Identifying User Issues and Request Types in Forum Question Posts Based on Discourse Analysis

Agnes Sandor
Xerox Research Centre Europe
6 chemin de Maupertuis
Meylan, France
agnes.sandor@xerox.com

Nikolaos Lagos
Xerox Research Centre Europe
6 chemin de Maupertuis
Meylan, France
nikolaos.lagos@xerox.com

Ngoc-Phuoc-An Vo
Xerox Research Centre Europe
6 chemin de Maupertuis
Meylan, France
an.vo@xrce.xerox.com

Caroline Brun
Xerox Research Centre Europe
6 chemin de Maupertuis, Meylan, France
caroline.brun@xerox.com

ABSTRACT

In this paper we propose the detection of user issues and request types in technical forum question posts with a twofold purpose: supporting up-to-date knowledge generation in organizations that provide (semi-) automated customer-care services, and enriching forum metadata in order to enhance the effectiveness of search. We present a categorization system for detecting the proposed question post types based on discourse analysis, and show the advantage of using discourse patterns compared to a baseline relying on standard linguistic features. Besides the detailed description of our method, we also release our annotated corpus to the community.

Keywords

on-line forum; discourse analysis; question post analysis.

1. INTRODUCTION

The wealth of the corresponding knowledge creates new opportunities and challenges for organizations seeking to automate parts of user support and customer care services. Forum users introduce new problems and solutions, related for instance to new devices, and they describe first hand user experiences with rich information in terms of which solutions are better than others and why. Being able to actually transform such noisy and, frequently, unstructured data in a form that is useful for the enterprise – for example in mining frequently discussed problems, identifying trends, enriching a corresponding knowledge base – is a great challenge.

As a first step towards tackling some of the above tasks we propose identifying and extracting the most important types of user issues and requests, and detect the sentences that convey them. Especially in a customer care setting, this can typically help in designing and developing the knowledge base that will be used for responding to user requests. Even more importantly, in cases where a (semi-)automated system is used, the workflow that will be followed will be different according to the type of question: e.g. if the question is related to troubleshooting then a dialog-based system will be provided that will try to identify a root cause and will use a corresponding knowledge base, while if the question is related to a property of the device (e.g. what is the resolution of the screen) an answer will be given based on a match (i.e. closer to question answering) to the knowledge base holding device specifications.

We propose the detection of two main types of sentences in question posts: **anomaly descriptions** and several types of **information request**. The detection of anomaly descriptions allows directing the workflow towards troubleshooting: the detection of a root cause, which leads to the choice of a solution from a solution database. The detection of various types of information request sentences helps finding dedicated knowledge bases that provide the answers: **"how to" questions** require instructions, enquiries about the **properties** of devices need device specification resources, and finally enquiries for **explanations** may be answered based on specific explanatory documents[1]. One sentence may include expressions denoting different information requests and thus can have multiple categories associated to it.

Besides organisations, the users of technical web forums could also benefit from the classification of question posts. They often search for answers or solutions to their problems among the existing posts, and the usual search method involves using keywords or forum tags that represent a normalised form of query concepts. While keywords and tags characterize the posts through the set of notions that they include, they do not indicate the question post types: Does the post describe an anomaly with the device? Does it enquire for a method to execute an operation, for some property of the device or for some explanation? Thus posts that share the same subject matter, i.e. that are characterised through the same set of keywords or tags, may differ in the motivation of the authors of the question posts.

Consider the question posts shown in Figure 1 that are automatically linked as "Related" in the AskDifferent forum in the StackOverflow site based on the (s)4(den)-ons they include.

Both question posts share the tag *Adobe Bridge* and mention *uninstallment*. However in the first post the author experiences anomalous behaviour ("*For some reason Adobe Bridge is opening on login, even though my Login items show that it is unchecked*") and asks a question concerning ways for correcting the anomaly ("*Without uninstalling Adobe Bridge how can I stop it opening on*

WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890568

[1] The relevance of these categories has been verified by our business partners in the context of a corresponding research project.

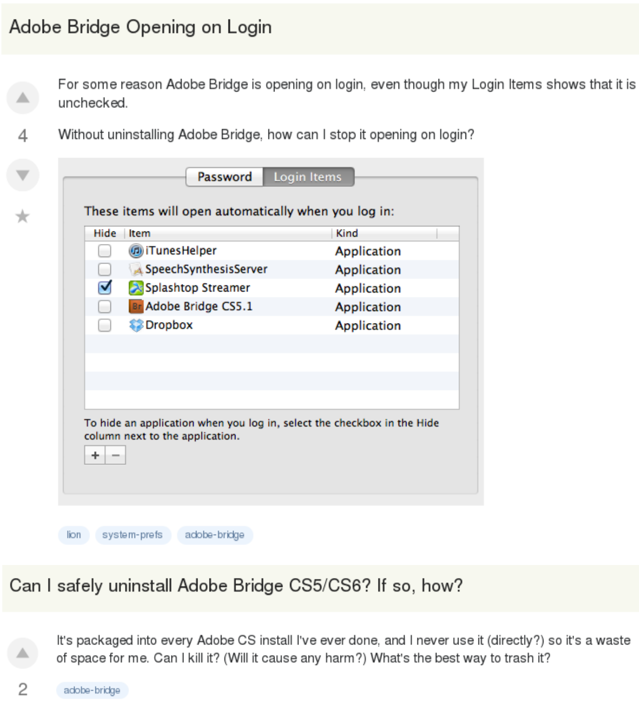

Figure 1. Question posts that share similar keywords and tags but refer to different intents

login?"), while the second post does not report any anomaly concerning the operation of the device; rather, the user would like to know if it is possible to carry out an operation, namely *uninstall Adobe Bridge*, and if so, if it has any consequences and how it should be done ("*Can I safely uninstall Adobe Bridge CS5/CS6? If so how? … Can I kill it? (Will it cause any harm?) What's the best way to trash it?*"). The users could access the best solution more rapidly if they could filter the posts according to this criterion.

The importance of question classification has already been recognized in the state-of-the-art (e.g. [2], [6]), however none of the existing methods dealt with the current setting and consequently with the corresponding classes we identify here, especially the one related to user issues. In addition, the use of discourse related, linguistically motivated features has not been used before in that context. Furthermore, in the paper we show that the advantage of our approach, when compared to supervised classification approaches used in related tasks in the state-of-art, is that our method does not require a large number of annotated examples and is (partially) domain-independent (although not genre-independent) thanks to the discourse-based features.

2. DATA

As a publicly available dataset, which would be appropriate for our purposes, does not exist, we developed an annotated corpus for our experiments.

To get sufficient data we used the December 2014 XML dump of the Ask Different website[2]. From the data dump we retrieved randomly 1000 question posts and also the tags assigned by community members to each post. Out of the 1000 posts, 150 posts were randomly selected and each of them was annotated independently by three annotators with the categories that interest us (the annotation guidelines provided to the annotators can be found at http://download.xrce.xerox.com/q4aps/guidelines.zip). The corresponding inter-annotator agreement is shown in Table 1. It is worth noting that the inter-annotator agreement is consistently lower for the category "Anomaly" compared to the rest of the categories. As it will be illustrated in Section 4.2, this is due to the complexity of the concept.

After this step was completed, and in order to develop the gold standard corpus, the three annotators were invited to find a final agreement. The above task resulted in 1150 annotations. Table 2 illustrates an example sentence for each annotation category.

Table 1. Inter-annotator agreement per category (Fleiss kappa)[3]

Category	annotators 1 + 2	annotators 1+ 3	annotators 2 + 3
ANOMALY (A)	0.635	0.652	0.622
EXPLANATION (E)	0	0.846	0
HOWTO (H)	0.796	0.768	0.791
PROPERTY (P)	0.796	0.882	0.82
OTHER (O)	0.784	0.829	0.793
Average	0.753	0.795	0.757

Table 2. Examples of sentences for each annotation category

Category	Example
ANOMALY	For some reason, Image Capture doesn't always update the list of devices after I plug in my iPhone.
HOWTO	How to automatically login to captive portals on OS X?
PROPERTY	What are the differences, feature-wise, between Safari for Mac and Safari for Windows?
EXPLANATION	How does "Find my Mac" work?

We randomly chose 100 posts as our training corpus and the remaining 50 posts as our test corpus. The distribution of the gold-standard annotations by category is shown in Figure 2. We can see that the distribution of the annotations is similar in the two corpora.

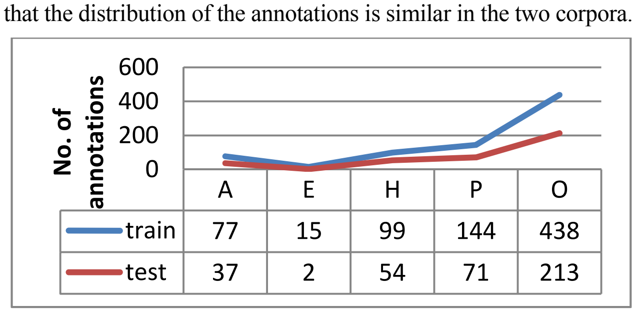

Figure 2. Distribution of annotations in the training and test corpora

To allow further research in this subject we release the annotated corpus, which can be accessed at http://download.xrce.xerox.com/q4aps/guidelines.zip.

[2]https://archive.org/details/stackexchange

[3] For the category "Explanation" there are two 0 values, because annotator 3 did not take into consideration this category. In the averages, the 0 values are not taken into account.

3. DISCOURSE ANALYSIS

We used the manually annotated sentences in the development corpus for identifying discourse patterns that signal the sentence types, and for implementing them as rules using the Xerox Incremental Parser (XIP) [1]. XIP is a convenient tool for this task since we could build on its general language analysis functions - it provides rich lexical, morpho-syntactic and dependency information-, as well as on its rule formalism that allows using and enriching linguistic features.

The complexity of the discourse patterns varies according to the sentence types: whereas the information request sentences, i.e. HOWTO, PROPERTY and EXPLANATION, are signaled by surface linguistic patterns, i.e. by patterns that contain a relatively small set of lexical elements and syntactic structures, the sentences conveying ANOMALY are much more heterogeneous, and needed a deeper analysis.

3.1 Information Request Patterns

The following simple surface patterns (underlined in the examples) characterize information request.

HOWTO = Direct or indirect question containing

- *how* + 1st person subject
  - *(1) How can I format Time Capsule?*
  - *(2) I use Mac OS X (10.7.5) I wonder how I can turn off screen only of my MacBook without closing the lid, without waiting for the screensaver.*
- *how to / way*
  - *(3) How to automatically login to captive portals on OS X?*
  - *(4) Also I am curious how to do thumbnails for the post*
  - *(5) Is there any way to force-enable it?*
- the lemma *do* having the word *what* as its direct object
  - *(6) What do I have to do to my tracks to get the iTunes Album Art?*

PROPERTY = Direct or indirect

- yes-no question
  - *(7) Are the devices returned in factory condition?*
  - *(8) We would like to know if Apple keeps track of iOS App opening by users.*
- Wh-question except for HOWTO
  - *(9) What could happen if I upgrade an iPod to iOS 5 and what is the likelihood it will happen)?*
  - *(10) How does Find my Mac work, since there's no GPS in a Mac?*

EXPLANATION = Direct or indirect question containing

- *why*
  - *(11) Why don't desktops stay in order?*
- Direct or indirect question where *cause* has a subject or an object
  - *(12) What can cause the phone to not automatically connect to a known access point?*

3.2 Anomaly Description Patterns

In the annotation guidelines we defined ANOMALY as follows: "An ANOMALY is a **deviation** from normal (correct, usual, expected, good, etc.) behavior of the **device-related reference** as described by the author of the post. This means that the annotation of ANOMALY does not need to be aware of actual, real or intended normal behavior, but it needs to capture the author's point of view."

As we mentioned, the discourse patterns signaling anomaly descriptions are less apparent on the surface, and moreover it may be difficult to distinguish them from descriptions of negative phenomena that represent normal behaviour. These are the two main issues our discourse model aims to account for.

Our model is based on the analysis of the meaning of ANOMALY. According to our definition above, anomaly descriptions consist of two basic meaning elements: a deviance-related element and a device-related reference. The basic ANOMALY pattern thus is the following:

ANOMALY = DEVIANCE + TERM

In the following sentences the surface indicators of DEVIANCE are underlined, and the TERMs are in bold:

(13) I'm seeing a ripple-like on an external display attached to my 2010 Mac Book Pro via VGA (Display adapter).

(14) For some reason, doesn't always update the list of after I plug in my iPhone.

(15) I have created a number of desktops in which to keep Safari, Mail, iCal etc. These are all kept in full-screen mode and I have unchecked the setting to automatically rearrange spaces, yet whenever I restart my the have changed order.

Whereas the TERMs are final elements, i.e. they are instantiated by lexical units, DEVIANCE may be conveyed either by final elements as in sentence (13) or by further complex discourse elements, as in sentences (14) and (15).

In the following sections we briefly describe our method for the acquisition of the TERMs, the discourse model underlying the anomaly descriptions and the implementation of the anomaly detection module as a set of rules.

3.3 Detection of Device-related Terms

In this section, we describe the method for extracting terms from the data we described above. We apply a general approach using jointly topic modeling and TF-IDF for finding the words which can be considered as domain specific terms, a method that has given good results on other cases [3], [5].

We use the corpus built from a collection of 1000 posts sample from the Ask Different forum (see Section 2). As pre-processing we remove stopwords using the standard list of stopwords in MALLET [9] as well as URLs. The method for extracting terms consists of three steps:

- We use the MALLET package to extract the list of related topics associated with related words.
- We compute TF-IDF to find the relevant unigrams in the corpus, which we postulate to be the domain-specific terms.
- We integrate the results of the steps above and filter a list of words considered as significant terms in the corpus.

In parallel to the detection system using the TERMs provided by the topic model, we also implemented a system where we used the forum tags as a proxy to the gold standard TERMs. The combination of both was kept for the rest of the experiments.

3.4 Discourse Model of Anomaly Descriptions

Sentences (13) through (15) represent three main discourse patterns of anomaly descriptions. The difference between the three patterns is the realization of the deviance-related element.

1. In example (13) DEVIANCE is indicated through the lexical unit *ripple-like*. Its meaning inherently involves anomaly, since *ripples* are not normal properties of a display. The discourse pattern matching such anomaly descriptions thus consists of two sentence elements corresponding to the two main elements of anomaly descriptions: DEVIANCE and TERM. In the ANOMALY pattern these two elements need to be in a syntactic dependency relation to ensure their semantic cohesion, as shown in Figure 3. The nature of the dependency as well as the order of the elements is irrelevant.

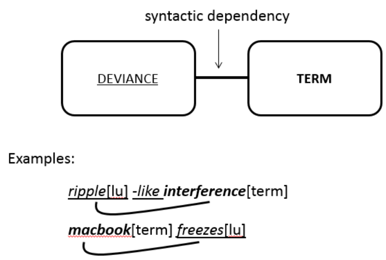

Figure 3. Pattern of anomaly descriptions constituted by syntactically related elements with particular semantic features

2. In example (14) the deviance involves the negative predicate *doesn't … update*. This sentence element, however does not inherently convey DEVIANCE, as illustrated by other sentences where the same negative event may describe a normal behaviour:

   (16) I believe this is an expected behavior. Although PWD <u>doesn't update</u> the path right away, if you do something else like cd .. or even cd . you can see that the path gets updated.

In sentence (14) the deviant character of the event is indicated in addition to NEGATION by *For some reason* and *not always.* These elements convey important aspects of the negative predicate, which indicate deviance from normal. *For some reason* indicates that the author does not know the cause of the event, which implies that she is uncertain if the event happens as expected, and *not always* indicates irregularity, which implies deviance from expected, regular behaviour. (cf. "deviation from normal (correct, usual, expected, good, etc.) behavior" in the definition.)

A more detailed analysis of such deviance indicators has yielded the following aspects:

- temporal (random, irregular, changing): *<u>sometimes</u>, <u>not always</u>, <u>random</u>, <u>unpredictable</u>, <u>not any more</u>, <u>stop</u>, (dis)appear,* etc.
- knowledge-related (uncertain knowledge of causes, of phenomena): *<u>for some reason</u>, <u>seems</u>,* etc.
- contrast: *<u>yet</u>, <u>but</u>, <u>even</u>, <u>however</u>,* etc.
- pragmatic (emphasis): *<u>keeps</u>, <u>constantly</u>, <u>simply</u>, <u>whenever</u>,* etc.

In this more complex discourse pattern of anomaly descriptions the DEVIANCE meaning is distributed between two elements: NEGATIVE + DEVIANCE ASPECT INDICATOR (DAI). The NEGATIVE element can be instantiated either by a grammatical operation (negation) or by lexical elements with negative meaning. NEGATIVE and TERM are in a syntactic dependency relation, which reflects their semantic cohesion. As for DAI, a syntactic dependency relation with either NEGATIVE or TERM is required if the DAI is instantiated by a main category lexical unit, otherwise co-occurrence is sufficient. The order of the elements is not relevant. These patterns are presented in Figure 4.

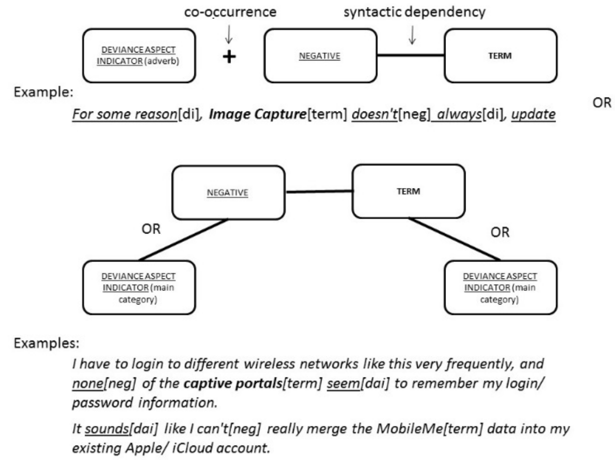

Figure 4. Pattern of anomaly descriptions constituted by syntactically related and co-occurring lexical, grammatical and discourse elements

3. In sentence (15) the deviant event is *have changed order*, which does not contain any negative meaning element. In this sentence this negative element is conveyed by *yet*, a CONTRAST INDICATOR OF DEVIANCE. These three elements (PREDICATE, TERM and CONTRAST INDICATOR OF DEVIANCE), however, do not necessarily describe an anomaly, like in the following sentence:

   (17) Isn't '720p' 1280 x 720, <u>yet</u> the or <u>is</u> <u>1024x768</u>?

The anomaly pattern in sentences like (15) contain an additional DEVIANCE ASPECT INDICATOR, like *<u>whenever</u>*, which indicates repetition in an emphatic way, as a sign of frustration. Thus this pattern requires the co-occurrence of two DEVIANCE ASPECT INDICATORs out of which one should convey CONTRAST, as well as a predication including a TERM, as illustrated in Figure 5. Similarly to the previous patterns, the order of the elements is not relevant.

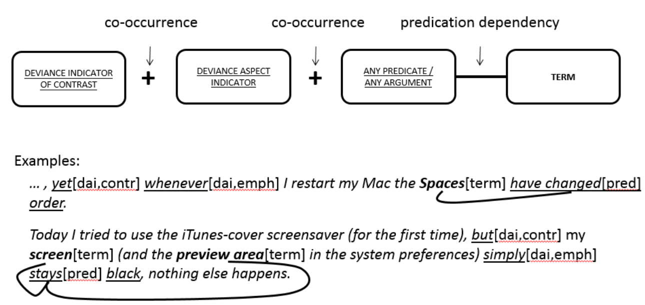

Figure 5. Pattern of anomaly descriptions constituted by multiple co-occurring DEVIANCE ASPECT INDICATORs and a predication containing a TERM

In summary our discourse model of ANOMALY is constituted by four final meaning elements (Figure 6): DEVIANCE ASPECT INDICATOR (randomness, uncertainty, etc.), CONTRAST INDICATOR OF DEVIANCE and any PREDICATE/ARGUMENT of TERM and TERM. They may all be present in anomaly descriptions, or the first three elements may merge in higher-level meaning elements: They may be merged in a single lexical unit conveying DEVIANCE, and CONTRAST INDICATOR OF DEVIANCE and any PREDICATE/ARGUMENT may be merged as NEGATION. The following schema represents the model:

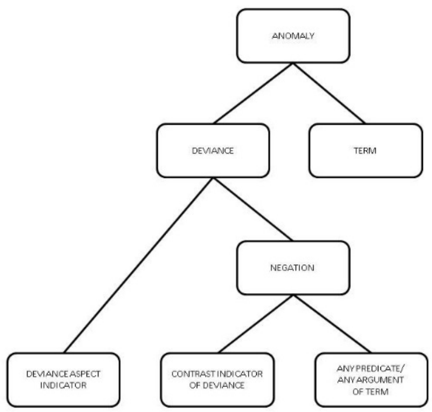

Figure 6. Discourse model of ANOMALY

In the sentences these meaning elements are instantiated as the lexical or syntactic units with various relationships. We'll discuss these points in the following section.

We do not claim that this model is exhaustive. However, we manually tested it on a number of random examples from different technical corpora, and we have not found an anomaly description that is not covered by it.

4. IMPLEMENTATION

We have implemented the discourse patterns in XIP using various features and rules.

The information request patterns are implemented by rules that recognize direct and indirect questions based on part-of-speech categories and syntactic features provided by XIP (e.g. the presence of an interrogative pronoun, verbs introducing indirect questions, like wonder; inverted order of subject and verb) as well as on the presence of lexical units (e.g. *how, how to, cause*).

The meaning elements of ANOMALY are implemented as a heterogeneous feature set used in syntactic and co-occurrence the rules. These features are listed in Table 3.

Apart from the lexical features the resources are readily provided by XIP: part of speech categories, dependency analysis and negation operation marks. At this stage the system uses the lexical resource of domain specific terminology acquired by topic modeling, as described in Section 3.2.3., and the lists of words that instantiate the remaining lexical features in the development corpus. The acquisition of the remaining lexical resources (deviance, negation) is left for future work, which is obviously a limitation for the coverage of the current system, although some of the words (e.g. *freeze*) are recurring.

Table 3. Linguistic features associated with the meaning elements

Meaning elements of ANOMALY	Feature
TERM	lexical
DEVIANCE	lexical
NEGATION	negation operation
	lexical
DISCOURSE INDICATOR	part of speech category (temporal adverbs)
	lexical
CONTRAST DISCOURSE INDICATOR	part of speech category (negative connectors)
PREDICATE	dependency (subject, object, subject attribute)

The rules specify the nature of the co-occurrence of the features in the sentences, which can be simple co-occurrence, any dependency relation or a specific dependency relation. Whenever the features are assigned to verbs, nouns or adjectives, at least one dependency relation between two elements is required. This dependency relation is only specified in sentence type 3 as predication (i.e. subject or object dependency), otherwise it can be any dependency. When the features are assigned to adverbs or connectors, co-occurrence signals sufficient cohesion among the elements of the patterns.

5. RELATED WORK

Mining forum posts has been recognized as an important task for various use cases. Here we provide an overview of the ones that use discourse related features.

[12] distinguishes between problem and solution posts where the forum structure does not indicate it. The authors train a CRF classifier based on discourse move annotated technical forum corpora. The classifier distinguishes between relevant discourse moves, which describe problems, problem queries, suggest solutions and resolution steps, and those that are irrelevant for the classification, like greetings and messages to the author. This work could be complementary to ours as we assume that the post is already classified as a question or resolution. We rather identify and consequently classify the sentences in the question posts that indicate one of our pre-defined categories.

Based on dialogue act tagging and coherence-based discourse analysis [6] and [15] identify and link problem and solution pairs in troubleshooting forum posts. Moreover, they label the links according to their relationship to the previous discourse act, as ADD, CONFIRMATION, CORRECTION, etc. One particular use of discourse markers is the detection of resolved problems proposed by Wang et al. (2012). Again in that case the use of discourse-related features is aiming at solving a different task.

Identifying and characterizing forum threads has also been an active research problem. Previous work mainly addresses the classification of troubleshooting threads. [2] distinguishes between specific vs. general problems, the complete or not complete initial post in the thread, and resolved or not resolved threads. [10] performs clustering of similar troubleshooting posts and builds hierarchies among post types.

Investigations that aim at identifying and typing sentences/sections in forum posts are closer in spirit to the work we are presenting in this paper. [13] coins this task as sentence extraction from forums. They have developed CRF and SVM classifiers to distinguish between sentences describing physical examination – which corresponds to problem formulation – and those describing medication – which corresponds to solutions. Their method, though, does not consider different sorts of discourse moves that indicate the two main categories. In the same area we can also include the [11], which extracts segments from documents that convey the "basic intent" of the author i.e. each segment corresponding to a different topic found in the document is defined as a basic intent. This could include for instance questions, problems and/or solutions. However, the authors do not provide any classification of the intent types.

A relevant field is that of question classification in question answering systems. [8] defines question classification as the task of predicting the entity type or category of the expected answer. However, traditional question answering does not deal with identifying and extracting the questions from unstructured text, but only typing them. Furthermore, the main body of work performed in question classification is based on a taxonomy proposed by [7], which is more oriented towards open domain information retrieval and does not include the categories that interest us here (with the most notable difference being of course the category "ANOMALY").

6. EVALUATION

We evaluated the performance of the system by comparing the automated classification results with the gold-standard classification of the 50 test question posts of the corpus. In order to assess the role of discourse analysis in the classification we also performed experiments on the different categories using a hybrid classifier.

6.1 Performance against the gold standard

The following table shows the results of our system in terms of precision, recall and F1 measures for the different categories:

Table 4. Performance of the system for the different categories

	ANOMALY	HOWTO	PROPERTY	EXPLAN	NULL	AVERAGE
PREC	0.531	0.803	0.814	0.166	0.869	0.637
REC	0.459	0.833	0.803	0.5	0.887	0.696
F1	0.495	0.818	0.808	0.333	0.877	0.666

These results indicate that the simple patterns of the HOWTO and the PROPERTY categories have captured fairly well the actual language patterns. The results for EXPLANATION cannot be considered as representative due to the fact that the test corpus only contained 2 gold-standard sentences in this category.

As discussed earlier the detection of the ANOMALY class is challenging due to the wide lexical and structural variety of the anomaly descriptions. Since our system implements part of the lexical resources required by the discourse patterns, the performance results are lower. We expect the improvement of the results by the injection of more lexical resources.

As a way of assessing the role of the discourse analysis patterns in the detection of the categories we performed experiments described in the following section using a hybrid classification system.

6.2 Experiments

6.2.1 Method

The system comprises two basic modules:

1. The syntactic parsing component based on XIP, to detect linguistically rich information (POS, syntactic dependencies, discourse patterns etc.)
2. A sentence classification module that associates predefined categories to sentences (a given sentence may have multiple categories associated to it).

The syntactic parsing components provide linguistic information used as features by the classification modules, which yield the final output. The machine learning classification components are based on the standard classification library liblinear [4].

The sentence classification module is used to assign categories to sentences. For each sentence, the module takes as input features the bag of words in the sentence as well as information provided by the syntactic parsing component. The output consists of a list of categories corresponding to each sentence associated with their probabilities. In the pre-processing stage, stop words are removed (determinants, conjunctions).

We use the L2-regularized logistic regression solver from the Liblinear library to train the classification model. Features include unigram, bigrams, POS, and discourse patterns extracted by the rule-based component presented above. Classification results are described in the following subsection.

6.2.2 Results

We first trained the classifier using various standard features (unigram, bigrams, part-of-speech). The best results were obtained by using bigrams and part-of-speech features. In order to test the role of the discourse patterns we added as a feature to the bigram and part-of-speech features the output of the classification by XIP. The following table shows the comparison of the results:

Table 5. Comparison of the output of two classifiers

		ANOMALY	HOWTO	PROPERTY	EXPLAN	NULL	AVERAGE
BIGRAM+POS	PREC	0.444	0.854	0.764	0	0.801	0.578
	REC	0.108	0.759	0.732	0	0.925	0.505
	F1	0.276	0.806	0.748	0	0.862	0.534
BIGRAM+POS+XIP	PREC	**0.5**	0.883	**0.854**	**1**	0.767	**0.8**
	REC	**0.27**	0.704	0.577	**0.5**	**0.971**	**0.604**
	F1	**0.385**	0.794	0.716	**0.75**	**0.869**	**0.703**

As we can see the discourse feature does not have a great effect on the HOWTO and PROPERTY categories. This is expected, because these categories are expressed with simple surface patterns, which the machine-learning algorithms capture. The result for EXPLANATION is much better using the discourse feature, however, as we mentioned above, this is not significant due to the few cases in the corpus.

However, the role of the discourse feature is apparent for the ANOMALY category: the precision is slightly better and the recall is more than the double. This result indicates the important role of discourse analysis in detecting this challenging category.

7. CONCLUSION

We presented a method for identifying, extracting, and typing user issues and requests in question posts of online discussion forums. As part of the method, we presented a discourse-based analysis of relevant information and described how corresponding features have been integrated in an implemented system. Our experiments show that discourse related features are especially useful when dealing with complex concepts such as the notion of anomalies expressed in question posts. We are releasing the dataset we used, together with our annotation. In the immediate future we plan to extend our method for identifying important information types in solution posts, and evaluating the domain-(in)dependence of discourse-related features across a number of other online forums.

8. REFERENCES

[1] Aït-Mokhtar, S., Chanod, J. P., & Roux, C. 2002. Robustness beyond shallowness: incremental deep parsing. Natural Language Engineering, 8(2-3), 121-144.

[2] Baldwin, T., Martinez, D., & Penman, R. B. 2007. Automatic thread classification for Linux user forum information access. In *Proceedings of the Twelfth Australasian Document Computing Symposium* (ADCS 2007) (pp. 72-9).

[3] Bolshakova, E., Loukachevitch, N., and Nokel, M. 2013. Topic models can improve domain term extraction. In *Proceedings of the 35th European conference on Advances in Information Retrieval* (ECIR'13). Serdyukov, P., Braslavski, P., Kuznetsov, S. O., Kamps, J., and Rüger, S. (Eds.). Springer-Verlag, Berlin, Heidelberg (pp. 684-687).

[4] Fan, R.-E. Chang, K.-W., Hsieh, C.-J., Wang, X.-R. and Lin. C.-J. 2008. LIBLINEAR: A Library for Large Linear Classification, J. Mach. Learn. Res. 9 (2008), 1871-1874.

[5] Habibi, M.; Popescu-Belis, A. 2015. Keyword Extraction and Clustering for Document Recommendation in Conversations. In Audio, Speech, and Language Processing, IEEE/ACM Transactions on , vol.23, no.4, (pp.746-759).

[6] Kim, S. N., Wang, L., & Baldwin, T. 2010. Tagging and linking web forum posts. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning* (pp. 192-202). Association for Computational Linguistics.

[7] Li, X. and Roth, D. 2002. Learning question classifiers. In Proceedings of the 19th international conference on Computational linguistics - Volume 1 (COLING '02), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 1-7. DOI=http://dx.doi.org/10.3115/1072228.1072378

[8] Loni, B. 2011. A survey of state-of-the-art methods on question classification.Literature Survey. Published on TU Delft Repository. http://repository.tudelft.nl/assets/uuid:8e57caa8-04fc-4fe2-b668-20767ab3db92/A_Survey_of_State-of-the-Art_Methods_on_Question_Classification.pdf

[9] McCallum, Andrew Kachites. 2002. MALLET: A Machine Learning for Language Toolkit. http://mallet.cs.umass.edu.

[10] Medem, A., Akodjenou, M. I., & Teixeira, R. 2009. Troubleminer: Mining network trouble tickets. In *Integrated Network Management-Workshops*, 2009. IM'09. IFIP/IEEE International Symposium on (pp. 113-119). IEEE.

[11] Mukherjee, S., & Joshi, S. Help Yourself: A Virtual Self-Assist Agent. 2014. WWW'14 Companion, April 7–11, 2014, Seoul, Korea.

[12] Raghavan, P., Catherine, R., Ikbal, S., Kambhatla, N., & Majumdar, D. 2010. Extracting Problem and Resolution Information from Online Discussion Forums. In COMAD (p. 77).

[13] Sondhi, P., Gupta, M., Zhai, C., & Hockenmaier, J. 2010. Shallow information extraction from medical forum data. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (pp. 1158-1166). Association for Computational Linguistics.

[14] Wang, L., Kim, S. N., & Baldwin, T. 2012. The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums. In COLING (pp. 2739-2756).

[15] Wang, L., Lui, M., Kim, S. N., Nivre, J., & Baldwin, T. 2011. Predicting thread discourse structure over technical web forums. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 13-25). Association for Computational Linguistics.