SemEval-2016 Task 5: Aspect Based Sentiment Analysis
Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos,
Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan
Zhao, Bing Qin, Orphée de Clercq, et al.

To cite this version:
Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, et al.. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. International Workshop on Semantic Evaluation, Jan 2016, San Diego, United States. pp.19 - 30, 10.18653/v1/S16-1002 . hal-01838537

HAL Id: hal-01838537
https://hal.archives-ouvertes.fr/hal-01838537
Submitted on 12 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
SemEval-2016 Task 5: Aspect Based Sentiment Analysis

Maria Pontiki¹, Dimitrios Galanis¹, Haris Papageorgiou¹, Ion Androutsopoulos¹,², Suresh Manandhar³, Mohammad AL-Smadi¹, Mahmoud Al-Ayyoub⁴, Yanyan Zhao⁵, Bing Qin⁵, Orphée De Clercq⁶, Véronique Hoste⁶, Marianna Apidianaki⁷, Xavier Tannier⁷, Natalia Loukachevitch⁸, Evgeny Kotelnikov⁹, Nuria Bel¹⁰, Salud Maria Jiménez-Zafra¹¹, Gülşen Eryiğit¹²

¹Institute for Language and Speech Processing, Athena R.C., Athens, Greece, ²Dept. of Informatics, Athens University of Economics and Business, Greece, ³Dept. of Computer Science, University of York, UK, ⁴Computer Science Dept., Jordan University of Science and Technology Irbid, Jordan, ⁵Harbin Institute of Technology, Harbin, Heilongjiang, P.R. China, ⁶LT3, Ghent University, Ghent, Belgium, ⁷LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Orsay, France, ⁸Lomonosov Moscow State University, Moscow, Russian Federation, ⁹Vyatka State University, Kirov, Russian Federation, ¹⁰Universitat Pompeu Fabra, Barcelona, Spain, ¹¹Dept. of Computer Science, Universidad de Jaén, Spain, ¹²Dept. of Computer Engineering, Istanbul Technical University, Turkey

Abstract

This paper describes the SemEval 2016 shared task on Aspect Based Sentiment Analysis (ABSA), a continuation of the respective tasks of 2014 and 2015. In its third year, the task provided 19 training and 20 testing datasets for 8 languages and 7 domains, as well as a common evaluation procedure. From these datasets, 25 were for sentence-level and 14 for text-level ABSA; the latter was introduced for the first time as a subtask in SemEval. The task attracted 245 submissions from 29 teams.

1 Introduction

Many consumers use the Web to share their experiences about products, services or travel destinations (Yoo and Gretzel, 2008). Online opinionated texts (e.g., reviews, tweets) are important for consumer decision making (Chevalier and Mayzlin, 2006) and constitute a source of valuable customer feedback that can help companies to measure satisfaction and improve their products or services. In this setting, Aspect Based Sentiment Analysis (ABSA) - i.e., mining opinions from text about specific entities and their aspects (Liu, 2012) - can provide valuable insights to both consumers and businesses. An ABSA method can analyze large amounts of unstructured texts and extract (coarse- or fine-grained) information not included in the user ratings that are available in some review sites (e.g., Fig. 1).

Sentiment Analysis (SA) touches every aspect (e.g., entity recognition, coreference resolution, negation handling) of Natural Language Processing (Liu, 2012) and as Cambria et al. (2013) mention “it requires a deep understanding of the explicit and implicit, regular and irregular, and syntactic and semantic language rules”. Within the last few years several SA-related shared tasks have been organized in the context of workshops and conferences focus-
ing on somewhat different research problems (Seki et al., 2007; Seki et al., 2008; Seki et al., 2010; Mitchell, 2013; Nakov et al., 2013; Rosenthal et al., 2014; Pontiki et al., 2014; Rosenthal et al., 2015; Ghosh et al., 2015; Pontiki et al., 2015; Mohammad et al., 2016; Recupero and Cambria, 2014; Ruppenhofer et al., 2014; Loukachevitch et al., 2015). Such competitions provide training datasets and the opportunity for direct comparison of different approaches on common test sets.

Currently, most of the available SA-related datasets, whether released in the context of shared tasks or not (Socher et al., 2013; Ganu et al., 2009), are monolingual and usually focus on English texts. Multilingual datasets (Klinger and Cimiano, 2014; Jiménez-Zafra et al., 2015) provide additional benefits enabling the development and testing of cross-lingual methods (Lambert, 2015). Following this direction, this year the SemEval ABSA task provided datasets in a variety of languages.

ABSA was introduced as a shared task for the first time in the context of SemEval in 2014; SemEval-2014 Task 4\(^1\) (SE-ABSA14) provided datasets of English reviews annotated at the sentence level with aspect terms (e.g., “mouse”, “pizza”) and their polarity for the laptop and restaurant domains, as well as coarser aspect categories (e.g., “food”) and their polarity only for restaurants (Pontiki et al., 2014). SemEval-2015 Task 12\(^2\) (SE-ABSA15) built upon SE-ABSA14 and consolidated its subtasks into a unified framework in which all the identified constituents of the expressed opinions (i.e., aspects, opinion target expressions and sentiment polarities) meet a set of guidelines and are linked to each other within sentence-level tuples (Pontiki et al., 2015). These tuples are important since they indicate the part of text within which a specific opinion is expressed. However, a user might also be interested in the overall rating of the text towards a particular aspect. Such ratings can be used to estimate the mean sentiment per aspect from multiple reviews (McAuley et al., 2012). Therefore, in addition to sentence-level annotations, SE-ABSA16\(^3\) accommodated also text-level ABSA annotations and provided the respective training and testing data. Furthermore, the SE-ABSA15 annotation framework was extended to new domains and applied to languages other than English (Arabic, Chinese, Dutch, French, Russian, Spanish, and Turkish).

The remainder of this paper is organized as follows: the task set-up is described in Section 2. Section 3 provides information about the datasets and the annotation process, while Section 4 presents the evaluation measures and the baselines. General information about participation in the task is provided in Section 5. The evaluation scores of the participating systems are presented and discussed in Section 6. The paper concludes with an overall assessment of the task.

2 Task Description

The SE-ABSA16 task consisted of the following subtasks and slots. Participants were free to choose the subtasks, slots, domains and languages they wished to participate in.

Subtask 1 (SB1): Sentence-level ABSA. Given an opinionated text about a target entity, identify all the opinion tuples with the following types (tuple slots) of information:

- **Slot1: Aspect Category.** Identification of the entity E and attribute A pairs towards which an opinion is expressed in a given sentence. E and A should be chosen from predefined inventories\(^4\) of entity types (e.g., “aptop”, “restaurant”), and attribute labels (e.g., “rating”, “quality”).

- **Slot2: Opinion Target Expression (OTE).** Extraction of the linguistic expression used in the given text to refer to the reviewed entity E of each E#A pair. The OTE is defined by its starting and ending offsets. When there is no explicit mention of the entity, the slot takes the value “null”. The identification of Slot2 values was required only in the restaurants, hotels, museums and telecommunications domains.

- **Slot3: Sentiment Polarity.** Each identified E#A pair has to be assigned one of the following polarity labels: “positive”, “negative”, “neutral” (mildly positive or mildly negative).

\(1\)http://alt.qcri.org/semeval2014/task4/
\(2\)http://alt.qcri.org/semeval2015/task12/
\(3\)http://alt.qcri.org/semeval2016/task5/
\(4\)The full inventories of the aspect category labels for each domain are provided in Appendix A.
An example of opinion tuples with Slot1-3 values from the restaurants domain is shown below: “Their sake list was extensive, but we were looking for Purple Haze, which wasn’t listed but made for us upon request!” → \{cat: “drinks#style_options”, trg: “sake list”, fr: “6”, to: “15”, pol: “positive”\}, \{cat: “service#general”, trg: “null”, fr: “0”, to: “0”, pol: “positive”\}. The variable cat indicates the aspect category (Slot1), pol the polarity (Slot3), and trg the target (Slot2); fr, to are the starting/ending offsets of trg.

**Subtask 2 (SB2): Text-level ABSA.** Given a customer review about a target entity, the goal was to identify a set of \{cat, pol\} tuples that summarize the opinions expressed in the review. cat can be assigned the same values as in SB1 (E#A tuple), while pol can be set to “positive”, “negative”, “neutral”, or “conflict”. For example, for the review text “The So called laptop Runs to Slow and I hate it! Do not buy it! It is the worst laptop ever “, a system should return the following opinion tuples: \{cat: “laptop#general”, pol: “negative”\}, \{cat: “laptop#operation_performance”, pol: “negative”\}.

**Subtask 3 (SB3): Out-of-domain ABSA.** In SB3 participants had the opportunity to test their systems in domains for which no training data was made available; the domains remained unknown until the start of the evaluation period. Test data for SB3 were provided only for the museums domain in French.

### 3 Datasets

#### 3.1 Data Collection and Annotation

A total of 39 datasets were provided in the context of the SE-ABSA16 task; 19 for training and 20 for testing. The texts were from 7 domains and 8 languages; English (en), Arabic (ar), Chinese (zh), Dutch (nl), French (fr), Russian (ru), Spanish (es) and Turkish (tr). The datasets for the domains of restaurants (rest), laptops (lap), mobile phones (phns), digital cameras (came), hotels (hote) and museums (muse) consist of customer reviews, whilst the telecommunication domain (tels) data consists of tweets. A total of 70790 manually annotated ABSA tuples were provided for training and testing; 47654 sentence-level annotations (SB1) in 8 languages for 7 domains, and 23136 text-level annotations (SB2) in 6 languages for 3 domains. Table 1 provides more information on the distribution of texts, sentences and annotated tuples per dataset.

The rest, hote, and lap datasets were annotated...
at the sentence-level (SB1) following the respective annotation schemas of SE-ABSA15 (Pontiki et al., 2015). Below are examples of annotated sentences for the aspect category “SERVICE#GENERAL” in EN (1), DU (2), FR (3), RU (4), ES (5), and TU (6) for the rest domain and in AR (7) for the NOTE domain:

1. Service was slow, but the people were friendly. → {trg: “Service”, pol: “negative”}, {trg: “people”, pol: “positive”}

2. Snelle bediening en vriendelijk personeel moet ook gemeld worden!! → {trg: “bediening”, pol: “positive”}, {trg: “personeel”, pol: “positive”}

3. Le service est impeccable, personnel agréable. → {trg: “service”, pol: “positive”}, {trg: “personnel”, pol: “positive”}

4. 2. Просервисничегонегативногонескажешь− быстро подходят, все улыбаются, подходят спрашивают, всё ли нравится. → {trg: “сервис”, pol: “neutral”}

5. También la rapidez en el servicio. → {trg: “servicio”, pol: “positive”}

6. Servisi hızlı valesi var. → {trg: “Servisi”, pol: “positive”}

7. ..الخدمة جيدة جدا و سريعة .. → {trg: “الخدمة”, pol: “positive”}

The LAPT annotation schema was extended to two other domains of consumer electronics, CAME and PHNS. Examples of annotated sentences in the LAPT (EN), PHNS (DU and CH) and CAME (CH) domains are shown below:

1. It is extremely portable and easily connects to WiFi at the library and elsewhere. → {cat: “LAPTOP#PORTABILITY”, pol: “positive”}, {cat: “LAPTOP#CONNECTIVITY”, pol: “positive”}

2. Apps starten snel op en werken vlot, internet gaat prima. → {cat: “SOFTWARE#OPERATION_PERFORMANCE”, pol: “positive”}, {cat: “PHONE#CONNECTIVITY”, pol: “positive”}

The offsets of the opinion target expressions are omitted.

The text-level (SB2) annotation task was based on the sentence-level annotations; given a customer review about a target entity (e.g., a restaurant) that included sentence-level annotations of ABSA tuples, the goal was to identify a set of {cat, pol} tuples that summarize the opinions expressed in it. This was not a simple summation/aggregation of the sentence-level annotations since an aspect may be discussed with different sentiment in different parts of the review. In such cases the dominant sentiment had to be identified. In case of conflicting opinions where the dominant sentiment was not clear, the “conflict” label was assigned. In addition, each review was assigned an overall sentiment label about the target entity (e.g., “RESTAURANT#GENERAL”, “LAPTOP#GENERAL”), even if it was not included in the sentence-level annotations.

3. 当然屏幕这么好 → {cat: “DISPLAY#QUALITY”, pol: “positive”}

4. 更轻便的机身也便于携带。 → {cat: “CAMERA#PORTABILITY”, pol: “positive”}

In addition, the SE-ABSA15 framework was extended to two new domains for which annotation guidelines were compiled: TELC for TU and MUSE for FR. Below are two examples:

1. #Internet kopuyor sürekli :( @turkcell → {cat: “INTERNET#COVERAGE”, trg: “Internet”, pol: “positive”}

2. 5€ pour les étudiants, ça vaut le coup. → {cat: “MUSEUM#PRICES”, “NULL”, “positive”}

The text-level (SB2) annotation task was based on the sentence-level annotations; given a customer review about a target entity (e.g., a restaurant) that included sentence-level annotations of ABSA tuples, the goal was to identify a set of {cat, pol} tuples that summarize the opinions expressed in it. This was not a simple summation/aggregation of the sentence-level annotations since an aspect may be discussed with different sentiment in different parts of the review. In such cases the dominant sentiment had to be identified. In case of conflicting opinions where the dominant sentiment was not clear, the “conflict” label was assigned. In addition, each review was assigned an overall sentiment label about the target entity (e.g., “RESTAURANT#GENERAL”, “LAPTOP#GENERAL”), even if it was not included in the sentence-level annotations.

3.2 Annotation Process

All datasets for each language were prepared by one or more research groups as shown in Table 2. The EN, DU, FR, RU and ES datasets were annotated using BRAT (Stenetorp et al., 2012), a web-based annotation tool, which was configured appropriately for the needs of the task. The TU datasets were annotated using a customized version of TURKSENT (Eryigit et al., 2013), a sentiment annotation tool for social media. For the AR and the CH data in-house tools\(^6\) were used.

\(^{6}\)The AR annotation tool was developed by the technical team of the Advanced Arabic Text Mining group at Jordan University of Science and Technology. The CH tool was developed by the Research Center for Social Computing and Information Retrieval at Harbin Institute of Technology.
Below are some further details about the annotation process for each language.

**English.** The SE-ABSA15 (Pontiki et al., 2015) training and test datasets (with some minor corrections) were merged and provided for training (rest and LAPT domains). New data was collected and annotated from scratch for testing. In a first phase, the rest test data was annotated by an experienced linguist (annotator A), and the LAPT data by 5 undergraduate computer science students. The resulting annotations for both domains were then inspected and corrected (if needed) by a second expert linguist, one of the task organizers (annotator B). Borderline cases were resolved collaboratively by annotators A and B.

**Arabic.** The norte dataset was annotated in repeated cycles. In a first phase, the data was annotated by three native Arabic speakers, all with a computer science background; then the output was validated by a senior researcher, one of the task organizers. If needed (e.g. when inconsistencies were found) they were given back to the annotators.

**Chinese.** The datasets presented by Zhao et al. (2015) were re-annotated by three native Chinese speakers according to the SE-ABSA16 annotation schema and were provided for training and testing (PHNS and CAME domains).

**Dutch.** The rest and phns datasets (De Clercq and Hoste, 2016) were initially annotated by a trained linguist, native speaker of Dutch. Then, the output was verified by another Dutch linguist and disagreements were resolved between them. Finally, the task organizers inspected collaboratively all the annotated data and corrections were made when needed.

**French.** The train (rest) and test (rest, muse) datasets were annotated from scratch by a linguist, native speaker of French. When the annotator was not confident, a decision was made collaboratively with the organizers. In a second phase, the task organizers checked all the annotations for mistakes and inconsistencies and corrected them, when necessary. For more information on the French datasets consult Apidianaki et al. (2016).

**Russian.** The rest datasets of the SentiRuEval-2015 task (Loukachevitch et al., 2015) were automatically converted to the SE-ABSA16 annotation schema; then a linguist, native speaker of Russian, checked them and added missing information. Finally, the datasets were inspected by a second linguist annotator (also native speaker of Russian) for mistakes and inconsistencies, which were resolved along with one of the task organizers.

**Spanish.** Initially, 50 texts (134 sentences) from the whole available data were annotated by 4 annotators. The inter-annotator agreement (IAA) in terms of F-1 was 91% for the identification of OTE, 88% for the aspect category detection (E#A pair), and 80% for opinion tuples extraction (E#A, OTE, polarity). Provided that the IAA was substantially high for all slots, the rest of the data was divided into 4 parts and each one was annotated by a different native Spanish speakers (2 linguists and 2 software engineers). Subsequently, the resulting annotations were validated and corrected (if needed) by the task organizers.
The TELC dataset was based on the data used in (Yıldırım et al., 2015), while the rest dataset was created from scratch. Both datasets were annotated simultaneously by two linguists. Then, one of the organizers validated/inspected the resulting annotations and corrected them when needed.

3.3 Datasets Format and Availability

Similarly to SE-ABSA14 and SE-ABSA15, the datasets of SE-ABSA16 were provided in an XML format and they are available under specific license terms through META-SHARE\(^9\), a repository devoted to the sharing and dissemination of language resources (Piperidis, 2012).

4 Evaluation Measures and Baselines

The evaluation ran in two phases. In the first phase (Phase A), the participants were asked to return separately the aspect categories (Slot1), the OTEs (Slot2), and the \{Slot1, Slot2\} tuples for SB1. For SB2 the respective text-level categories had to be identified. In the second phase (Phase B), the gold annotations for the test sets of Phase A were provided and participants had to return the respective sentiment polarity values (Slot3). Similarly to SE-ABSA15, F-1 scores were calculated for Slot1, Slot2 and \{Slot1, Slot2\} tuples, by comparing the annotations that a system returned to the gold annotations (using micro-averaging). For Slot1 evaluation, duplicate occurrences of categories were ignored in both SB1 and SB2. For Slot2, the calculation for each sentence considered only distinct targets and discarded “null” targets, since they do not correspond to explicit mentions. To evaluate sentiment polarity classification (Slot3) in Phase B, we calculated the accuracy of each system, defined as the number of correctly predicted polarity labels of the (gold) aspect categories, divided by the total number of the gold aspect categories. Furthermore, we implemented and provided baselines for all slots of SB1 and SB2. In particular, the SE-ABSA15 baselines that were implemented for the English language (Pontiki et al., 2015), were adapted for the other languages by using appropriate stopword lists and tokenization functions. The baselines are briefly discussed below:

SB1-Slot1: For category (E#A) extraction, a Support Vector Machine (SVM) with a linear kernel is trained. In particular, \(n\) unigram features are extracted from the respective sentence of each tuple that is encountered in the training data. The category value (e.g., “service\#general”) of the tuple is used as the correct label of the feature vector. Similarly, for each test sentence \(s\), a feature vector is built and the trained SVM is used to predict the probabilities of assigning each possible category to \(s\) (e.g., \{“service\#general”, 0.2\}, \{“restaurant\#general”, 0.4\}). Then, a threshold\(^{10}\(t\) is used to decide which of the categories will be assigned\(^{11}\) to \(s\). As features, we use the 1,000 most frequent unigrams of the training data excluding stopwords.

SB1-Slot2: The baseline uses the training reviews to create for each category \(c\) (e.g., “service\#general”) a list of OTEs (e.g., “service\#general” \(\rightarrow\) \{“staff”, “waiter”\}). These are extracted from the (training) opinion tuples whose category value is \(c\). Then, given a test sentence \(s\) and an assigned category \(c\), the baseline finds in \(s\) the first occurrence of each OTE of \(c\)’s list. The OTE slot is filled with the first of the target occurrences found in \(s\). If no target occurrences are found, the slot is assigned the value “null”.

SB1-Slot3: For polarity prediction we trained a SVM classifier with a linear kernel. Again, as in Slot1, \(n\) unigram features are extracted from the respective sentence of each tuple of the training data. In addition, an integer-valued feature\(^{12}\) that indicates the category of the tuple is used. The correct label for the extracted training feature vector is the corresponding polarity value (e.g., “positive”). Then, for each tuple \{category, OTE\} of a test sentence \(s\), a feature vector is built and classified using the trained SVM.

SB2-Slot1: The sentence-level tuples returned by the SB1 baseline are copied to the text level and duplicates are removed.

\(^9\)The data are available at: http://metashare.ilsper.gr:8080/repository/search/?q=semeval+2016

\(^{10}\)The threshold \(t\) was set to 0.2 for all datasets.

\(^{11}\)We use the –b 1 option of LibSVM to obtain probabilities.

\(^{12}\)Each E#A pair has been assigned a distinct integer value.
Table 3: English REST results for SB1.

| Slot1 | Slot2 | Slot3 |
|-------|-------|-------|
| Dom.  | F-1   | F-1   | F-1   |
| EN/   | NLANG./U/73.031 | NLANG./U/72.34 | NLANG./U/52.607 |
| REST  | NileT./U/72.886  | AUEB.-/U/70.441 | XREC/C/48.891  |
|       | BUTkn./U/72.396 | UWB/C/66.096   | NileC./U/88.126 |
|       | AUEB.-/U/71.537 | GTI/U/66.553   | IHS-R./U/83.935 |
|       | BUTkn./C/71.494 | Senti./C/66.545 | ECNU/C/83.586  |
|       | SYSU/U/70.869   | bunji/U/64.882  | AUEB.-/U/83.236 |
|       | XREC/C/68.701   | NLANG./C/63.861 | XREC/C/37.795  |
|       | UWB/U/68.203    | DMIS/C/63.495  | IHS-R./U/55.608 |
|       | INSIG./U/68.108 | XRCE/C/61.98   | UWB/C/34.846   |
|       | ES/7U/67.979    | AUEB.-/C/61.552 | TGB/C/34.163   |
|       | UWB/C/67.817    | UWate./U/57.067 | DMIS/C/39.796  |
|       | GTI/U/67.714    | KnowC./U/56.816* | DMIC/C/38.976  |
|       | AUEB.-/C/67.35  | TGB/C/55.054*  | baseline./C/37.795 |
|       | NLANG./C/65.563 | BUAP/U/50.253  | IHS-R./U/34.846 |
|       | LeeHu./C/65.455 | baseline./C/44.071 | SeemGo/U/30.667 |
|       | TGB/C/63.919*   | IHS-R./U/43.808 | BUAP/U/18.428  |
|       | IIT-T./U/63.051 | IIT-T./U/42.603 | SeemGo/U/18.428 |
|       | DMIS/U/62.583   | bunji/U/60.145 | IHS-R./U/34.846 |
|       | DMIC/C/61.754   | baseline./C/59.928 | TGB/C/34.163   |
|       | IIT-T./U/61.227 | UFAL/U/59.3    | DMIS/C/79.977  |
|       | bunji/U/60.145  | INSIG./C/58.303 | DMIS/C/79.627  |
|       | baseline./C/59.928 | IHS-R./U/55.034 | IHS-R./U/78.696 |
|       | UFAL/U/59.3     | IHS-R./U/53.149 | Senti./U/78.114 |
|       | INSIG./C/58.203 | SeemGo/U/50.737 | LeeHu./C/78.114 |
|       | IHS-R./U/55.034 | UWate./U/49.73 | baseline./C/76.484 |
|       | IHS-R./U/53.149 | CENNL./C/40.578 | bunji/U/76.251 |
|       | SeemGo/U/50.737 | BUAP/U/37.29    | SeemGo/U/72.992 |
|       | UWate./U/49.73  | CENNL./C/40.578 | AKTSKI/U/71.111 |

SB2-Slot3: For each text-level aspect category \(c\), the baseline traverses the predicted sentence-level tuples of the same category returned by the respective SB1 baseline and counts the polarity labels (positive, negative, neutral). Finally, the polarity label with the highest frequency is assigned to the text-level category \(c\). If there are no sentence-level tuples for the same \(c\), the polarity label is determined based on all tuples regardless of \(c\).

The baseline systems and evaluation scripts are implemented in Java and are available for download from the SE-ABSA16 website\(^{13}\). The LibSVM package\(^{14}\) (Chang and Lin, 2011) is used for SVM training and prediction. The scores of the baselines in the test datasets are presented in Section 6 along with the system scores.

5 Participation

The task attracted in total 245 submissions from 29 teams. The majority of the submissions (216 runs) were for SB1. The newly introduced SB2 attracted 29 submissions from 5 teams in 2 languages (en and sp). Most of the submissions (168) were runs for the rest domain. This was expected, mainly for two reasons; first, the rest classification schema is less fine-grained (complex) compared to the other domains (e.g., lapt). Secondly, this domain was supported for 6 languages enabling also multilingual or language-agnostic approaches. The remaining submissions were distributed as follows: 54 in lapt, 12 in phns, 7 in came and 4 in hote.

---

\(^{13}\)http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

\(^{14}\)http://www.csie.ntu.edu.tw/~cjlin/libsvm/
| Lang./ Dom. | Slot1 | Slot2 | {Slot1,Slot2} | Slot3 | Acc. |
|------------|-------|-------|---------------|-------|------|
| ES/REST    | GTI/U/70.588 | GTI/C/68.515* | TGB/C/41.219* | IIT-T./U/83.582 | GTI/C/36.379 |
|            | GTI/C/70.027 | GTI/U/68.387 | TGB/C/55.764* | TGB/C/82.09* | TGB/C/77.799 |
|            | TGB/C/63.551* | IIT-T./U/64.338 | TGB/C/51.914 | INSIG./C/79.571 | basel/C/77.799 |
|            | UWB/C/61.968 | TGB/C/41.219* | TGB/C/51.914 | basel/C/77.799 | basel/C/77.799 |
|            | INSIG./C/61.37 | IIT-T./U/59.899 | TGB/C/51.914 | basel/C/77.799 | basel/C/77.799 |
|            | IIT-T./U/59.062 | TGB/C/51.914 | TGB/C/51.914 | basel/C/77.799 | basel/C/77.799 |
|            | UFAL/U/58.81 | TGB/C/51.914 | TGB/C/51.914 | basel/C/77.799 | basel/C/77.799 |
|            | basel/C/54.686 | TGB/C/51.914 | TGB/C/51.914 | basel/C/77.799 | basel/C/77.799 |
| FR/REST    | XRCE/C/61.207 | XRCE/C/65.316 | XRCE/C/45.455 | XRCE/C/78.26 | UWB/C/75.262 |
|            | IIT-T./U/57.875 | IIT-T./U/56.667 | basel/C/45.455 | UWB/C/75.262 | basel/C/67.4 |
|            | IIT-T./C/57.033 | IIT-T./U/66.667 | basel/C/45.455 | UWB/C/75.262 | basel/C/67.4 |
|            | INSIG./C/53.592 | XRCE/C/65.316 | basel/C/45.455 | UWB/C/75.262 | basel/C/67.4 |
|            | basel/C/52.609 | XRCE/C/65.316 | basel/C/45.455 | UWB/C/75.262 | basel/C/67.4 |
|            | UFAL/U/49.928 | XRCE/C/65.316 | basel/C/45.455 | UWB/C/75.262 | basel/C/67.4 |
| RU/REST    | UFAL/U/64.825 | basel/C/49.308 | basel/C/39.441 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | INSIG./C/62.802 | basel/C/39.441 | basel/C/39.441 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | IIT-T./U/62.689 | Daniii./U/33.472 | Daniii./U/33.472 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | IIT-T./C/58.196 | Daniii./U/30.618 | Daniii./U/30.618 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | basel/C/55.882 | Daniii./U/30.618 | Daniii./U/30.618 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | Daniii./C/39.601 | Daniii./U/30.618 | Daniii./U/30.618 | MayAnd/U/77.923 | INSIG./C/73.166 |
|            | Daniii./C/38.692 | Daniii./C/30.618 | Daniii./C/30.618 | MayAnd/U/77.923 | INSIG./C/73.166 |
| DU/REST    | TGB/C/60.153* | TGB/C/56.986 | TGB/C/45.167* | TGB/C/77.814* | INSIG./C/75.041 |
|            | INSIG./C/56 | TGB/C/51.775* | TGB/C/50.64 | TGB/C/77.814* | INSIG./C/75.041 |
|            | IIT-T./U/55.247 | TGB/C/51.775* | TGB/C/50.64 | TGB/C/77.814* | INSIG./C/75.041 |
|            | IIT-T./C/54.98 | TGB/C/51.775* | TGB/C/50.64 | TGB/C/77.814* | INSIG./C/75.041 |
|            | UFAL/U/53.876 | TGB/C/51.775* | TGB/C/50.64 | TGB/C/77.814* | INSIG./C/75.041 |
|            | basel/C/42.816 | TGB/C/51.775* | TGB/C/50.64 | TGB/C/77.814* | INSIG./C/75.041 |
| TU/REST    | UFAL/U/61.029 | basel/C/28.152 | basel/C/18.806 | IIT-T./U/84.277 | INSIG./C/75.041 |
|            | basel/C/58.896 | basel/C/28.152 | basel/C/18.806 | IIT-T./U/84.277 | INSIG./C/75.041 |
|            | IIT-T./U/56.627 | basel/C/28.152 | basel/C/18.806 | IIT-T./U/84.277 | INSIG./C/75.041 |
|            | IIT-T./C/55.728 | basel/C/28.152 | basel/C/18.806 | IIT-T./U/84.277 | INSIG./C/75.041 |
|            | INSIG./C/49.123 | basel/C/28.152 | basel/C/18.806 | IIT-T./U/84.277 | INSIG./C/75.041 |
| AR/HOTE    | INSIG./C/52.114 | basel/C/30.978 | basel/C/30.978 | INSIG./C/82.719 | basel/C/76.421 |
|            | UFAL/U/47.302 | basel/C/30.978 | basel/C/30.978 | INSIG./C/82.719 | basel/C/76.421 |
|            | basel/C/40.336 | basel/C/30.978 | basel/C/30.978 | INSIG./C/82.719 | basel/C/76.421 |

Table 4: REST and HOTE results for SB1.

An interesting observation is that, unlike SE-ABSA15, Slot1 (aspect category detection) attracted significantly more submissions than Slot2 (OTE extraction); this may indicate a shift towards concept-level approaches. Regarding participation per language, the majority of the submissions (156/245) were for en; see more information in Table 5. Most teams (20) submitted results only for one language (18 for en and 2 for ru). Of the remaining teams, 3 submitted results for 2 languages, 5 teams submitted results for 3-7 languages, while only one team participated in all languages.

6 Evaluation Results

The evaluation results are presented in Tables 3 (SB1: rest-en), 4 (SB1: rest-es, fr, ru, du, tu & hote-ar), 5 (SB1: lap, cama, phns), and 7 (SB2)\(^{15}\). Each participating team was allowed to submit up to two runs per slot and domain in each phase; one constrained (C), where only the provided training data could be used, and one unconstrained (U), where other resources (e.g., publicly available

\(^{15}\)No submissions were made for sb3-muse-fr & sb1-tecl-tu.
English 27 156
Arabic 3 4
Chinese 3 14
Dutch 4 16
French 5 13
Russian 5 15
Spanish 6 21
Turkish 3 6

| Language | Teams | Submissions |
|----------|-------|-------------|
| English  | 27    | 156         |
| Arabic   | 3     | 4           |
| Chinese  | 3     | 14          |
| Dutch    | 4     | 16          |
| French   | 5     | 13          |
| Russian  | 5     | 15          |
| Spanish  | 6     | 21          |
| Turkish  | 3     | 6           |

| Language | Teams | Submissions |
|----------|-------|-------------|
| EN/      |       |             |
| LAPT     |       |             |
| NLANG./U | 51.937|             |
| AUEB.-/U | 49.105|             |
| SYSU/U   | 49.076|             |
| BUTk./U  | 48.396|             |
| UWB/C    | 47.891|             |
| BUTk./C  | 47.527|             |
| UWB/U    | 47.258|             |
| NileT./U | 47.196|             |
| NLANG./C | 46.728|             |
| INSIG./U | 45.863|             |
| AUEB./C  | 45.629|             |
| IIT-T./U | 43.913|             |
| LeeHu./C | 43.754|             |
| IIT-T./C | 42.609|             |
| SeemGo/U | 41.499|             |
| INSIG./C | 41.458|             |
| bunji/U  | 39.586|             |
| IHS-R-U  | 39.024|             |
| basel./C | 37.481|             |
| UFWAL/U  | 26.984|             |
| CENNL./C | 26.908|             |
| BUAP/U   | 26.787|             |
| CH/      |       |             |
| CAME     |       |             |
| UWB/C    | 36.345|             |
| INSIG./C | 25.581|             |
| basel./C | 18.434|             |
| SeemGo/U | 17.757|             |
| CH/      |       |             |
| PHNS     |       |             |
| UWB/C    | 22.548|             |
| basel./C | 17.03 |             |
| INSIG./C | 16.286|             |
| SeemGo/U | 10.43 |             |
| DU/      |       |             |
| PHNS     |       |             |
| INSIG./C | 45.551|             |
| IIT-T./U | 45.443|             |
| IIT-T./C | 45.047|             |
| basel./C | 33.55 |             |

Table 5: Number of participating teams and submitted runs per language.

As revealed by the results, in both SB1 and SB2 the majority of the systems surpassed the baseline by a small or large margin and, as expected, the unconstrained systems achieved better results than the constrained ones. In SB1, the teams with the highest scores for Slot1 and Slot2 achieved similar F-1 scores (see Table 3) in most cases (e.g., EN/REST, ES/REST, DU/REST, FR/REST), which shows that the two slots have a similar level of difficulty. However, as expected, the {Slot1, Slot2} scores were significantly lower since the linking of the target expressions to the corresponding aspects is also required. The highest scores in SB1 for all slots (Slot1, Slot2, {Slot1, Slot2}, Slot3) were achieved in the EN/REST; this is probably due to the high participation and to the lower complexity of the REST annotation schema compared to the other domains. If we compare the results for SB1 and SB2, we notice that the SB2 scores for Slot1 are significantly higher (e.g., EN/REST, ES/REST, FR/REST) even though the respective annotations are for the same (or almost the same) set of texts. This is due to the fact that it is easier to identify whether a whole text discusses an aspect c than finding all the sentences in the text discussing c. On the other hand, for Slot3, the SB2 scores are lower (e.g., EN/REST, ES/REST, RU/REST, EN/LAPT) than the respective SB1 scores. This is mainly because an aspect may be discussed at different points in a text and often with different sentiment. In such cases a system has to identify the dominant sentiment, which usually is not trivial.

7 Conclusions

In its third year, the SemEval ABSA task provided 19 training and 20 testing datasets, from 7 domains and 8 languages, attracting 245 submissions from 29 teams. The use of the same annotation guidelines for domains addressed in different languages gives the opportunity to experiment also with cross-lingual or language-agnostic approaches. In addition, SE-ABSA16 included for the first time a text-
Table 7: Results for SB2.

| Lang./Dom. | Slot1 | Slot3 |
|------------|-------|-------|
| EN/REST    | GTI/U/83.995 | UWB/B/81.931 |
|            | UWB/U/80.965 | ECNU/U/81.436 |
|            | UWB/U/80.163 | UWB/C/80.941 |
|            | bunji/U/79.777 | ECNU/C/78.713 |
|            | basel/C/78.711 | basel/C/74.257 |
|            | SYSU/U/68.841 | bunji/U/70.545 |
|            | SYSU/U/68.841 | bunji/C/66.584 |
|            |               | GTI/U/64.109 |
| ES/REST    | GTI/C/84.192 | UWB/C/77.185 |
|            | GTI/U/84.044 | basel/C/74.548 |
|            | basel/C/74.548 |               |
| RU/REST    | basel/C/84.792 | basel/C/70.6 |
|            | basel/C/84.792 | basel/C/70.6 |
| DU/REST    | basel/C/70.323 | basel/C/73.228 |
| TU/REST    | basel/C/72.642 | basel/C/57.407 |
| AR/HOTE    | basel/C/42.757 | basel/C/73.216 |
| EN/LAPT    | UWB/C/60.45 | ECNU/U/75.046 |
|            | UWB/U/59.721 | UWB/U/75.046 |
|            | bunji/U/54.723 | UWB/C/74.495 |
|            | basel/C/52.685 | basel/C/73.028 |
|            | SYSU/U/48.889 | ECNU/C/67.523 |
|            | SYSU/U/48.889 | bunji/C/62.202 |
|            |               | bunji/U/60 |
|            |               | GTI/U/58.349 |
|            |               |               |

Acknowledgments

The authors are grateful to all the annotators and contributors for their valuable support to the task: Konstantina Papanikolau, Juli Bakagianni, Omar Qwasmeh, Nesreen Alqasem, Areen Magableh, Saja Alzoubi, Bashar Talalha, Zekui Li, Binbin Li, Shengqiu Li, Aaron Gevaert, Els Lefever, Cécile Richard, Pavel Blinov, Maria Shatalova, M. Teresa Martín-Valdivia, Pilar Santolaria, Fatih Samet Çetin, Ezgi Yıldırım, Can Özbeay, Leonidas Valavanis, Stavros Giorgis, Dionysios Xenos, Panos Theodor-akos, and Apostolos Rousas. The work described in this paper is partially funded by the projects EOX GR07/3712 and “Research Programs for Excellence 2014-2016 / CitySense-ATHENA R.I.C.”. The Arabic track was partially supported by the Jordan University of Science and Technology, Research Grant Number: 20150164. The Dutch track has been partly funded by the PARIS project (IWT-SBO-Nr. 110067). The French track was partially supported by the French National Research Agency under project ANR-12-CORD-0015/TransRead. The Russian track was partially supported by the Russian Foundation for Basic Research (RFBR) according to the research projects No. 14-07-00682a, 16-07-00342a, and No. 16-37-00311mol_a. The Spanish track has been partially supported by a grant from the Ministerio de Educación, Cultura y Deporte (MECD - scholarship FPU014/00983) and REDES project (TIN2015-65136-C2-1-R) from the Ministerio de Economía y Competitividad. The Turkish track was partially supported by TUBITAK-TEYDEB (The Scientific and Technological Research Council of Turkey – Technology and Innovation Funding Programs Directorate) project (grant number: 3140671).

References

Marianna Apidianaki, Xavier Tannier, and Cécile Richard. 2016. A Dataset for Aspect-Based Sentiment Analysis in French. In Proceedings of the International Conference on Language Resources and Evaluation.

Erik Cambria, Björn W. Schuller, Yunqing Xia, and Catherine Havasi. 2013. New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2):15–21.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. ACM TIST, 2(3):27.

Judith Chevalier and Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. J. Marketing Res, pages 345–354.

Orphée De Clercq and Véronique Hoste. 2016. Rude waiter but mouthwatering pastries! An exploratory study into Dutch Aspect-Based Sentiment Analysis. In Proceedings of the 10th International Conference on Language Resources and Evaluation.

Gülşen Eryigit, Fatih Samet Cetin, Meltem Yank, Turkcell Global Bilgi, Tanel Temel, and Ilyas Çiçekli.
Gayatree Ganu, Noemie Elhadad, and Amélie Marian. 2009. Beyond the Stars: Improving Rating Predictions using Review Text Content. In Proceedings of WebDB.

Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes. 2015. SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. In Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, Colorado.

Salud M. Jiménez-Zafra, Giacomo Berardi, Andrea Esuli, Diego Marcheggiani, María Teresa Martín-Valdivia, and Alejandro Moreo Fernández. 2015. A Multi-lingual Annotated Dataset for Aspect-Oriented Opinion Mining. In Proceedings of Empirical Methods in Natural Language Processing, pages 2533–2538.

Roman Klinger and Philipp Cimiano. 2014. The USAGE Review Corpus for Fine Grained Multi Lingual Opinion Analysis. In Proceedings of the Ninth International Conference on Language Resources and Evaluation, Reykjavik, Iceland.

Patrik Lambert. 2015. Aspect-Level Cross-lingual Sentiment Classification with Constrained SMT. In Proceedings of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, 2015, Beijing, China, pages 781–787.

Bing Liu. 2012. Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Natalia Loukachevitch, Pavel Blinov, Evgeny Kotelnikov, Yulia Rubtsova, Vladimir Ivanov, and Elena Tutubalina. 2015. SentiRuEval: Testing Object-oriented Sentiment Analysis Systems in Russian. In Proceedings of International Conference Dialog.

Julian J. McAuley, Jure Leskovec, and Dan Jurafsky. 2012. Learning Attitudes and Attributes from Multi-aspect Reviews. In 12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012, pages 1020–1025.

Margaret Mitchell. 2013. Overview of the TAC2013 Knowledge Base Population Evaluation English Sentiment Slot Filling. In Proceedings of the 6th Text Analysis Conference, Gaithersburg, Maryland, USA.

Saif M Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In Proceedings of the 10th International Workshop on Semantic Evaluation, San Diego, California.

Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. SemEval-2013 Task 2: Sentiment Analysis in Twitter. In Proceedings of the 7th International Workshop on Semantic Evaluation, Atlanta, Georgia.

Stelios Piperidis. 2012. The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions. In Proceedings of the 8th International Conference on Language Resources and Evaluation.

Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Ireland.

Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, Colorado.

Diego Reforgiato Recupero and Erik Cambria. 2014. Eswc’14 challenge on concept-level sentiment analysis. In Semantic Web Evaluation Challenge - SemWebEval 2014 at ESWC 2014, Anissaras, Crete, Greece, pages 3–20.

Sara Rosenthal, Alan Ritter, Preslav Nakov, and Veselin Stoyanov. 2014. SemEval-2014 Task 4: Sentiment Analysis in Twitter. In Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Ireland.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. SemEval-2015 Task 10: Sentiment Analysis in Twitter. In Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, Colorado.

Josef Ruppenhofer, Roman Klinger, Julia Maria Strüß, Jonathan Sonntag, and Michael Wiegand. 2014. IGGSA Shared Tasks on German Sentiment Analysis (GESTALT). In Workshop Proceedings of the 12th Edition of the KONVENS Conference, pages 164–173.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-Hsi Chen, Noriko Kando, and Chin-Yew Lin. 2007. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proceedings of the 6th NTCIR Workshop, Tokyo, Japan.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and Noriko Kando. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Proceedings of the 7th NTCIR Workshop, Tokyo, Japan.

Yohei Seki, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, and Noriko Kando. 2010. Overview of Multilingual Opinion Analysis Task at NTCIR-8: A Step Toward Cross
Lingual Opinion Analysis. In *Proceedings of the 8th NTCIR Workshop, Tokyo, Japan*, pages 209–220.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of Empirical Methods in Natural Language Processing*, pages 1631–1642, Stroudsburg, PA.

Pontus Stenetorp, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. BRAT: A Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

Ezgi Yıldırım, Fatih Samet Çetin, Gülşen Eryiğit, and Tanel Temel. 2015. The impact of nlp on turkish sentiment analysis. *TÜRKİYE BİLİŞİM VAKFI BİLGİ-SAYA-RI ve MÜHENDİSLİĞİ DERGİSİ*, 7(1 (Basılı 8).

Kyung Hyan Yoo and Ulrike Gretzel. 2008. What Motivates Consumers to Write Online Travel Reviews? *J. of IT & Tourism*, 10(4):283–295.

Yanyan Zhao, Bing Qin, and Ting Liu. 2015. Creating a Fine-Grained Corpus for Chinese Sentiment Analysis. *IEEE Intelligent Systems*, 30(1):36–43.

### Appendix A. Aspect inventories for all domains

#### Table 8: Laptops.

**Entity Labels**

- CAMERA, DISPLAY, KEYBOARD, CPU, PORTS, MEMORY, POWER_SUPPLY, BATTERY, MULTIMEDIA DEVICES, HARDWARE, SOFTWARE, OS, WARRANTY, SHIPPING, SUPPORT, COMPANY

**Attribute Labels**

- GENERAL, PRICE, QUALITY, DESIGN FEATURES, OPERATION PERFORMANCE, USABILITY, PORTABILITY, CONNECTIVITY, MISCELLANEOUS

#### Table 9: Mobile Phones.

**Entity Labels**

- PHONE, DISPLAY, KEYBOARD, CPU, PORTS, MEMORY, POWER_SUPPLY, HARD_DISK, MULTIMEDIA DEVICES, BATTERY, HARDWARE, SOFTWARE, OS, WARRANTY, SHIPPING, SUPPORT, COMPANY

**Attribute Labels**

- Same as in Laptops (Table 8) with the exception of PORTABILITY that is included in the DESIGN FEATURES label and does not apply as a separate attribute type.

#### Table 10: Digital Cameras.

**Entity Labels**

- CAMERA, DISPLAY, KEYBOARD, MOUSE, MOTHERBOARD, CPU, FANS COOLING, PORTS, MEMORY, POWER_SUPPLY OPTICAL DRIVES, BATTERY, GRAPHICS, HARD_DISK, MULTIMEDIA DEVICES, HARDWARE, SOFTWARE, OS, WARRANTY, SHIPPING, SUPPORT, COMPANY

**Attribute Labels**

- GENERAL, PRICE, QUALITY, DESIGN FEATURES, OPERATION PERFORMANCE, USABILITY, PORTABILITY, CONNECTIVITY, MISCELLANEOUS

#### Table 11: Restaurants.

**Entity Labels**

- RESTAURANT, FOOD, DRINKS, AMBIENCE, SERVICE, LOCATION

**Attribute Labels**

- GENERAL, PRICES, QUALITY, STYLE OPTIONS, DESIGN FEATURES, MISCELLANEOUS

#### Table 12: Hotels.

**Entity Labels**

- HOTEL, ROOMS, FACILITIES, ROOM AMENITIES, SERVICE, LOCATION, FOOD DRINKS

**Attribute Labels**

- GENERAL, PRICE, COMFORT, CLEANLINESS, QUALITY, STYLE OPTIONS, DESIGN FEATURES, MISCELLANEOUS

#### Table 13: Telecommunications.

**Entity Labels**

- TELECOM OPERATOR, DEVICE, INTERNET, CUSTOMER SERVICES, APPLICATION SERVICE

**Attribute Labels**

- GENERAL, PRICE, INVOICE, COVERAGE, SPEED, CAMPAIGN ADVERTISEMENT, MISCELLANEOUS

#### Table 14: Museums.

**Entity Labels**

- MUSEUM, COLLECTIONS, FACILITIES, SERVICE, TOUR GUIDING, LOCATION

**Attribute Labels**

- GENERAL, PRICES, COMFORT, ACTIVITIES, ARCHITECTURE, INTEREST, SETUP, MISCELLANEOUS