

ANALYSIS OF PUBLIC STANCE ON TOURISM DESTINATIONS IN SREM/SRIJEM REGION

Olivera Grljević¹ 
Saša Bošnjak² 
Veselin Pavličević³ 
Nataša Pavlović⁴ 

DOI: <https://doi.org/10.31410/tmt.2019.267>

Abstract: *Online reviews posted on social media are a rich source of the customers' voice. In this chapter, online reviews about various tourism destinations and attractions in Srem/Srijem region are used to analyze public stance and to uncover sources of satisfaction and dissatisfaction of visitors. Online reviews as unstructured data cannot be directly used for analysis, but require extensive data transformation and preparation which is done through annotation of data. Annotation is a process of enrichment of texts with specific meta-data. In presented research during the annotation process, the dataset, i.e. the corpus of collected reviews was labeled with information on sentiment, aspects, discourse functions, and information on a function different words have in expressions of opinion, sentiment or attitude within a review. The authors present their approach to annotation, the evaluation of the quality of conducted work, as well as the analysis of the resulting dataset, i.e. annotated corpus.*

Keywords: *Online reviews, User-generated content, Annotation, Corpus development, Corpus analysis, Srem/Srijem region.*

1. INTRODUCTION

Rapid use of the Internet and social media has changed the way people share and search for information about tourism and hospitality products, services, or entities. Traditional word of mouth is replaced with e-word of mouth (eWoM) (Gruen et al., 2006), meaning that people prior to conducting purchases or making a visit consult online sources of information. The most valuable source of eWoM is user-generated content in the form of online reviews. People freely share information on personal customer experience or recommend products and services to others, and as such social media has become a rich source of the customer's voice. Many studies have proved the relevance of social media from customers' point of view. Customers appreciate and believe more in the opinion of other people than in advertisements of companies (Berthon et al., 2012; Pitt et al., 2002). Customers have equal trust in online comments and reviews as they have in personal recommendations of friends (Park et al., 2007; Gligorijevic & Luck, 2012). Online reviews are in third place according to the influence on purchasing decisions (Yang et al., 2015). Given the influence, online reviews have on the shaping of public opinion and the richness of information hidden in them, modern tourism and hospitality entities should not circumvent their analysis.

¹ University of Novi Sad, Faculty of Economics in Subotica, Segedinski put 9-11, Subotica, Serbia

² University of Novi Sad, Faculty of Economics in Subotica, Segedinski put 9-11, Subotica, Serbia

³ University of Novi Sad, Faculty of Economics in Subotica, Segedinski put 9-11, Subotica, Serbia

⁴ Tourism Organisation of Vojvodina, Bulevar Mihajla Pupina 6/IV, Novi Sad, Serbia

In this paper, we have analyzed public opinion expressed on the Internet about various tourism destinations in Srem and Srijem (from now on: S-S) region to support the enhancement of continental tourism. To do this, we have collected user-generated online reviews from the Internet directed to four different types of tourism destinations: Nature, Nature reserves, Museums, and Monasteries. Online reviews are unstructured data as they can vary in length, types of characters, language, or other characteristics. To introduce certain structures into such data collection, we have conducted annotation of collected data following the MATTER methodology (Pustejovsky & Stubbs, 2012; Stubbs, 2012). A collection of textual data is referred to as a corpus. In our research the corpus comprises online reviews. Annotation of a corpus is a process of enriching textual data with linguistic, semantic, or other meta-data useful for further analytical goals. Annotation is conducted manually by adding labels over certain parts of texts. Four annotators conducted manual annotation of our corpus. Annotators were instructed to mark text with the following attributes: *sentiment polarity*, *aspect*, *discourse function*, various *functions of expressions*. *Sentiment polarity* indicates sentiment conveyed in online review. It can be positive, negative, neutral, or it can convey both positive and negative sentiment. *Aspect* indicates the target of the sentiment. *Discourse function* is content dependent and it indicates usage of particular expressions in various contexts. The annotators were instructed to consider and mark three *function* types on a word level, if a word conveys positive or negative sentiment, and negation. Evaluation of the quality of conducted annotations was done using *Fleiss' kappa* and *agr* metrics (Fleiss, 1971; Fleiss et al., 2003; Wiebe et al., 2005; Wissler et al., 2014). These metrics provide an evaluation of inter-rater agreement. Annotated corpora are usually used in sentiment analysis, opinion mining, or analysis of variations and changes in language, which will be a part of our future work. In this paper, we have used an annotated corpus for statistical analysis of information added over expressions of opinion or personal stance to uncover underlying sources of satisfaction or dissatisfaction of visitors in the S-S region. Enrichment of online reviews achieved through the annotation process enabled thorough analysis of public stance. The conducted analysis provides insights about preferred and disliked aspects of each tourism destination from the perception of visitors, which represent valuable information on potential directions for improvement (through analysis of negative stance) and optimization of marketing activities for a particular destination (through analysis of positive stance).

The paper is structured as follows. In section Related work, we give an overview of relevant literature that corroborates our approach. Section 3 provides an overview of our methodology, data collection, data annotation, evaluation of the quality of annotations, i.e. gold standard corpus, and statistical overview of annotations. In Section 4, we provide analysis of our corpus and public stance on different tourism destinations in the S-S region, as well as discussion of results. In the last section, we provide conclusions and further research directions.

2. RELATED WORK

An online review can be considered as an opinion or expression of the sentiment of one person. In (Liu 2012), opinion is defined as subjective impression provided by some person, h , in a given time, t , which reflects that person's positive, negative or neutral opinion or emotions, s , which is directed towards something, e and a . From this, opinion is defined as quintuple in (1):

$$(e_p, a_{ij}, S_{ijkl}, h_k, t_l) \quad (1)$$

where e_i represents an entity, which can be a product, service, problem, person, organization, or event. Each entity has different parts or subparts described by some attributes, so-called aspects a_{ij} . The entity with its aspects is often called opinion target. One online review can comprise the opinion of multiple persons, so-called opinion holders h_k or sources of sentiment. Each opinion is expressed in a particular time t_l . A sentiment value of expressed opinion of the holder h_k about aspect a_{ij} of the entity e_i in a given time t_l is represented by s_{ijkl} . Sentiment can obtain one value: positive, negative, or neutral. The value of a sentiment is referred to as sentiment polarity.

The usefulness of social media as a data source of public opinion has been proven by various research. Online reviews can be used to discover patterns and trends in data (Liu et al., 2007; Bollen et al., 2011; Van de Kauter et al., 2015), as well as to uncover consumer habits and preferences and to give directions for proactive innovation of products or services (Hu & Liu, 2004; Kim et al., 2014; D'Avanzo & Pilato 2015). Sentiment analysis and opinion mining represent a computational study of people's opinions, sentiment, emotion, or attitudes (Liu, 2012). Classification, as a supervised learning algorithm, is the basic approach to sentiment analysis. Supervised learning algorithms require datasets with clearly distinguished and labeled examples of each class. Online reviews are unstructured data as they can vary in length, writing style, language, or other attributes. Also, they are usually missing labels on sentiment polarity, opinion targets, or other important information for further analysis. Therefore, to use online reviews in sentiment analysis, they need to be extensively pre-processed and prepared. In terms of sentiment analysis, as a leading approach to the analysis of online reviews and public opinion, one possibility for structuring online reviews is an annotation of a corpus. Corpus is a collection of any textual data. Corpus annotation or labeling can be conceptualized as a process of enrichment of a corpus by adding different linguistic meta-data or information of interest, which will serve to the theoretical or practical objective, over the whole document (online review), each sentence, or a particular word (Hovy & Lavid, 2010; Wissler et al., 2014). Depending on the purpose of annotation, information that will be added to the corpus can vary from syntactic, lexical to semantic information. Annotation is a laborious and time-consuming task, but it provides data tailored to a particular domain and analytical task.

Corpus analysis can be carried out to analyze language and its natural occurrence (Garrote et al., 2013; Lin, 2017; Iosif et al., 2018), for discourse analysis (Iruskieta et al., 2013; Rumenapp, 2016; Henricson & Nelson 2017), or to research grammatical and lexical aspects within a corpus (Chung, 2011; González & Ramos, 2013; Hammond, 2015; Tang, 2016; Vass, 2017). Also, corpus annotation creates a fundamental knowledge base for the training of machine learning algorithms that will implement sentiment analysis (Hovy & Lavid, 2010). To study sentiment expressed in online reviews and its effects on behavior of individuals or online reputation, such as in research presented in (Ganu et al., 2009; Dalal & Zaveri, 2014; Yang et al., 2015; D'Avanzo & Pilato, 2015), one must develop domain-specific corpora, i.e. data must be collected for particular domain, and annotation approach must be adjusted to particular task. The need for the development of domain-dependent corpora exists because knowledge and conclusions derived from their analysis cannot directly be transferred to another domain. Domain dependence is partly the result of a change in vocabulary. Identical terms can carry a completely different sentiment in different domains (Pang & Lee, 2008) and models developed for one product/service type often express different performance when applied on a set of reviews of another product/service (Dave et al., 2003; Reinstein & Snyder, 2005). Such domain-specific corpora is developed and widely used for sentiment analysis in various domains: different technical products (Hu & Liu, 2004; Broß, 2013; Kim et al., 2014), automotive domain (Kessler et al., 2010), or movie industry (Pang et al., 2002; Dave et al., 2003). Tourism and

hospitality is a particularly interesting domain for sentiment analysis. In (Broß, 2013; Wachsmuth et al., 2014) corpus is developed for sentiment analysis of hotel services. Customer ratings (star ratings and customer ratings) were used to analyze a relationship they have with online sales (Öğüt & Onur, 2012) or prices (Lei & Law, 2015). The corpus developed, and presented in our paper comprises online reviews on various tourism destinations in the S-S region. It is developed primarily with the goal of the development of an automated sentiment analysis model. For detail analysis of discourse, functions used to express sentiments in online reviews written in Serbian. However, in this paper, we have opted for a different approach to analysis. We have used annotation data, labels which were added over online reviews, to analyze public stance and gain insights on preferred and less preferred aspects of various tourism destinations. Such analysis gives valuable insights on directions for improvements, as well as emphasized positive aspects of tourism destinations, which could be used as part of marketing strategies or offers to tourists.

3. METHODOLOGY

The subject of this research is the evaluation of public stance on various tourism destinations in the S-S region to identify which are preferred aspects of each destination or category of destination and which aspects are not preferred by visitors. Identification of advantages and disadvantages, as well as the analysis of preferences of visitors, is useful for the improvement of marketing and management strategies and for identification of aspects which need improvements. Specific objectives of our research could be formulated as questions. What are the main drawbacks of each analyzed tourism destination from the visitor's perspective? What aspects of tourism destinations in the S-S region do visitors prefer?

To address these objectives, we have chosen to gather online reviews as they represent a publicly available source of people's opinion and to transform them into a useful dataset for such analysis. In this process, we have conducted data collection, data preparation, and annotation of data as suggested in MATTER (Model, Annotate, Test, Train, Evaluate, Revise) methodology described in the papers (Pustejovsky & Stubbs, 2012; Pustejovsky & Moszkowicz, 2012; Stubbs, 2012). Within this methodology, the authors described methodological steps required to obtain and create a corpus, to annotate such corpus with informational attributes, and to evaluate the quality of created corpus. Data collection and data preparation are presented in Section 3.1. The definition of annotation attributes, description, and illustration of annotation attributes are presented in Section 3.2. Evaluation of the quality of conducted annotations is presented in Section 3.3 while statistical analysis of annotated corpus is presented in Section 3.4.

3.1. Data collection and preparation

For the study presented in this paper we have collected publicly available online reviews that target specific destinations in the S-S region. Given the availability of data, online reviews were collected from Google reviews. Our research is focused on three types of destinations, which we refer to as *Destination categories*: Nature reserves, Monasteries and well-known churches in S-S region, and Museums. For each destination category we have collected reviews for particular destinations, which we refer to as *Sub destination*. Within destination category *Nature reserve*, we have collected reviews about four sub destinations: Fruška Gora, Obedska bara, Zasavica, and Borkovac lake. Given that Fruška Gora is mostly discussed we have extracted these reviews as individual destination category. The collection comprises reviews about 14 sub destinations in *Monastery* (and churches) destination category: Besenovo, Divsa, Grgeteg, Jazak, Petkovic, Privina Glava,

Sisatovac, Staro Hopovo, Velika Remeta, Vrdnik, Novo Hopovo, Krusedol, Orthodox church Indjija, and The Church of St. Demetrius. *Museums* destination category comprises reviews about the following sub destinations in S-S region are: Archaeological site Sirmium, Imperial palace Sirmium, Sava Sumanovic memorial house, Sava Sumanovic art gallery, Art museum Sid, Museum of beekeeping, Museum of Srem, City museum Vinkovci, Ilok Castle, Museum Vucedol, Vukovar museum, and Castle Eltz. For each destination, we have collected all available online Google reviews. Collected reviews were pre-processed for further annotation and analysis in the following manner. Spelling and grammar errors were checked and corrected accordingly. Query sentences and historical facts were removed from the dataset given they do not reflect subjective stance; rather, they are used to pose a question or state some information. Data was cleansed from excessive whitespaces and special characters, which are not part of the written language.

In addition to text of a review, we kept several other information: author of the review, relevance of the review which is expressed as the total number of Google reviews a reviewer has previously provided, temporal information, and the number of likes provided by other people which indicates their agreement with what has been said. Reviews are often written in languages other than Serbian, and we have automated translation of such reviews using Python script. Therefore, the resulting structure of collected reviews contains the original text of the review, as well as translation. Such pre-processed data with all the relevant information is structured into XML file using Python script.

3.2. Data annotation

Online reviews collected from social media or the Internet, in general, had to be adequately prepared for further analysis, which is usually conducted by adding meta-data to different parts of texts (the whole online review, each sentence, or a word). Each meta-data label used to indicate some element, function, or information is called an annotation. Conducted annotations must be exact and relevant to the task in order to achieve effective analysis and training of algorithms in terms of sentiment analysis. Therefore, annotation of the corpus represents a critical point in the development of applications for natural language processing (Pustejovsky & Stubbs, 2012). In our research, the annotation was conducted on document-, sentence-, and word-level. Document-level means the labels were added for the overall review. Sentence-level refers to the annotation of each sentence a review comprises. Word-level could refer to a particular word, phrase, or expression. For each level, we have defined different annotation attributes, as explained in the following this section.

On *document-level* annotators were instructed to mark one and only one *Destination category* and *Sub-destination* (see Section 3.1 for more detail) a review refers to. Also, annotators should mark *Sentiment polarity* expressed in overall review, which can obtain one and only one value per review: positive, negative, both (meaning that within review a reviewer is stating mixed sentiment both positive and negative), or neutral.

On *sentence-level* annotators was instructed to label each sentence a review comprises with the following annotation attributes: *Discourse function*, *Aspect*, and *Sentiment polarity* (positive, negative, both, or neutral) *on the sentence level*.

Discourse is a coherent, structured group of the sentence. It can refer to monolog or dialog. Regardless of a product or a service type, there is a fixed set of that universally applicable discourse

functions. The author of (Broß, 2013), has collected 30 discourse functions and analyzed their applicability on two different domains, digital camera, and hotel reviews. During such an analysis, the author has selected 16 applicable discourse functions. In our research, we have adopted most of the discourse function introduced in (Broß, 2013) and made a small modification to adapt it to our needs and domain. Table 1 illustrates the final list of discourse functions used in our research with a brief explanation of their meaning and one illustrative example. Annotation attribute *Discourse function* was mandatory for each sentence. Annotators were instructed to annotate, label, each sentence with one, and only one *Discourse function*. The annotator was limited to choose a function from a set of predefined discourse functions presented in Table 1.

Table 1. List of predefined discourse functions with short descriptions

Function	Description	Example sentence
Advantage	The reviewer mentions the presence of a feature or part of the destination or tourism attraction that he/she considers as an advantage.	There are marked hiking trails for those who are not frequent visitors or those who go without a guide.
Advice	The reviewer advises on circumventing a particular problem, or on the improvement of the service, ambient, surroundings.	Someone could make good money opening a place like that there!
Comparison	The reviewer compares the destination with another similar destination or experience.	Something completely different from all other museums and numerous wineries at Fruška gora.
Conclusion	The reviewer concludes and summarizes his/her comments and potentially expresses a recommendation.	Overall great picnic area!
Expectation or Requirement	The reviewer describes what he expects from the visit, destination, or tourism attraction or describes his requirements to concerning the destination or attraction or any of its aspects.	Epic place, however far exceeded all our expectations!
Fact	The reviewer provides factual information about the destination or tourist attraction, or its aspects (facts can have sentiment polarity, although this is a rare case).	The place has historical significance.
Lack or Problem	The reviewer mentions the absence of a feature or part of the destination or tourism attraction or describes a problem that is encountered with the destination or attraction during the visit or one of its aspects.	There is no parking.
Suggestion	The reviewer offers some suggestions to future visitors, such as who should visit the destination, type of destination, e.g., the destination is suitable for families, power walks, or similar.	If you like history and culture, you must visit one of many monasteries.
Personal Content	The reviewer provides personal information.	I'm a photographer.
Purchase or Usage	The reviewer mentions his personal procedure of purchasing a certain service on a particular destination or tourism attraction, e.g., where and when he bought the ticket (this does not refer to the customer service) or describes a concrete situation when the particular destination or tourism attraction was visited, potentially commenting on his/her experience.	They should be ashamed, entrance to the winery with short storytelling costs 400 RSD, with wine tasting 700.
Sentiment	The reviewer explicitly expresses his/her opinion or feelings. This refers to explicit praise or condemns of a certain aspect or overall destination.	Very beautiful!
Summary	The reviewer summarizes the advantages or disadvantages of the destination or tourism attraction. The summary is usually in the form of enumeration.	Hiking trails, lake for swimming and fishing, public pool, dissent hotel, and a couple of dissent restaurants won't disappoint you.
Other	Any other discourse function that is not listed above.	Blossomed the first snowdrops.

Source: Authors adaptation of discourse functions presented in (Broß, 2013)

Attribute *Aspect* reflects a subject of a comment or a target of the sentiment. Aspects are dependent on the destination type or service provided at a particular tourism destination. Based on the initial insight into collected online reviews, we have defined a set of aspects appropriate for our data, domain, and the goal of future analysis. Table 2 provides an overview of defined

aspects with a short description and illustrative example for each. Given the nature of the data, the main aspects which reviewers write about are *Nature*, *Monasteries (and churches)*, and *Museums (and castles)* in general. We refer to these aspects as hi-level aspects as they are at the highest hierarchy and are applicable if a reviewer writes a general impression, hence we denote them as *NatureGeneral*, *MonasteryGeneral*, and *MuseumsGeneral*. In addition to these general aspects, we provide more fine-grained aspects: *Cleanliness*, *Facility*, *Recreation*, *Location*, *Parking*, *Service*, *Price*, and *Decoration (Landscaping)*. We refer to these aspects as low-level aspects. Annotation attribute *Aspect* was mandatory for each sentence, meaning that annotators were obliged to assign one and only one aspect to each sentence a review comprises.

Table 2. List of predefined aspects with short descriptions

Attribute value	Description	Example
NatureGeneral	A reviewer writes his/her general impression regarding nature, natural resources, picnic area, facilities, or similar. Such a sentence does not contain particular information on other low-level aspects, such as cleanliness or accessibility of the location.	A wonderful time in nature.
MonasteryGeneral	A reviewer writes his/her general impression regarding the monastery or church. Such a sentence does not contain particular information on other low-level aspects, such as service in terms of availability of the priests or in terms of quality of the mass.	A beautiful monastery.
MuseumsGeneral	A reviewer writes his/her general impression regarding museums, castles, memorial houses, or similar. Such a sentence does not contain particular information on other low-level aspects, such as parking, price of the ticket, or similar concrete information.	A place to visit.
Cleanliness	Cleanliness is applicable in terms of all three hi-level aspects. If a reviewer writes about the cleanliness of nature, picnic area, monastery, church, museum, or castle (from the inside or about the cleanliness of its surroundings), this value is applicable.	Unordered, neglected, and unused.
Facility	Facility refers to space or equipment necessary for doing something. For instance, a facility could refer to access for disabled to a particular location, existing of a resting area in terms of natural resources, or similar.	The faucets are broken; there is nowhere to sit.
Recreation	Availability of recreational facilities. Particularly applicable if a reviewer writes about nature or picnic areas.	Swimming pool, walking trails, fishing, restaurant, wooded area, everything is beautiful.
Location	If a reviewer writes about the location of a particular destination, whether it is easily reachable or it is convenient, this value for aspect is applicable.	About ten meters from the river Sava.
Parking	If the reviewer explicitly writes about the parking, this value is applicable.	In front is spacious parking.
Service	Sentences with the mention of any kind of services in the review should be marked with this value. In case of nature, nature reserve, picnic areas services could refer to availability of restaurants in the area and the quality of service for instance, in case of monastery or churches service could apply to mass, souvenir shops or similar, or in case of museums or castle the service could refer to availability of staff or any type of services which particular tourism attraction offers.	Very nice museum with helpful staff and nice exhibitions.
Price	If the reviewer explicitly mentions the price in the sentence, it should be marked with a price.	Expensive tariff.
Decoration/ Landscaping	If the reviewer writes about the decoration of a monastery, for instance, or landscaping, this value is applicable for the attribute <i>Aspect</i> .	Frescoes and iconostases are among the most beautiful in the Balkans.
Other	When no other aspect is applicable.	I was at a summer celebration.

Source: Authors

Sentiment polarity on sentence-level defines the polarity of expressed opinion (positive, negative, or both) or indicates whether the sentence is a fact (neutral). Annotators were instructed to assign one and only one value to annotation attribute *Sentence sentiment polarity*. This attribute was also mandatory, meaning that each sentence had to be assigned appropriate sentiment polarity.

On *word-level*, annotators were instructed to assign to relevant words a function: positive sentiment expression, negative sentiment expression, intensifier, neutralizer, inverter, negation, and negation scope. Table 3 provides an overview of these functions, their brief description, and representative examples. Annotator was instructed to strive to a minimal unit of text for these annotations; however, if there are cases when several consecutive words denote particular function annotators are allowed to assign value to such expressions (two or more words). Also, one word could hold multiple functions. Annotators were instructed to analyze each word to identify relevant words.

Table 3. Word functions with description

Attribute	Usage	Example
PositiveExpression	Words are conveying positive sentiment.	Nice, Beautiful, Wonderful
NegativeExpression	Words are conveying negative sentiment.	Ugly, Bad, Unpleasant
Intensifiers	Words that amplify expressed sentiment.	Very, Extremely
Neutralizers	Words that decrease expressed sentiment.	Somewhat, Little
Inverters	Words that invert the sentiment of the next word. Observing word sequence in the sentence, the next word is a word that is placed after the inverter and belongs to the same sentence.	But, However
Negation keywords	A word is indicating negation.	Not
Scope of negation	The sequence of words that negation keyword influences and changes their meaning.	[not] <i>very nice place</i>

Source: Authors

When the annotator identifies positive or negative sentiment in a sentence, the annotator is instructed to identify the word(s) which bear sentiment. These words are called sentiment expressions, and they convey the overall sentiment (*PositiveExpression* or *NegativeExpression*). *Intensifiers* and *Neutralizers* are words that do not convey clear sentiment when observed individually but are used to emphasize the sentiment expressed with the word which follows. Combined with other words, these words increase (intensifiers) the positive or negative sentiment or decrease (neutralizers) sentiment, both positive and negative. One sentence can contain more than one such word, and annotators are instructed to annotate each word individually. Also, intensifiers and neutralizers are always single words; however, in a sequence of words, more than one consecutive word can be marked as an intensifier or neutralizer. Annotators were instructed to mark each word individually. *Inverters* are words which indicate the shift of sentiment from positive to negative or from negative to positive. Mostly inverters refer to negation; however, they can also be certain types of conjunctions, such as *mada* (though), *premda* (though), *ali* (but), *a* (but). These conjunctions introduce the shift of sentiment. *Negation* inverters the sentiment polarity of those word(s) which follow the negation keyword. If negation is combined with positive sentiment expressions, it will change its sentiment to negative, and vice versa. Annotation of negation is a two-step process. Firstly, annotators should mark negation keywords and then negation span. *Negation keyword* refers to negation words that are separated from remaining of the text, therefor negation of verbs. *Negation span* refers to a set of words whose meaning is changed by the influence of the negation keyword. Negation span usually spans to the punctuation sign, to another negation, conjunction, or a certain entirety in a sentence.

3.3. Evaluation of the annotation quality

Our corpus comprises 1901 online reviews. Annotation was conducted in MAE (Multi-document Annotation Environment) tool⁵ by four annotators. Firstly, annotators were trained for this job in two months. During this phase, annotators were provided with thorough guidelines with explanations, examples, and were instructed to ask about any misunderstandings. Data used in this phase was not used in any consequent phases. The training phase was also useful to identify areas of improvement in annotation guidelines due to ambiguities. When all the issues were resolved, the next phase of annotation was initiated when all annotators were provided with the same set of reviews (200 reviews, i.e. 11% of the corpus) and were instructed to conduct annotations independently without communication. These annotations were used to evaluate the quality of conducted work and the potential to build a trustworthy dataset for training algorithms and data analysis. For this purpose, as suggested in (Pustejovsky & Stubbs, 2012; Wiebe et al., 2005), we calculated inter-rater agreement using Fleiss' kappa (for more detail, please refer to (Fleiss, 1971; Fleiss et al., 2003)) and agr metric expressed as Precision, Recall, and F-measure (for more details on these metrics, please refer to (Baeza-Yates & Ribeiro-Neto, 1999; Kessler et al., 2010; Kessler & Nicolov, 2009; Wiebe et al., 2005)). These metrics are suitable for the evaluation of annotations conducted by more than two annotators. The values of inter-rater agreement point to reliability of annotations, harmonization of annotators, and give pointers for the following work: whether it is necessary to go back on the training phase or the annotation process can carry on for the remaining of the corpus. The approach proposed in (Landis & Koch, 1977) and illustrated in Figure 1 was used for interpretation of the resulting values.

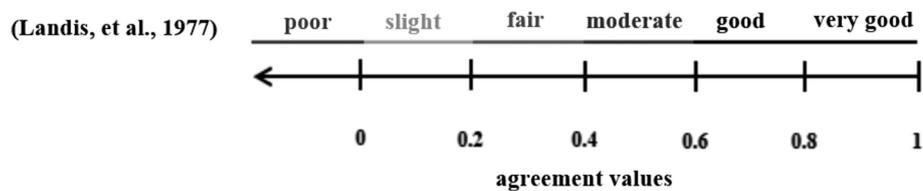


Figure 1. Interpretation of agreement values

Source: Authors based on (Landis & Koch, 1977)

```
<?xml version="1.0" encoding="UTF-8"?>
<TourismSrijem>
- <TEXT>
  <![CDATA[Predivan manastir. Divan ikonostas i ostaci fresaka. U manastiru su mošti cara Lazara. Divan ambijent. ]]>
</TEXT>
- <TAGS>
  <DocumentLevel SentimentPolaritet="positive" SubDestination="Vrdnik"
    DestinationCategory="Monastery" OznakaKomentara="49" komentar=" " text="Predivan
    manastir. Divan ikonostas i ostaci fresaka. U manastiru su mošti cara Lazara. Divan
    ambijent." spans="0~106" id="D0"/>
  <SentenceLevel text="Predivan manastir." spans="0~18" id="S0"
    SentenceSentiment="positive" Aspect="MonasteryGeneral"
    DiscourseFunction="Sentiment"/>
  <SentenceLevel text="Divan ikonostas i ostaci fresaka." spans="20~53" id="S1"
    SentenceSentiment="positive" Aspect="Decoration" DiscourseFunction="Sentiment"/>
  <SentenceLevel text="U manastiru su mošti cara Lazara." spans="55~88" id="S2"
    SentenceSentiment="neutral" Aspect="MonasteryGeneral" DiscourseFunction="Fact"/>
  <SentenceLevel text="Divan ambijent." spans="90~105" id="S3"
    SentenceSentiment="positive" Aspect="MonasteryGeneral"
    DiscourseFunction="Sentiment"/>
  <WordLevel text="Predivan" spans="0~8" id="W0" WordType="PositiveExpression"/>
  <WordLevel text="Divan" spans="20~25" id="W1" WordType="PositiveExpression"/>
  <WordLevel text="Divan" spans="90~95" id="W2" WordType="PositiveExpression"/>
</TAGS>
</TourismSrijem>
```

Figure 2. Illustration of annotation

Source: Authors

⁵ Multi-document Annotation Environment. (MAE). Available at: <http://keighrim.github.io/mae-annotation/>

Based on the values presented in the following sub-sections, we could conclude that annotators achieved good or very good agreement on most of the annotation attributes and that we can proceed to the next steps. The last phase of the annotation process refers to the independent annotation of the remaining of the corpus. Figure 2 illustrates the result of the annotation of one online review.

3.3.1. *Inter-rater agreement on document level*

To evaluate agreement among four annotators on document-level, we used Fleiss' kappa statistic (Fleiss, 1971; Fleiss et al., 2003), which is appropriate in cases of more than two annotators with a proportional number of annotations. According to (Landis & Koch, 1977), Figure 1, we can conclude that annotators achieved almost perfect agreement while annotating *Destination category* and *Sub destination* with kappa = 0.996 and kappa = 0.978, respectively, meaning that annotators agreed in 99.6% and 97.8% of cases. While annotating Sentiment polarity annotators agreed in 72.8% cases, i.e. kappa = 0.728. These values indicate good and satisfactory agreement among annotators and that they can proceed to independent annotations of the remaining of the corpus.

3.3.2. *Inter-rater agreement on sentence level*

To evaluate agreement among four annotators on the sentence level, we used Fleiss' kappa statistic. According to (Landis & Koch, 1977), Figure 1, we can conclude that annotators achieved good agreement while annotating *Aspect* and *Sentence sentiment polarity* with kappa = 0.735 and kappa = 0.738, respectively, meaning that annotators agreed in 73.5% and 73.8% of cases. While annotating *Discourse function*, annotators achieved moderate agreement, i.e. kappa = 0.482. If we consider agreement among 3 annotators (eliminating the second annotator with the most deviations) on *Discourse function*, the agreement rises for 5.3%, and kappa is 0.535. Error analysis has provided insight into the values annotators had the most difficulties to distinguish. These are discourse functions *Advantage*, *Suggestion*, *Conclusion*, *Summary*, and *Other*.

3.3.3. *Inter-rater agreement on word level*

To evaluate agreement among four annotators on word-level, we used *agr* metric, which is appropriate in case of multiple annotators with different numbers of annotations for the given annotation attribute (Wiebe et al., 2005). Analyzing agreement among four annotators on *Positive expressions*, *Negative expressions*, *Negation keyword*, and *Negation scope* we could observe that annotators have achieved almost perfect agreement when it comes to *Negation keyword* (F-measure = 0.949), *Positive expressions* (F-measure = 0.910), and *Negative expressions* (F-measure = 0.795, i.e. 0.8). Good agreement is achieved for *Negation scope* (F-measure = 0.769). Analyzing agreement on *Inverters*, *Neutralizers*, and *Intensifiers* we could observe that annotators have achieved good agreement on *Inverters* (F-measure = 0.759) and borderline good on *Intensifiers* (F-measure = 0.664), while a moderate agreement is achieved on *Neutralizers* (F-measure = 0.529).

3.4. Statistical analysis of annotated corpus

The collected corpus comprises 1901 documents (reviews) with a total vocabulary of 16281 words from which 3870 words are unique (4% of overall vocabulary). This points to the lack of variations in the vocabulary used to express an opinion regarding tourism destinations and attractions, which can facilitate the identification of opinionated expressions. The longest docu-

ment has 127 words (followed by a review with 122 words, 108 words, 104, and 91 words – top five documents according to their length), while the shortest has only one word.

The statistics also point to the number of sentences in our corpus (2746) and the number of words in our corpus with different functions (5369). The number of annotated words naturally differ from the size of vocabulary, given that word-level annotations refer to words which are indicators of sentiment or indicators of different functions (sentiment expression, intensifiers, neutralizers, and so on), while vocabulary size is the overall number of words in our corpus and therefore vocabulary also encompasses words without any particular function.

On document-level annotators marked Destination category and Sentiment polarity. In total they made 940 (49.45%) annotations of Nature destination category, 331 (17.41%) annotations of Natural reserve, 325 (17.10%) annotations of Monastery, and 305 (16.04%) annotations of Museum. Positive sentiment is annotated in 1745 (91.79%) documents, negative is marked over 57 (3%) documents, both in 99 (5.21%) documents, while neutral is not present in the corpus. We can observe that sentiment is highly skewed towards positive sentiment, which is not uncommon for online reviews (Broß, 2013; Park et al., 2015). Document-level sentiment gives insight into the prevailing sentiment of the overall review and, as such, is not sufficient to draw conclusions on peculiarities of satisfaction and dissatisfaction of visitors of particular destinations.

On sentence-level annotators made 2746 different annotations: 1260 for Nature, 541 for Natural reserve, 492 for Monastery, and 453 for Museum. Naturally, the most sentence-level annotations are present in destination category *Nature*, given that most reviews are collected from this category.

Discourse function gives insight into a type of opinion, comment, or review people provide. The distribution of annotations over discourse functions as a percentage of their share in the total number of annotations is as follows: Advantage 5.30%, Advice 1.82%, Comparison 0.44%, Conclusion 0.58%, Expectation and Requirement 0.40%, Fact 5.05%, Lack or Problem 3.81%, Suggestion 16.01%, Personal Content 0.22%, Purchase or Usage 2.61%, Sentiment 55.74%, Summary 5.95%, Other 2.07%. Based on the provided data, we can conclude that people most often express their overall sentiment towards a certain destination or tourism attraction, then they offer suggestions for improvements, emphasize advantages, or summarize the overall impression. Figure 3 illustrates the distribution of discourse function over different destination categories. We can observe that people mostly use facts in reviews about monasteries. Also, while writing about monasteries, people tend to make comparisons, which is less common for other types of destinations. People tend to give more often bits of advice and suggestions while reviewing Fruška gora national park and other nature reserves than they do in reviews of other destination categories. People either give others advice on what to visit and how to spend time, suggestions are commonly given on how to overcome some deficiency they have identified or how to overcome some negative aspects to make it more enjoyable for tourists.

Another important attribute on the sentence-level annotation is *Aspect*. The distribution of each *Aspect* as their share in a total number of annotations is as follows: NatureGeneral 44.92%, MonasteryGeneral 16.23%, MuseumGeneral 12.27%, Cleanliness 1.53%, Facility 5.74%, Recreation 5.81%, Location 2.25%, Parking 0.15%, Service 5.56%, Price 0.80%, Decoration/Landscaping 0.44%, Other 4.32%. We can observe that certain aspects are underrepresented, such as *Parking*, *Decoration and landscaping*, and *Price*. This points to the possibility of joining certain aspects. In future work, the following aspects could be joined: 1) *Location* and *Parking*, and

2) *Service, Price, Cleanliness, and Decoration*. Figure 4 illustrates the distribution of aspects over different destination categories. Based on the presented data, we can conclude that *Aspects* have different importance in reviews of different destination categories. *Facility* and *Recreation* are more important for destinations and attractions in nature, while the kind of service provided to visitors is more important at museums and nature reserves.

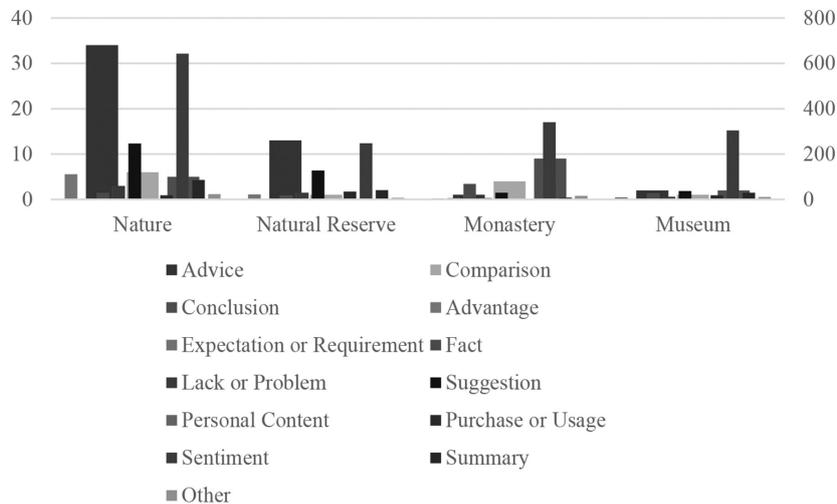


Figure 3. Discourse functions according to destination category
 Source: Authors

Annotators were labelling sentiment on sentence-level as well. Positive sentiment is marked in 2340 (85.21%) sentences, negative in 176 (6.41%) sentences, both in 82 (2.99%) sentences, and neutral in 148 (5.39%) sentences. Reviews are highly skewed towards positive sentiment, which is not uncommon for online reviews (Broß, 2013; Park et al., 2015). Figure 5 illustrates the distribution of sentiment polarity on sentence level over destination categories. Based on the presented data, we can observe that nature is more criticized than monasteries and museums.

On word-level, we had various annotation attributes. The overview of the total number of annotations, as well as the number of unique annotations, is provided in Table 4.

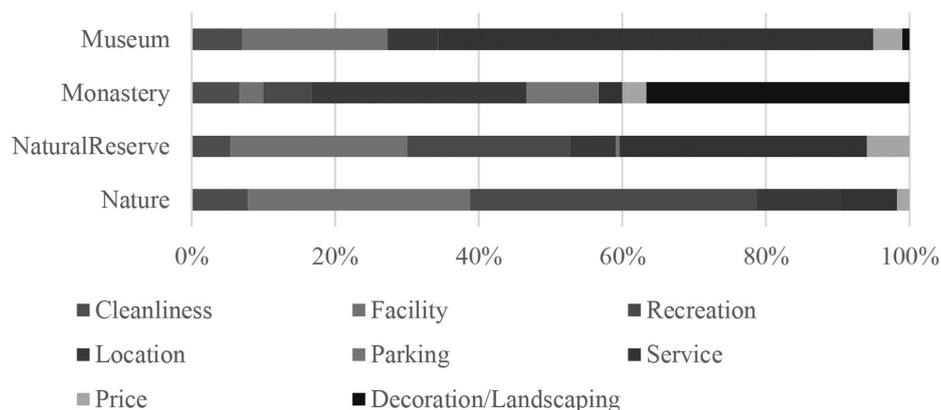


Figure 4. Distribution of aspects over destination categories
 Source: Authors

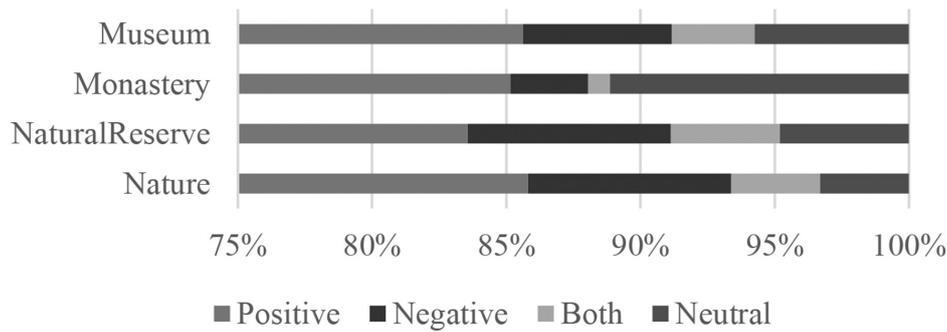


Figure 5. Overview of sentence sentiment annotations

Source: Authors

Table 4. Overview of annotations on word level

	Total number of annotations	Number of unique annotations
Sentiment expressions: 4225		
Positive	3893	1049
Negative	332	231
Functional words: 781		
Intensifiers	577	129
Neutralizers	121	25
Inverters	83	15
Negation		
Negation keywords	170	21

Source: Authors

We can observe that people use richer vocabulary while expressing negative attitudes, 69.55% of used negative sentiment expressions (231) are unique in our corpus, while only 26.95% (1049) of annotated positive sentiment expressions are unique. The rest is repeated throughout the corpus. People use more often words that intensify expressed sentiment compared with the words used to downtone the sentiment. Inverters are used to shift expressed sentiment from one to the other polarity. Although they occur quite often in our corpus, only 15 different words play this role. The same applies to negation. Negation is detected 170 times in our corpus, while only 21 different words are used for this purpose. These annotations are not used in the analysis presented in the rest of the paper. However, they are used to develop valuable resources for sentiment analysis in the form of dictionaries. Dictionaries are deployed in the dictionary- and rule-based sentiment analysis, where hand crafted linguistic rules are used to map occurrences in online reviews with the values from dictionaries. Based on such mappings and word order in sentences, automated systems can generate the overall sentiment polarity expressed in reviews.

4. ASSESSMENT OF PUBLIC OPINION ON TOURISM ATTRACTIONS AND DESTINATIONS IN SREM/SRIJEM REGION

In the research presented in this paper, we have used annotations made over the corpus for further analysis in order to draw insights about public stance on various tourism destinations and attractions in the S-S region.

4.1. Analysis of the overall public stance

On document level, we have reviews from four different categories: monasteries, nature, nature reserves, and museums, and in the following of this section, we offer more detailed analysis according to sub-destinations and sentiment expressed towards sub-destinations in each destination category. Such analysis provides an overview of the overall public stance on tourism attractions in the S-S region.

Figure 6 illustrates the distribution of public sentiment over each analyzed location from the Monasteries category. We can notice that people predominantly have a positive attitude towards monasteries. However, to some extent, criticism is present in the case of several monasteries. Such criticism is also enveloped in a mixed review where people usually state an overall positive impression and point to some negative aspect within. In these terms, negativity within mixed reviews is mostly present in Krušedol, where people criticize luxurious elements on the monastery and commercialization of the monastery. Also, people, in general, resent the insufficiently transparent signage for reaching the monastery as well as the poor organization of the parking space or its absence. More details on the discussed aspects are presented in Section 4.2.

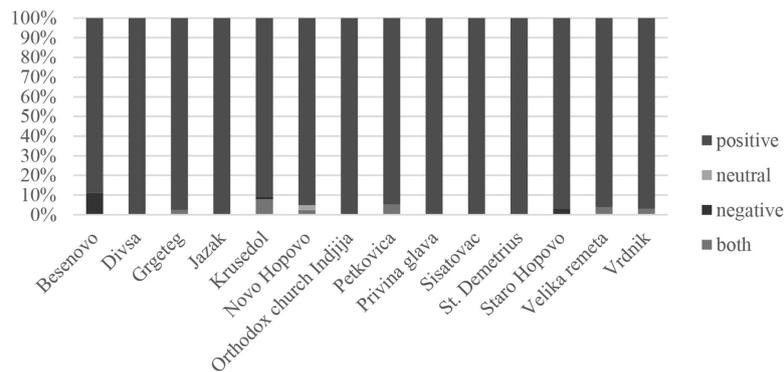


Figure 6. Distribution of public sentiment over a particular monastery in the Srem/Srijem region

Source: Authors

Figure 7 illustrates the distribution of reviews according to polarity over various destinations from the category *Museum* in Srem/Srijem region. We can observe from the presented data that there are more negative and mixed reviews than in the category of *Monasteries*. Ilok castle, Eltz, and Beekeeping museums are three destinations with the most negative reviews within this category. Mostly people criticize services in terms of price or hospitality. More details on the discussed aspects are presented in Section 4.2. Mixed reviews are present in most reviews.

Analysis of reviews according to their sentiment polarity and various destinations from the category of *Nature reserves* in the S-S region reveals the following distribution: 10% of Borkovac lake reviews comprise both positive and negative sentiment while it is not commented in purely negative manner; 3.13% of reviews of Obedska bara are purely negative and 4.69% comprise both sentiments; 2.14% of Zasavica reviews are purely negative and 5.35% comprise both sentiments. We can observe that on occasion Obedska bara and Zasavica have been commented in a purely negative manner, while all three sub destinations have reviews with mixed sentiments, meaning that people have expressed both positive and negative sentiments within one comment. People have commented in negative manner facilities, safety on facilities, as well as the servic-

es. More details on the discussed aspects are presented in Section 4.2. Public stance towards Fruška gora is predominantly positive, 90.02% of comments comprise positive sentiment. However, there is approximately 10% of comments containing negativity, and these are annotated as either both (5.41% reviews containing both positive and negative sentiment) or negative (4.35% of reviews). Only small portion of reviews (0.21%) is neutral.

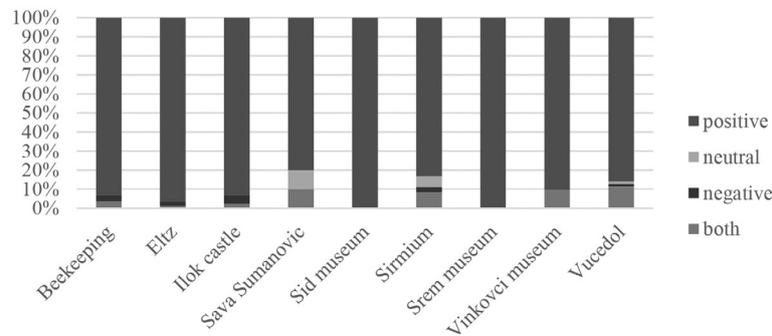


Figure 7. Distribution of public sentiment over particular museums in Srem/Srijem region

Source: Authors

4.2. Analysis of public stance according to different aspects

Although the analysis on document level is useful to extract insights on general public perception and sentiments, it does not offer sufficient information nor details about expressed sentiments. Analysis of sentence-level annotations allows more detailed observation of public opinion and extraction of insights about aspects mostly discussed in a positive and negative context.

Figure 8 illustrates public opinion according to aspects discussed in reviews from the *Monasteries* category. We can observe presence of two hi-level aspects MonasteryGeneral and NatureGeneral since most of the monasteries are situated in the nature of Fruška Gora. When writing their general impression about monasteries, reviewers are on occasion negative. Based on data, we can conclude that public stance is always negative towards pricing, that people often discuss a lack of parking or organization of parking space in a negative manner, and often the target of criticism is luxurious decoration in monasteries. Facilities are often subject to mixed sentiment, meaning the people mention both positive and negative aspects of facilities.

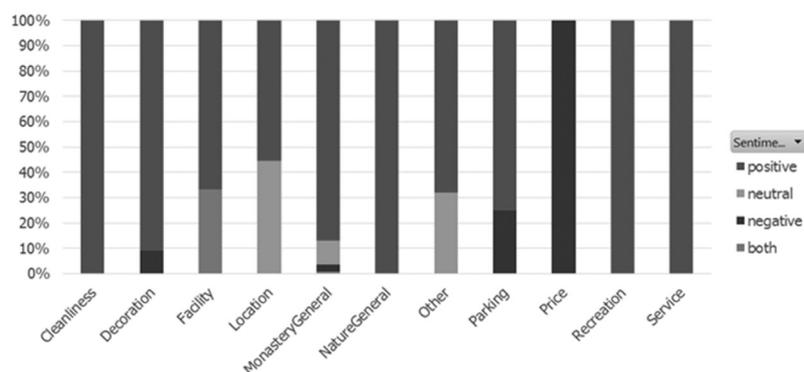


Figure 8. Distribution of public sentiment according to discussed aspect – category Monastery

Source: Authors

Figure 9 offers insight into the discourse function used to express sentiments. We are particularly interested in the way negativity is expressed as negative comments serve as direction for improvements. Negativity directed towards monasteries in Srem/Srijem region is expressed within identified lacks and problems and if visitor's expectations are not fulfilled.

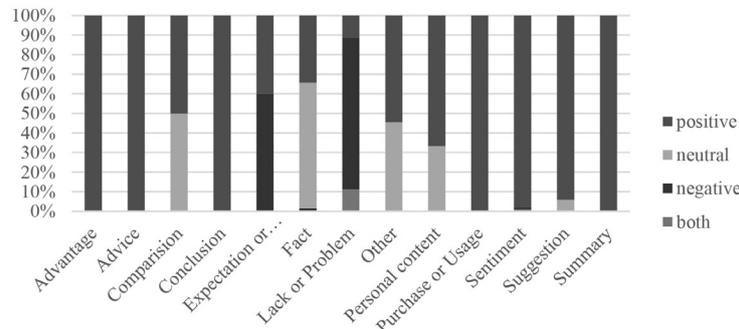


Figure 9. Category Monastery: an overview of discourse function according to expressed sentiment

Source: Authors

Figure 10 illustrates the public stance towards museums in Srem/Srijem region. We can observe presence of all three hi-level aspects: MuseumsGeneral, NatureGeneral, and MonasteryGeneral. Reviewers on occasion express their general impression about the museum in negative manner, while other two hi-level aspects are mentioned in purely positive way. Based on the presented data, we can observe that people are mostly unsatisfied with the cleanliness of museums, available facilities, price, and to some extent, services. Often people just state what the location of the museum is, and that particular museum charges a certain amount without expressing the particular sentiment, and such sentences are marked as neutral.

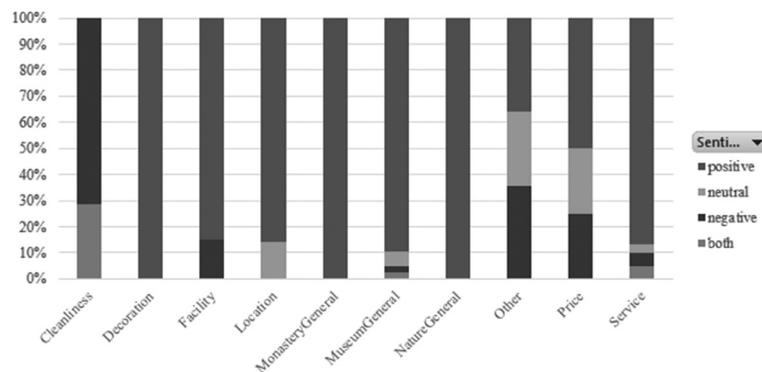


Figure 10. Distribution of public sentiment according to discussed aspects – category Museums

Source: Authors

Figure 11 illustrates the discourse functions used to express sentiments towards museums. Based on the presented data, we can conclude that negativity is mostly encompassed within pieces of advice. This is not surprising given that people often write reviews offering advice or their idea on how to overcome some issues they have experienced. Also, negativity is part of identified lacks and problems or purchasing experience. Other discourse functions are not used to express negativity, such as conclusion, comparison, advantage, or have been used to some extent in mixed reviews.

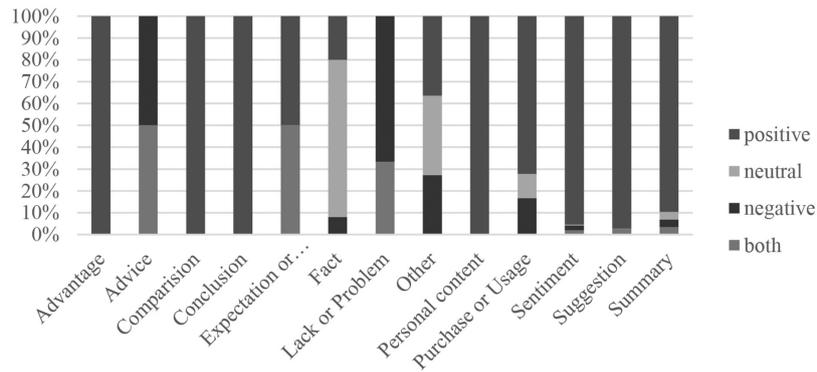


Figure 11. Category Museums: an overview of discourse function according to expressed sentiment

Source: Authors

Figure 12 illustrates sentiment expressed towards different aspects of *Nature reserves* (Zasavica, Obedska bara, and Borkovac lake). We can observe only one hi-level aspect – NatureGeneral which is on occasion discussed in negative manner. Based on available data, we can see that the only aspect which is not discussed in a negative context is parking. Cleanliness, service and price are three aspects with the most negative comments, followed by facilities.

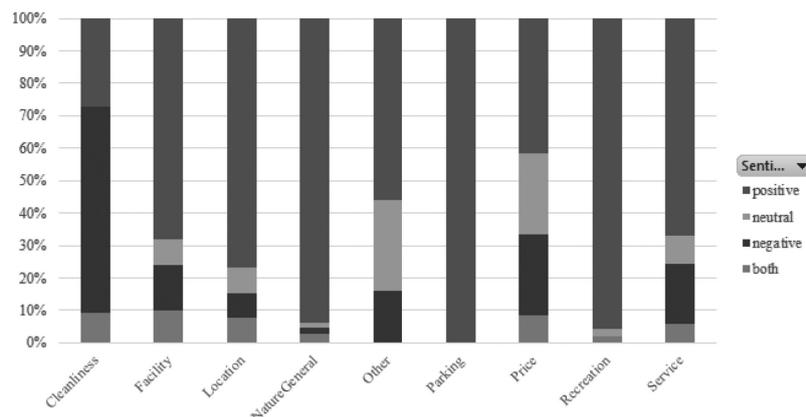


Figure 12. Distribution of public sentiment according to discussed aspects – a category Nature reserve

Source: Authors

Figure 13 illustrates the discourse function used to express sentiment in online reviews targeting nature reserves. Expectations and requirements, as well as lack or a problem, are completely negative, meaning the people use these discourse functions to state the negative attitude. The other discourse functions used to express negativity are purchase and advice.

Category nature encompassed online reviews about Fruška gora. Figure 14 indicates presence of two hi-level aspects: MonasteryGeneral, which is discussed in purely positive manner, and NatureGeneral, which is on occasion discussed in negative manner. In reviews about Fruška Gora, negativity is mostly directed towards cleanliness, prices, facilities, and services, Figure 14. These aspects are commented by visitors in a purely negative manner or in a mixture of positive and negative comments.

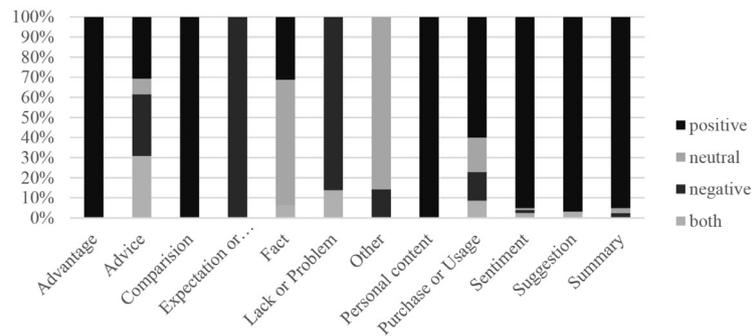


Figure 13. Category Nature reserve: an overview of discourse function according to expressed sentiment

Source: Authors

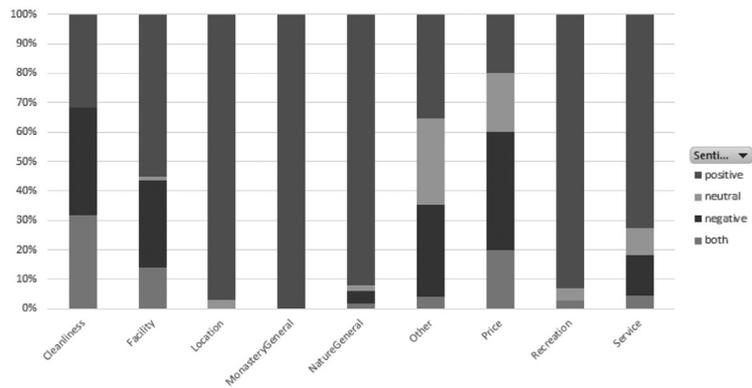


Figure 14. Distribution of public sentiment according to discussed aspects – category Nature/Fruška gora

Source: Authors

Figure 15 presents the distribution of reviews of Fruška gora according to expressed sentiments and discourse functions. As in other categories, lack or problem and expectation or requirement are key discourse functions used to express negativity. Often when people offer advice, it is also an indicator of their dissatisfaction given that people write advice to state their opinion on how to overcome some issues. When a review concludes with an expression of negative sentiment, it is mostly an indicator of the highly dissatisfied visitor. This discourse function is used for expression of negativity only in reviews from category nature, i.e. reviews about Fruška gora.

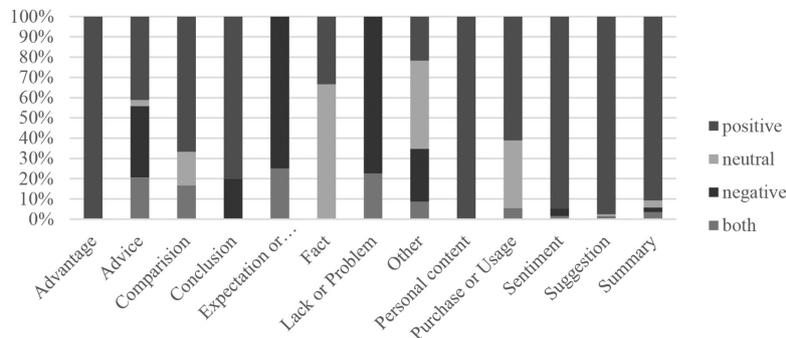


Figure 15. Category Nature: an overview of discourse function according to expressed sentiment

Source: Authors

Overall, we can observe differences in public stance depending on the category of destination, which is discussed. The price is the subject of negative comments in reviews from all four destination categories. However, the price is perceived completely in a negative manner in reviews about monasteries, while in reviews of other types of destinations, it is part of positive or mixed reviews as well. Category Monasteries differs from other types of destination in the way people perceive parking space and decoration. These aspects are predominantly negative within this category of reviews, while they are not commented in such a way in other types of destinations. Not only that, the decoration is discussed negatively in reviews from category Monasteries, but also decoration appears only in reviews from this destination category. If people discuss cleanliness in reviews about Museums or Nature reserves, they do it in a purely negative way. Cleanliness is also often discussed negatively in reviews about Nature, but in this category, contrary to Museums and Nature reserves, it is also sometimes praised. In reviews about museums, nature reserves, and nature, people most often criticize facilities and services. Another deviation from other categories is a predominantly negative perception of location in reviews from category Nature reserves. Location is not discussed in this manner in reviews from other destination categories interestingly, if people comment on parking in reviews of Nature reserves, they do it in a purely positive manner, while all other aspects in this destination category are, to some extent, discussed in a negative manner.

Analysis of discourse functions according to sentiment and destination category pointed to *Lack and problem* and *Expectation and requirements* as two essential discourse functions for expressing a negative stance in all four destination categories. From the point of automation and development of automated systems for detection and analysis of public stance, it would be beneficial to learn patterns of composing of these two key discourse functions. Advice is used in purely negative reviews in category Museums, and to some extent, pieces of advice are negative in reviews from category Nature reserves and Nature. Purchase and usage have revealed as a useful discourse function for detection of negativity expressed towards museums and nature reserves, while Conclusion and Comparisons are present only in negative reviews about nature. These are useful insights for future research as they can be used to model different discourse functions in terms of the way sentiment is expressed within and to generate linguistic rules useful for automated detection of such discourse functions.

5. CONCLUSION

In this paper, we have described and presented methodological steps for creating an annotated corpus as a critical part of natural language analysis and the development of automated systems for the detection and analysis of public sentiments. We have collected 1910 online reviews about 30 destinations and attractions in Srem/Srijem region that we have grouped into four major destination categories: Monasteries, Museums, Nature, and Nature reserves. Such a collection of reviews is called a corpus. The corpus has been pre-processed in terms of data cleansing, spelling mistakes, interrogative and factual reviews were removed, and reviews written in foreign languages were translated into Serbian. The annotation process, or the process of labelling collected reviews, represents the key step in the transformation of collected reviews into a useful and meaningful dataset for further analysis and development of automated sentiment analysis systems. Corpus was labelled with different annotation attributes on three levels: at the level of the overall review, i.e. document-level, at the sentence level, and at the level of individual words or expressions. In this way, corpus of reviews on destinations and attractions in Srem/Srijem region was significantly enriched with useful information on sentiment, locality, aspect which is discussed in review, discourse function which reveals how particular expressions are used in various occasions, as well as information on different word functions (intensifiers, neutralizers,

inverters, negation, positive sentiment expression, negative sentiment expression). The annotation was carried out by four annotators. Evaluation of their work and the quality of annotation indicated that annotators are in a good or a very good agreement for most of the annotation attributes, which was the indicator of high quality of conducted work.

Resulting annotations were used for the analysis of public stance on various localities in the Srem/Srijem region. Overall, the public stance is positive given that most reviews are labeled with the positive sentiment on document-level (91.79%), as well as the majority labels for sentiment polarity on sentence-level are positive (85.21%). Such distribution of sentiment polarity in a corpus of online reviews is not atypical. Positive impressions could be useful for conceptualizing offers to future visitors, and they can serve as an indicator of a preferred characteristic of the locality in the Srem/Srijem region, which could be emphasised in the offer or within the marketing message. However, from the business point of view, it is more important to address the negativity and dissatisfaction of visitors in order to neutralize the negative stance of visitors and negativity within the message they share. The current distribution of sentiment polarity within our corpus is a clear indicator of the level of analysis we need to observe when addressing negative public stance. Document-level is not sufficient as it reveals insights about the overall public stance, and it is predominantly positive. Sentence-level is more useful as it can reveal insights about aspects tackled in negative sentences, as well as the insights about discourse functions used to express negativity. In the analysis and interpretation of public stance on localities in the Srem/Srijem region, we have used annotations made on sentence-level. Such analysis revealed sources of dissatisfaction among visitors and potential directions for improvements. Visitors mostly criticize prices and facilities, and these aspects are common for all four categories of destinations. However, we have uncovered some peculiarities. If people write about parking and decoration of monasteries, they will do it in a negative manner. These two aspects are discussed in a negative manner only in this category of destinations. Cleanliness and services are very important to visitors to museums, nature reserves, and Fruška gora, and they are extensively criticized by visitors. Location is discussed in a negative manner only in reviews of nature reserves. We have also identified that people emphasize negative aspects when addressing some lacks, problems, expectations, or requirements which were not fulfilled, and also when they offer advice on how to overcome some issues. These are valuable insights into discourse functions used in the tourism domain to express a negative stance, as well as for future analysis of discourse function and the way linguistic constructions are used to express sentiments. Such analysis will provide a basis for crafting linguistic rules which could be used as a part of rule-based sentiment analysis systems, which can automatically identify sentiment expressions in texts and conduct their analysis. Although the analysis provided in this paper is fully based on annotation attributes added over collected reviews, the resulting corpus with annotations for sentiments and aspects represents a valuable resource for sentiment analysis of texts written in the Serbian language, given that such corpora lack for our language. Therefore, the resulting annotated corpus will be used for the development of sentiment analysis system based on machine learning algorithms for automated detection of sentiments and aspects in reviews written in Serbian about various tourism destinations in Srem/Srijem region, as well as for further research in the domain of natural language processing and sentiment analysis.

ACKNOWLEDGMENT

This paper is the result of project Virtual and Cultural Tourism, ViCTour, HR-RS47, co-financed by INTERREG IPA CBC Croatia-Serbia 2014-2020 program and Provincial Secretariat of Finance, Autonomous Province of Vojvodina, 102-401-337/2017-02-4-81-28.

REFERENCES

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Professional.
- Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, 55 (3), 261-271.
- Bollen, J., Mao, H., & Zeng, X.J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.
- Broß, J. (2013). *Aspect-oriented sentiment analysis of customer reviews using distant supervision techniques*. Berlin: Freie Universität Berlin PhD thesis.
- Chung, S.F. (2011). Uses of term- in Malay: A corpus-based study. *Journal of Pragmatics*, 43(3), 799-813. doi: 10.1016/j.pragma.2010.10.004
- Dalal, M.K., & Zaveri, M.A. (2014). Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges. *Applied Computational Intelligence and Soft Computing*, Article ID 735942, 9 pages. doi:10.1155/2014/735942.
- D'Avanzo, E., & Pilato, G. (2015). Mining social network users' opinions to aid buyers' shopping decisions. *Computers in Human Behavior*, 51, 1284-1294.
- Dave, K., Lawrence, S., & Pennock, D.M. (2003). *Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews*. Paper presented at the 12th international conference on World Wide Web, Budapest, Hungary, 20-24 May.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-82.
- Fleiss, J. L., Levin, B., & Paik, M.C. (2003). *Statistical Methods for Rates and Proportions*. Wiley.
- Ganu, G., Elhadad, N., & Marian, A. (2009). *Beyond the stars: Improving rating predictions using Review Text Content*. Paper presented at the 12th International Workshop on the Web and Databases (WebDB 2009), 9, 1-6. Providence, Rhode Island, USA, 28 June.
- Garrote, M., Kimura, C., Matsui, K., Sandoval, M.A., & Takamori, E. (2013). C-ORAL-JAPON: Corpus of Spontaneous Spoken Japanese. *Corpus Linguistics and Linguistic Theory*, 11(2), 373-392. doi: 10.1515/cllt-2013-0004.
- Gligorijevic, B., & Luck, E. (2012). *Engaging Social Customers –Influencing New Marketing Strategies for Social Media Information Sources*. In Contemporary Research on E-business Technology and Strategy (pp. 25-40). Springer Berlin Heidelberg.
- González, A.O. & Ramos, A.M. (2013). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish. *Procedia - Social and Behavioral Sciences*, 95(2013), 563-570.
- Gruen, T. W., Osmonbekov, T., & Czaplewski, A. J. (2006). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research*, 59(4), 449-456.
- Hammond, M. (2015). Predicting the gender of Welsh nouns. *Corpus Linguistics and Linguistic Theory*, 12(2), 221-261. doi: 10.1515/cllt-2015-0001.
- Henricson, S., & Nelson, M. (2017). Giving and receiving advice in higher education. Comparing Sweden-Swedish and Finland-Swedish supervision. *Journal of Pragmatics*, 109, 105-120. doi: <https://doi.org/10.1016/j.pragma.2016.12.013>.
- Hovy, E., & Lavid, J. (2010). Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics. *International Journal of Translation*, 22, 13-36.
- Hu, M. & Liu, B. (2004). *Mining and summarizing customer reviews*. Paper presented in the Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04, 168-177, Seattle, WA, USA, 22-25 August.

- Iosif, E., Klasinas, I., Athanasopoulou, G., Palogiannidi, E., Georgiladakis, S., Louka, K., & Potamianos, A. (2018). Speech understanding for spoken dialogue systems: From corpus harvesting to grammar rule induction. *Computer Speech and Language*, 47, 272-297. doi: <https://doi.org/10.1016/j.csl.2017.08.002>.
- Iruskieta, M., Ilarraza deDiaz, A., & Lersundi, M. (2013). Establishing criteria for RST-based discourse segmentation and annotation for texts in Basque. *Corpus Linguistics and Linguistic Theory*, 11(2), 303-334. doi:10.1515/cllt-2013-0008.
- Kessler, J.S., Eckert, M., Clark, L., & Nicolov, N. (2010). *The 2010 ICWSM JDPa sentiment corpus for the automotive domain*. Paper presented at the 4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010), Washington, DC, USA, 23-26 May.
- Kessler, J. S., & Nicolov, N. (2009). *Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations*. Proceedings of the Third International AAAI Conference on Weblogs and Social Media, (pp. 90-7).
- Kim, J., Choi, D., Hwang M., & Kim, P. (2014). Analysis on Smartphone Related Twitter Reviews by Using Opinion Mining Techniques. *Advanced Approaches to Intelligent Information and Database Systems, Studies in Computational Intelligence*, 551, 205-212.
- Landis, R.J., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-74.
- Lei, S., & Law, R. (2015). Content Analysis of TripAdvisor Reviews on Restaurants: A Case Study of Macau. *Journal of tourism*, 16(1), 17-28.
- Lin, Y.L. (2017). Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. *Journal of Pragmatics*, 117, 155-167. doi: <https://doi.org/10.1016/j.pragma.2017.06.014>.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Liu, Y., X. Huang, A., & Yu, X. (2007). ARSA: *A Sentiment-Aware Model for Predicting Sales Performance Using Blogs*. Paper presented at the SIGIR'07, 607-614, Amsterdam, Netherlands, 23-27 July.
- Öğüt, H., & Onur, T.B.K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry, *The Service Industries Journal*, 32(2), 197 – 214.
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?: Sentiment classification using machine learning techniques*. Paper presented at the ACL-02 Conference on Empirical Methods in Natural Language Processing, 10, 79-86, University of Pennsylvania, Philadelphia, PA, USA, 6-7 July.
- Park, D.-H., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 11 (4), 125-148.
- Park, S., Lee, W., & Moon, I.-C. (2015). Efficient extraction of domain specific sentiment lexicon with active learning. *Pattern Recognition Letters*, 56, 38-44.
- Pitt, L. F., Berthon, P. R., Watson, R. T., & Zinkhan, G. M. (2002). The Internet and the birth of real consumer power. *Business Horizons*, 45(4), 7-14.
- Pustejovsky, J., & Moszkowicz, J. (2012). *The Role of Model Testing in Standards Development: The Case of ISO-Space*. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12) (pp. 3060-3063). Istanbul, Turkey: European Language Resources Association (ELRA).

- Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning - a Guide to Corpus-Building for Applications*. O'Reilly Media Inc.
- Reinstein, D.A., & Snyder, C.M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie reviews. *Journal of Industrial Economics*, 53, 27-51.
- Rumenapp, J.C. (2016). Analyzing discourse analysis: Teachers' views of classroom discourse and student identity. *Linguistics and Education*, 35, 26-36. doi: 10.1016/j.linged.2016.04.002.
- Stubbs, A. (2012). *Developing Specifications for Light Annotation Tasks in the Biomedical Domain*. Proceedings of the Third Workshop on Building and Evaluating Resources for Biomedical Text Mining, (p. 71). Istanbul, Turkey.
- Tang, X. (2016). Lexeme-based collexeme analysis with DepCluster. *Corpus Linguistics and Linguistic Theory*, 13(1), 165-202. doi:10.1515/cllt-2015-0007.
- Van de Kauter, M., Breesch, D., & Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications*, 42, 4999–5010.
- Vass, H. (2017). Lexical verb hedging in legal discourse: The case of law journal articles and Supreme Court majority and dissenting opinions. *English for Specific Purposes*, 48(2017), 17-31.
- Wachsmuth, H., Trenkmann, M., Stein, B., Engels, G., & Palakarska, T. (2014). *A Review Corpus for Argumentation Analysis*. Paper presented at the 15th International Conference on Computational Linguistics and Intelligent Text Processing, Kathmandu, Nepal, 06-12 April.
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3), 165-210.
- Wissler, L., Almashraee, M., Monett, D., & Paschke, A. (2014). *The Gold Standard in Corpus Annotation*. Proceedings of 5th IEEE Germany Student Conference, IEEE GSC 2014, Passau, Germany.
- Yang, C.-S., Chen, C.-H., & Chang, P.-C. (2015). Harnessing consumer reviews for marketing intelligence: a domain-adapted sentiment classification approach. *Information Systems and e-Business Management*, 13 (3), 403-419.