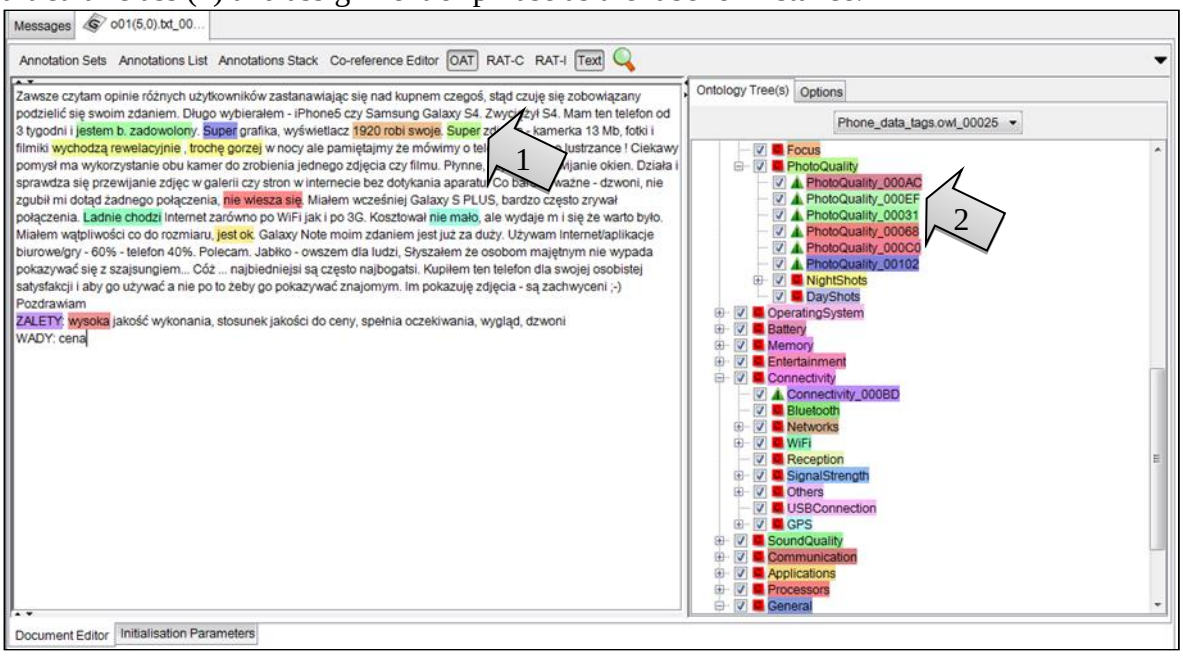

Figure 9. Process of tagging opinion with instances from ontology

Source: own elaboration in GATE application

Next step of analysis requires special dictionary containing words that can carry sentiment and numerical values representing polarity of each word (or phrase) including its direction and strength. For English language such dictionaries can be found in the Internet. SentiWordNet is one of them (Ohana and Tierney, 2009; Esuli and Sebastiani, 2006). For Polish language there is no sentiment dictionary available.

```
<rdf:Description rdf:about="http://uek.krakow.pl/phone#Heat_0008E">
      <rdf:type rdf:resource="http://uek.krakow.pl/phone#Heat"/>
      <rdfs:label rdf:datatype="&xsd;string" xml:lang="pl">nagrzewa się</rdfs:label>
      <sentiment rdf:datatype="&xsd;int">-4</sentiment>
</rdf:Description>

Heat_0008E – Individual
Heat – Class
nagrzewa się – label (string from opinion)
-4 – Datatype Property (sentiment)
```

Figure 10. Part of .owl file describing single phrase in ontology.

Source: own elaboration

For the purpose of current experiment all tagged phrases were collected and to each of them sentiment as value from range [-10;10] was assigned. Then those values were put into ontology as instance attribute called sentiment. Figure 10 presents definition of one instance

in ontology after tagging and sentiment assignment. Class of particular instance is pointed, its name (serialized by the program) and values of label and sentiment attributes.

To conduct feature-based sentiment analysis using ontology-based approach the sentiment values assigned to all instances of particular feature class must be aggregated. Two most frequently tagged of them were chosen and their sentiment was calculated. Table 1 presents sentiment values determined for phrases tagged in different opinions and total sentiment for **Battery** and **BuildQuality** features. Three different aggregation methods were chosen: average mean, median and dominant.

Table 1. Feature based analysis of examined opinions

	Battery (9 instances)	**BuildQuality (8 instances)**
	9	8
	-5	8
	6	8
	9	-5
	5	8
	-3	8
	5	-3
	-4	8
	-5	
Sentiment [-10,10] – average	1,(8)	5
Sentiment [-10,10] – median	5	8
Sentiment [-10,10] – dominant	9	8

Source: own elaboration

The results in Table 1 reflects results presented on Figure **5**. However the experiment was conducted on much smaller sample. It seems that median gives quite good results but to generalize this conclusion bigger sample must be examined.

4.4. PATTERN-BASED APPROACH EXPERIMENT

Pattern-based approach implies using some kind of rules or patterns based on regular expressions. The first thing that had to be done during research in this area was to find the application enabling creation and interpretation of that kind of rules of patterns. In literature authors found application created by Institute of Computer Science of Polish Academy of Sciences.

Spejd (ang. ***S**hallow **P**arsing **a**nd **D**ezambiguation **E**ngine* – spade[3], ang. ***S**hallow **P**arsing and **E**minently **J**udicious **D**isambiguation* - spejd[4]) is a tool for simultaneous morphosyntactic disambiguation and shallow parsing. The Spejd formalism is essentially a cascade of regular grammars. Unlike in the case of other shallow parsing formalisms, the rules of the grammar allow for explicit morphosyntactic disambiguation statements, independently or in connection with structure-building statements, which facilitate the task of the shallow parsing of ambiguous and/or erroneous input (Buczyński and Wawer, 2008). For shallow parsing it uses National Corpus of Polish (pl. *Narodowy Korpus Języka Polskiego*)

[3] At first, application was going to be called Spade, but there exist another shallow parsing application under this name. Then authors decided to polonaise its name phonetically. Sometimes they are also using spade symbol (♠) instead of application name.

[4] When the application name was changed also the meaning of abbreviation was changed.

Each rule defined in Spejd has following structure (Przepiórkowski and Buczyński, 2007; Buczyński and Przepiórkowski, 2008):

Rule „Rule name”
Left: left context
Match: regular expression describing matching phrase
Right: right context
Eval: operations to be done

Designations used in rules are presented in Table 2.

Table 2. Designations used in rules

National Corpus of Polish	
lexeme	**Denotation**
noun	subst
adjective	adj
adverbial	adv
verb not in past form	fin
particle-adverbial	qub
Operators	
operator	**Meaning**
[…]	single token
attribute~value	exists attribute with given value
Operations	
operation	**Action**
unify	leaves tokens with the same values of certain attributes.
alter	modifies token of its part
group	group tokens

Source: own elaboration on the basis from Spejd manual

Firstly Spejd was used in research to tag parts of speech and to left only nouns. In this step also stemming was conducted. On this basis the most frequently appearing features were chosen for experiment. Figure 11 presents rule used for this purpose.

```
Rule "SUBST"
Match:  [pos~subst];
Eval:   group(SUBST,1,1, 0.base);
```

Figure 11. Spejd rule filtering nouns

Source: own elaboration

In next stage of research two types of rules were defined:

1. rules modifying sentiment of single words
2. rules combining identified features with their polarized neighbors

Buczyński and Wawer in their papers define four types of sentiment modifying structures (Buczyński and Wawer, 2008):

- Negation – reversing the polarity
- Nullification – expressing lack of certain quality or property
- Limitation – expressing that positive or negative sentiment has only a very limited expand
- Negative modification – combination of negative adjective with positive noun

Figure 12 blow presents example of negation. Words sentiment is reversed (multiplied by -1)

```
Rule    "Negative -"
Match:  [orth~nie/i] [sen>0];
Eval:   alter(2, sen=sen*-1,);
```

Figure 12. Spejd rule for negation

Source: own elaboration

Another rule modifying sentiment is its strengthening (pl. *wzmocnienie*). There are some words expressing that something was very good or bad. Figure 13 presents example of that kind of rule. It required additional variable ($wzmocnienie) defining words that can strengthen the sentiment of other words.

```
Define wzmocnienie = [
base~bardzo|szczerze|gorąco|naprawdę|zdecydowanie];

Rule    "Wzmocnienie -"
Match:  $wzmocnienie [sen<0];
Eval:   alter(2, sen=sen*1.2,);
```

Figure 13. Spejd rule for strengthening with variable definition

Source: own elaboration

In the research few rules combining identified features with their polarized neighbors were defined. Figure 14 below presents exemplary one. It groups feature with verb and adverbial following it.

```
Rule "SUBST FIN QUB? ADV? ADV ADJ?"
Between: [orth~"się"];
Match:  [base~"bateria|cena|wygląd|jakość|wyświetlacz|e
kran|aparat|funkcjonalność|ekran|wykonanie|funkcja|zdję
cie|aplikacja|obudowa|obsługa|dźwięk|rozdzielczość|proc
esor|GPS|design"][pos~fin][pos~qub]?[pos~adv]?[pos~adv]
[pos~adj]?;
Eval:   group(SFQAA,1,5,0.orth);
```

Figure 14. Spejd rule grouping feature with verb and polarized adjective

Source: own elaboration

Also on this stage two dictionaries were used. First of them contained positive and second negative words. All words were in their base form. To each word there was sentiment assign. Values of sentiment were integer numbers in the range [-10;-1] for negative and [1;10] for positive dictionary. Both dictionaries were created for the purpose of this research and expanded during experiment.

Authors of Spejd described in their paper dictionaries they were using (Buczyński and Wawer, 2008) but in the time of this research those dictionaries were unavailable. Their construction was different than construction of dictionaries used in this research. When they

found a new word which was not in dictionary but its base form was there – they were adding new word to the dictionary. In this research only base forms of words were put into dictionaries so before sentiments assignment stemming was necessary.

Some words couldn't be put into any of dictionaries because they have different polarity depending on the context. They must be analyzed separately. Those are words like: high/low, fast/slow, long/short etc. Figure 15 presents exemplary rules proceeded by definition if variables $wysoka (high) nad $niska (low). In this rule they are combined with feature quality (pl. *jakość*).

```
Define niska =
    [base~"niski|nieduży|niewielki|niewysoki"];

Define wysoka =
    [base~"wielki|duży|wysoki"];
Rule "jakość +"
Match: $wysoka [base~jakość];
Eval: word(2, sen=5, 0.orth);

Rule "jakość -"
Match: $niska [base~jakość];
Eval: word(2, sen=-5, 0.orth);
```

Figure 15. Spejd rule dealing with polarized adjective which sentiment depends on the context.

Source: own elaboration

Figure 16 below presents exemplary opinion. Following pictures presents parts of xml file containing results of rules operations on exemplary opinion.

```
Telefon naprawdę super aczkolwiek ma minusy.z plusów
to wyświetlacz bardzo czytelny i z dużymi kątami
widzenia.aparat przy dobrym oświetleniu robi zdjęcia
lepsze od mojego 8mpx kompaktu, gorzej wychodzą przy
słabym oświetleniu. Masa funkcji dobrze działających
np. Barometr, higrometr, sterowanie ruchem dłoni bez
dotknięcia telefonu i wiele innych do których jeszcze
nie doszedłem. Z minusów to trochę palcujący się
ekran. Brak funkcji głosowych w języku polskim.
Słabszy zasięg sieci (se xperia arc s na tej samej
karcie sim miała lepszy zasięg) mikro sim to kolejny
minus jak dla mnie. Smart fon ma więcej zalet niż wad
i śmiało mogę go polecić.Jest trochę wielki i chwile
trwało zanim przyzwyczaiłem się do jego 5 cali,, teraz
jak biorę do ręki jakiegoś 4 calowego to jakiś taki
maluszek z niego:-)
wysoka jakość wykonania, spełnia oczekiwania, wygląd
cena, słabszy zasięg,
```

Figure 16. Exemplary opinion before parsing

Source: own elaboration

On Figure 17 there are results of two rules actions. First of them is strengthening. Originally word super has polarity equal to 8 in dictionary. Multiplying it by 1,2 gives value

9,6. Second rule groups noun telefon (ang. *phone*) with particle-adverbial and polarized adjective describing it. On Figure 18 rule "jakość+" was used. It identified words wysoka jakość (ang. high quality), combined them into group and assigned sentiment equal to 5. Figure 19 presents group of words słaby (ang. *weak*) and zasięg (ang. *coverage*). Polarized adjective describes attitude of opinions author to noun representing feature.


```
<tok id="a9a52" string-range="string-range(p-1,0,7)">
<orth>Telefon</orth>
<lex><base>telefon</base><ctag>subst:sg:nom:m3</ctag></lex>
<lex><base>telefon</base><ctag>subst:sg:acc:m3</ctag></lex>
</tok>
<tok id="a9a53" string-range="string-range(p-1,8,8)">
<orth>naprawdę</orth>
<lex><base>naprawdę</base><ctag>qub</ctag></lex>
</tok>
<tok id="a9a54" string-range="string-range(p-1,17,5)">
<orth>super</orth>
<lex><base>super</base><ctag>adj:sg:nom:m1:pos:sen=9.6</ctag></lex>
<lex><base>super</base><ctag>adj:sg:nom:m2:pos:sen=9.6</ctag></lex>
<lex><base>super</base><ctag>adj:sg:nom:m3:pos:sen=9.6</ctag></lex>
<lex><base>super</base><ctag>adj:sg:nom:f:pos:sen=9.6</ctag></lex>
<lex><base>super</base><ctag>adj:sg:nom:n1:pos:sen=9.6</ctag></lex>
<lex><base>super</base><ctag>adj:sg:nom:n2:pos:sen=9.6</ctag></lex>
</tok>
```

Figure 17. Part of xml file containing parsed exemplary opinion

Source: own elaboration

```
<syntok id="a9bc4" rule="jakość +">
<orth>wysoka jakość</orth>
<lex><base>wysoka jakość</base><ctag>subst:sg:nom:f:sen=5</ctag></lex>
<lex><base>wysoka jakość</base><ctag>subst:sg:acc:f:sen=5</ctag></lex>
<tok id="a9ba8" string-range="string-range(p-1,803,6)">
<orth>wysoka</orth>
<lex><base>wysoki</base><ctag>adj:sg:nom:f:pos</ctag></lex>
<lex><base>wysoki</base><ctag>adj:sg:voc:f:pos</ctag></lex>
</tok>
<tok id="a9bad" string-range="string-range(p-1,810,6)">
<orth>jakość</orth>
<lex><base>jakość</base><ctag>subst:sg:nom:f</ctag></lex>
<lex><base>jakość</base><ctag>subst:sg:acc:f</ctag></lex>
</tok>
</syntok>
```

Figure 18. Part of xml file containing parsed exemplary opinion

Source: own elaboration

```
<tok id="a9bc0" string-range="string-range(p-1,862,7)">
<orth>słabszy</orth>
<lex><base>słaby</base><ctag>adj:sg:nom:m1:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:nom:m2:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:nom:m3:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:acc:m3:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:voc:m1:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:voc:m2:com:sen=-4</ctag></lex>
<lex><base>słaby</base><ctag>adj:sg:voc:m3:com:sen=-4</ctag></lex>
</tok>
<tok id="a9bc1" string-range="string-range(p-1,870,6)">
<orth>zasięg</orth>
<lex><base>zasięg</base><ctag>subst:sg:nom:m3</ctag></lex>
<lex><base>zasięg</base><ctag>subst:sg:acc:m3</ctag></lex>
```

Figure 19. Part of xml file containing parsed exemplary opinion

Source: own elaboration

The next step in this experiment was to count total sentiment of each indicated feature. For this purpose authorial application in Java was used. For each model the averages of sentiments assigned to particular features were counted. Results are in Table 3. Empty cells signify lack of polarized phrases connected to particular feature. Column positive/negative shows if attitude of customers to certain feature was positive or negative. Column trend shows how this sentiment changes through consecutive models.

Table 3. Aggregated sentiments of chosen features of exemplary smartphones.

Samsung Galaxy / Feature	S II	S III	S4	S5	positive/negative	trend
aparat (camera)	5,00	5,16	5,41	8,25		
bateria (battery)	-1,10	-1,68	-0,07	4,50		
cena (price)		5,45	9,60			
dźwięk (sound)	5,00	6,00	10,00			
ekran (display)	3,25	5,81	6,66	6,25		
funkcjonalność (functionality)	5,00	5,83	6,25	6,50		
jakość (quality)	5,00	4,78	0,38	-1,50		
obudowa (case)		-5,00	-8,00			
procesor (CPU)	4,80	2,50	5,00			
rozdzielczość (resolution)		5,20	5,00			
wygląd (look)	5,00	6,06	6,01	5,33		

Source: own elaboration

Comparing results achieved in this experiment with previous ones we have confirmation that battery is disadvantage of Samsung Galaxy S Smartphones while look is their advantage. Most features are getting better in newer models. However the quality of latest model is perceived as worse, even negative.

5. CONCLUSIONS AND FURTHER RESEARCH PLANS

Experiments conducted during this research shown that three described approaches give similar results. However ontology-based approach and pattern-based approach needs a lot of preparations. Word-based approach was used as supporting method in other experiments. It helped to improve the results and make the less time-consuming.

All examined opinions were in first form (pros, cons and summary). It is good to analyze separately pros and cons using one method and summary using another method.

Possibility of combining ontology-based approach and pattern-based approach is promising. It could have influence on time consumed by each method and its automation.

In further research authors are going to automate the process of ontology creation using patterns. The set of opinions will also be expanded as well as sentiment dictionaries. Opinions are full of informal languages, abbreviations and emoticons. Taking them into consideration could also give interesting results.

6. BIBLIOGRAPHY

Buczyński, A. and Przepiórkowski, A. (2008). ♠ Demo: An Open Source Tool for Partial Parsing and Morphosyntactic Disambiguation. *Proceedings of LREC 2008.*

Buczyński, A. and Wawer, A. (2008). Automated classification of product review sentiments in Polish. *Intelligent Information Systems*, strony 213–217.

Buczyński, A. and Wawer, A. (2008). Shallow parsing in sentiment analysis of product reviews. *Proceedings of the Partial Parsing workshop at LREC 2008*, (strony 14-18).

Esuli, A. and Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource. *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, (strony 417-422).

Haider, S. Z. (2012). An *Ontology based Sentiment Analysis. A Case Study.* University of Skövde.

Jainarain, R. (2012). *Attributes that influence Generation-Y consumers in their choice of Smartphone.* University of Pretoria.

Kontopoulos, E., Berberidis, C., Dergiades, T. and Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems withApplicatins*(40), strony 4065-4074.

Liu, B. (2007). *Web DataMining. Exploring Hyperlinks, Contents, and Usage Data.* Heidelberg: Springer-Verlag Berlin .

Lula, P., and Wójcik, K. (2011). Sentiment analysis of consumer opinions writen in Polish. *Economics and Management*(16), pp. 1286-1291.

Ohana, B. and Tierney, B. (2009). Sentiment Classification of Reviews Using SentiWordNet. *IT&T Conference.* Dublin: Dublin Institute of Technology.

Przepiórkowski, A. and Buczyński, A. (2007). ♠: Shallow Parsing and Disambiguation Engine. *Proceedings of the 3rd Language & Technology Conference.* Poznań.

Sam, K. M. and Chatwin, C. (2013, Grudzień). Ontology-Based Sentiment Analysis Model of Customer. *International Journal of e-Education, e-Business, e-Management and e-Learning, 3*(6), strony 477-482.

Wei, W. and Gulla, J. A. (2010). Sentiment Learning on Product Reviews via Sentiment Ontology Tree. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (strony 404–413). Uppsala, Sweden: Association for Computational Linguistics.

Wikipedia EN - Sentiment_analysis. (2014, 10 01). Pobrano z lokalizacji Wikipedia EN: http://en.wikipedia.org/wiki/Sentiment_analysis

Wikipedia EN - Tag_cloud. (2014, 10 1). Pobrano z lokalizacji Wikipedia EN: http://en.wikipedia.org/wiki/Tag_cloud

Wójcik, K.and Tuchowski, J. (2012). Comparison analysis of chosen approaches to sentiment analysis. In *IT for practice* (pp. 187-192). Frýdek Místek: VŠB - Technická univerzita Ostrava.

Wójcik, K., and Tuchowski, J. (2013). Sentiment Analysis of Opinions about Hotels Extracted from the Internet. In P. Lula, B. Mikuła, & A. Jaki, *Knowledge - Economy - Society. Global and Regional challenges of the 21st Century Economy* (pp. 755-771). Kraków: Foundation of the Cracow University of Economics.