A Multi-faceted Approach to Query Intent Classification

Cristina González-Caro[1] and Ricardo Baeza-Yates[2]

[1] Universidad Autónoma de Bucaramanga.
Avenida 42 No. 48 - 11, Bucaramanga, Colombia
cgonzalc@unab.edu.co
[2] Yahoo! Research Barcelona.
Diagonal 177, 9th floor. 08018, Barcelona, Spain
ricardo.baeza@acm.org

Abstract. In this paper, we report results for automatic classification of queries in a wide set of facets that are useful to the identification of query intent. We also investigate whether combining multiple facets can help to increase the predictability of the facets. Our hypothesis is that the performance of single-faceted classification of queries can be improved by introducing information of multi-faceted training samples into the learning process. We test our hypothesis through performing multi-faceted classification of queries based on the combination of correlated facets. Our experimental results show that the combination of correlated facets can effectively improve the quality of the classification results. Since most of the previous works in query intent classification are oriented to the study of single facets, these results are a first step to an integrated query intent classification model.

1 Introduction

As the Web continues to increase both in size and complexity, Web search is a ubiquitous service that allows users to find information, resources, and activities. However, as the Web evolves so do the needs of the users. Nowadays the needs of the users involve more complex interests that go beyond to the traditional *informational queries*. For example, many users may want to perform a particular commercial transaction, locate a special service, etc. Thus, it is important for Web-search engines, not only to continue answering effectively informational queries, but also to be able to identify and provide accurate results for the new types of queries.

Attending to the challenge of the new search paradigms, Web-search engines try to improve the quality of their results by adopting a number of different strategies. For instance, diversification of search results aims at providing a list of diversified results that cover different interpretations of ambiguous queries [8, 21, 1]. The objective of diversification is to identify ambiguous queries and present the best results for each meaning of the those queries. The first step towards this goal, is to identify the type of the query. In this respect, all the recent efforts to describe and identify the intent of the uses query are of great value [12, 27, 11]. Although there is a lot of work in the topic of identifying query intent, most of this previous work is based on the analysis of only one possible facet of the query. The most common of these facets are the *topic category* and the *type of query intent*; mainly based on Broder's taxonomy [4]. However, one can argue that classifying a query with respect to one facet may improve the classification with respect to another facet. For example, knowing that a query topic is *art* increases the prior that the query intent is *informational*. Similarly, knowing that a query topic is *electronics* increases the prior that the query intent is *transactional*. Motivated by the previous example, we argue that identification of the query intent is a multi-faceted problem. We show that by treating the problem as such we can significantly improve the accuracy of the classification problem.

In this paper, we explore the automatic classification of queries in a wide set of facets that are useful to the identification of query intent. We consider the set of facets that have been well analyzed in our prior work [9, 5]. We also investigate whether combining multiple facets can help to increase the predictability of the facets.

Our contributions to the topic of query-intent classification include:

- We present results for the automatic classification of a large set of queries into a comprehensive set of facets, showing the feasibility of an automatic faceted-classification of the query intent.
- We propose a multi-faceted classification of queries based on the combination of correlated facets. We evaluate the performance of each combination of facets and its impact on each individual facet.
- We compare the results of multi-faceted classification with the conventional faceted classification to determine if the combination of related facets can improve the identification of the intent of the user queries.
- We provide an extensive experimental evaluation showing that the combination of facets proposed in this paper can significantly improve the quality of the classification results.

To the best of our knowledge, this is the first work that explore automatic multi-faceted classification of user's query intent. Previous work has considered the multi-faceted classification of the query intent an open research problem [12].

2 Related Work

According to Cool and Belkin [7], users engage in multiple information seeking behavior within the context of accomplishing a single task. Therefore, it is important to have Web-search systems able to support multiple information seeking behaviors, and multiple interactions with the information. In this direction, many efforts have been devoted in trying to understand the intent of user queries. The understanding of user queries has been conducted from different facets. The first approaches to identify query intent include classifying queries based on the topic [3, 23, 16]. Topical associations of queries are important because they allow to place the queries in a particular context. A second line of work has focused on the classification of queries regarding the type of intent. In this case, the intent of the query refers to the type of resource associated with the query. The first taxonomy of query intent, proposed by Broder [4], defines three types of query intent: informational, navigational and transactional. Rose and Levinson [22] extended Broder's taxonomy by adding hierarchical sub-categories for informational and transactional queries. Based on these early taxonomies many works have attempted automatic classification of queries [15, 17, 27, 12, 11]. Other approaches to identifying query intent consider facets like geographic locality [10, 14], time sensitivity [13, 18], and ambiguity [24, 26]. However, each of these works has been restricted to the analysis of only one facet. Meanwhile, there have been few works attempting to classify query intent in more than one facet. Baeza-Yates et al. [2] presented an approach for automatic classification of queries into topic and query intent (informational, not informational and ambiguous). Nguyen and Kan [19]analyzed a set of four facets (ambiguity, authority sensitivity, temporal sensitivity and spatial sensitivity). However, although they presented a query log analysis of the four facets, they only provided automatic classification results for one of the facets: authority sensitivity. None of these studies implement multi-faceted classification of query intent. They only implement automatic classification of query intent based on the information of individual facets.

3 Experimental Design

In this research, we leveraged prior work reported in [5, 9]. We adopted the data-set processed in these works. The data-set is composed of 4726 queries extracted from a query-log of TodoCL[1], a real case search engine. Each query was represented as a term-weight vector compounded by the terms appearing in the selected Web pages for such query. The queries were manually classified into a set of nine facets, that can be used for the identification the query intent. These facets are:

Genre{*News, Business, Reference, Community*}: this facet provides a generic context to the user's query intent, and can be thought as a meta-facet.

Topic {*Adult & Sex, Arts & Culture, Beauty & Style, Cars & Transportation, Charity, Computers & Internet, Education, Entertainment, Music & Games, Finance, Food & Drink, Health, Home & Garden, Industrial Goods & Services, Politics & Government, Religion & belief systems, Science & Mathematics, Social Science, Sports, Technology & Electronic, Travel, Undefined, Work*}: a list of topics built from the first level of categories offered by ODP[2], Yahoo![3], and Wikipedia[4].

Task {*Informational, Not Informational, Both*}[2]: this facet is related with the type of resource associated with the query.

[1] www.todocl.com
[2] www.dmoz.org
[3] www.yahoo.com
[4] en.wikipedia.org/

Objective {*Resource, Action*}. Represents if the query is aimed to do some action or to obtain a Resource.

Specificity {*Specific, Medium, Broad*}: this facet describes how specialized is a query.

Scope {*Yes, No*}: the scope aims at capturing whether the query contains polysemic words or not.

Authority Sensitivity {*Yes, No*}: through this facet it is possible to determining whether the query is designed to retrieve authoritative and trusted answers [19]

Spatial Sensitivity {*Yes, No*}: this facet reflects the interest of the user to get a resource related to an explicit spatial location.

Time Sensitivity {*Yes, No*}: this facet captures the fact that some queries require different results when posed at different times [19].

The facets genre, objective, specificity and scope are considered for first time for query intent classification. These facets provide valuable information for the identification of the user query intent and it is important evaluate the feasibility of their automatic prediction.The rest of the facets have been considered in previous works for query intent classification (see Section 2) and it is also interesting to evaluate the predictability of these facets and their relation with the rest of the facets.

4 Predicting Individual Facets

In this section we use the labeled data-set described in Section 3 to train user intent prediction models. We trained a single support vector machine (SVM) classifier for each facet. The software used to implement SVM was LIBSVM [6]. We selected the one-against-one multi-class strategy and the radial basis function kernel for the SVM algorithm. In order to validate and measure the performance of the prediction, we split the data into training and unseen test sets by using percentage-fractions of 50/50 and 70/30. For the experiments, three metrics were considered: the recall, the precision and the f-measure.

Table 1. Performance evaluation of automatic prediction for Task.

Task	50%/50%			70%/30%		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Informational	0.7037	0.9889	0.8223	0.7227	0.9915	0.8360
Not Informational	0.8408	0.2670	0.4053	0.8917	0.2948	0.4431
Both	0.9167	0.0550	0.1038	0.8571	0.0526	0.0992
Average	0.8204	0.4370	0.4438	0.8238	0.4463	0.4594

The results for the automatic prediction of the facets are reported from Table 1 to Table 5. As we can see from these tables, in general we obtain good results for estimating the facets. There is not too much difference between the

Table 2. Performance evaluation of automatic prediction for Objective.

Objective	50%/50% Precision	50%/50% Recall	50%/50% F-Measure	70%/30% Precision	70%/30% Recall	70%/30% F-Measure
Action	0.9451	0.1673	0.2843	0.9375	0.2007	0.3306
Resource	0.8116	0.9973	0.8949	0.8235	0.9964	0.9017
Average	0.8783	0.5823	0.5896	0.8805	0.5985	0.6162

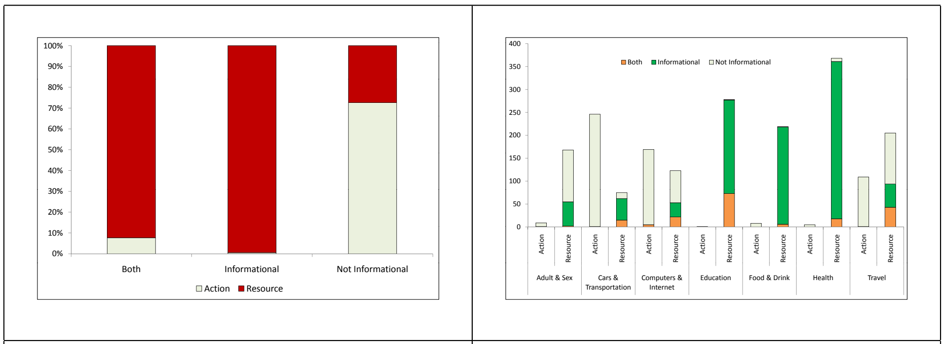
results obtained with the automatic classifiers based on training-sets of 50% of queries and those classifiers based on training-sets of 70% of queries. That is, the prediction's performance is good for different quantities of training data.

The best results are for the facets task and objective (see Table 1 and Table 2). The average precision for these facets is 0.822 and 0.879 respectively. In the case of task, the classifier was most useful for predict *Informational* and *Not Informational* queries, where we can observe a good balance of precision and recall (see f-measure values). For queries in *Both* category, we maintain high precision (0.88 on average) at the expense of recall, the recall could be improved. In the case of objective, the classifier is very effective, specially to distinguish *Resource* queries (f-measure=0.89 on average). For *Action* queries the precision is also very good (0.94 on average), although the f-measure is not as high as for *Resource* queries. In general, the automatic classification results for task and objective show the feasibility of the prediction of these facets. This is important, be able to correctly identify the task and the objective of queries give us an insight into the query intent and the type of resource associated with this intent. In Figure 1-A we show the distribution of queries along the task and the objective. As we can observe, *Informational* queries have a clear orientation towards *Resource*-objective and *Not Informational* queries are more oriented towards *Action*-objective. An interesting point is that *Ambiguous* queries are also oriented towards *Resource*-objective. This finding suggest that queries with an *Action*-objective are less ambiguous than queries with a *Resource*- objective. Since most of the queries with an *Action*-objective belong to the *Not Informational*- task we can say that the ambiguity of *Not Informational* queries is low.

Table 3. Performance evaluation of automatic prediction for Genre.

Genre	50%/50% Precision	50%/50% Recall	50%/50% F-Measure	70%/30% Precision	70%/30% Recall	70%/30% F-Measure
Business	0.9146	0.2884	0.4386	0.8984	0.3159	0.4675
Community	0.5649	0.9867	0.7185	0.5765	0.9836	0.7269
Reference	0.8033	0.1038	0.1839	0.8537	0.1203	0.2108
Average	0.7609	0.4597	0.4470	0.7762	0.4733	0.4684

The Table 3 shows the performance results for the facet genre. We obtain good genre identification performance for *business*, *community* and *reference*, SVM classifiers yielded good overall precision, around 0.76 (f-measure=0.46 on



A. B.

Fig. 1. (A.) Distribution of queries into facets Task and Objective. and (B.)Distribution of queries into facets Topic, Task and Objective.

average). These three genre categories are the most representative categories of the facet, they group together 97.4% of the total of queries. For the category *news* the classifier obtains a f-measure of zero, which may be caused by the availability of a small number of queries belonging to this genre category (56 queries in the training set, 122 of the total of queries). Apart of the *news* category, the performance of prediction for the facet genre is good.

Table 4. Performance evaluation of automatic prediction for Topic.

	50%/50%			70%/30%		
Topic	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Adult & Sex	0.7692	0.2000	0.3175	0.7000	0.2414	0.3590
Arts & Culture	0.5769	0.0938	0.1613	0.5556	0.1020	0.1724
Cars & Transportation	0.9107	0.3778	0.5340	0.8800	0.2857	0.4314
Computers & Internet	0.8600	0.3308	0.4778	0.8261	0.3167	0.4578
Education	0.8043	0.2483	0.3795	0.8529	0.3222	0.4677
Entertainment. Music & Games	0.1502	0.8743	0.2564	0.5474	0.4685	0.5049
Finance	0.6622	0.3858	0.4876	0.2149	0.8052	0.3393
Food & Drink	0.8667	0.2364	0.3714	0.9500	0.2923	0.4471
Health	0.6383	0.6667	0.6522	0.7347	0.4800	0.5806
Home & Garden	0.7692	0.1538	0.2564	0.5600	0.1667	0.2569
Politics & Government	0.4483	0.4437	0.4460	0.4903	0.5549	0.5206
Religion & Belief Systems	1.0	0.1224	0.2182	1.0	0.1515	0.2632
Science & Mathematics	0.4279	0.6394	0.5127	0.3846	0.6765	0.4904
Social Science	0.4375	0.0753	0.1284	0.3636	0.0690	0.1159
Travel	0.7083	0.2252	0.3417	0.7333	0.2472	0.3697
Average	0.6687	0.3382	0.3694	0.6529	0.3453	0.3851

For the facet topic, the classifier provides predictions for fifteen of the twenty topics that were considered for the classification. The Table 4 shows the automatic classification results of this facet. Overall, the precision is good, over 0.6 for most of the topics. The best precision values are for the topics: *Adult & Sex, Cars & Transportation, Computers & Internet, Education, Food & Drink* and

Table 5. Performance evaluation of automatic prediction for Authority Sensitivity, Spatial Sensitivity, Time Sensitivity, Scope and Specificity.

Facet	50%/50%			70%/30%		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Authority Sen.	0.8570	0.5193	0.4865	0.8441	0.5245	0.4961
Spatial Sen.	0.7526	0.6140	0.5471	0.7618	0.6494	0.5913
Time Sen.	0.4884	0.50	0.4941	0.4873	0.50	0.4936
Scope	0.49450	0.50	0.49723	0.4958	0.50	0.4979
Specificity	0.5126	0.3412	0.3101	0.5060	0.3444	0.3160

Health. These topics group an important number of queries and since have interesting connections with the other facets (see Figure 1-B). The topics for which the classifier does not provide predictions are: *Beauty & Style, Industrial Goods & Services, Sports, Technology & Electronic* and *Work*. All these topics have a very small concentration of queries (less than 50 queries each one in the sample set). This small concentration of queries suggest that these topics should be reconsidered in the list of topics. For example, the topic *Work* could be contained in the topic *Finance*.

The Table 5 shows the performance of the prediction for the facets authority sensitivity, spatial sensitivity, time sensitivity, scope and specificity.The results for this group of facets is good. Some of the facets obtain better results than the others, but overall results are balanced. The facets spatial sensitivity and authority sensitivity show the best precision results. Be able to identify correctly these two facets is important because the search results for spatially-sensitive and authority-sensitive queries, must be both relevant to the query and valid for the associated location and for the authoritative requirement. The performance for Time sensitivity, scope and specificity is similar, on average. These facets are a complement for the other facets, that is, these facets add information to the context of the query (e.g. scope informs if the query has polysemic words) but they do not reveal the intent of the user for themselves.

5 Combining Multiple Facets

Although the set of facets that we are studying here are different dimensions of the user's intent and not all of them are necessarily correlated, we are interested to study how the combination of multiple facets in the classification process can improve the performance of the prediction of the facets. Based on the log analysis made [9], we selected two groups of related facets and perform a multi-faceted classification of the queries with these facets. We present the results of the multi-faceted classification and then we make a comparison of these results with the results of the single-faceted classification. Our hypothesis is that the performance of the single-faceted classifiers can be improved by introducing the information of multi-faceted training samples into the learning procedure.

5.1 Multi-label Classification

We address the problem of classify a query into a set of relevant facets as a multi-label classification problem. Multi-label classification is an extension of traditional multi-class classification problem in which its classes are not mutually exclusive and each sample may belong to several classes simultaneously. Unlike multi-class classification, which each instance is assumed to be classified into only one category from a set of predefined categories, the objective of multi-label classification is to predict a set of relevant labels for a given input. For detailed information about multi-label classification, the interested reader may refer to [25].

Since support vector machine have shown good generalization ability in different single-label multi-class problems, is also one of the most used techniques to resolve multi-label problem [20]. We used the multi-label classification tool of LIBSVM[5] to build multi-label classifiers for our group of facets. We selected the label combination method.

5.2 Genre-Objective Combination

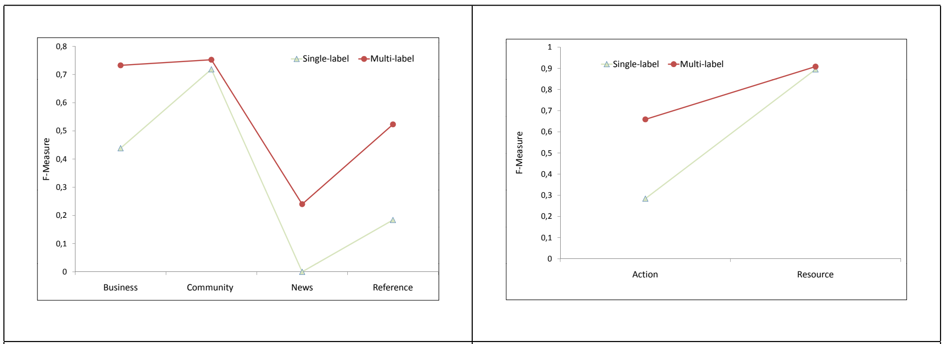
According to [9], two of the most correlated facets are genre and objective. For this reason we selected this combination of facets to test multi-label classification. In order to obtain comparable results, the training and test sets used for multi-label classification are the same data-sets used to perform single-label classification. That is, we used the same group of queries and the only variation are the training-set labels. In this case, each query was marked with two values, genre and objective respectively.

Table 6. Multi-label classification results based on the combination Genre-Objective for the facet Genre.

	50%/50%		
Genre	Precision	Recall	F-Measure
Business	0.7620	0.7052	0.7325
Community	0.7204	0.7874	0.7524
News	0.2727	0.2143	0.2400
Reference	0.5561	0.4936	0.5230
Average	0.5778	0.5501	0.5620

The Table 6, shows the multi-label classification results for the facet genre. Overall, multi-label classification outperform single-label classification for all categories of genre. Specially we appreciate important improvements in the recall for the small categories (i.e, categories with less representation of queries in the sample test) like *news* and *reference*. For *reference*, single-label classification yielded good precision result (0.82 on average) but the recall value is not very good (0.11

[5] available at http://www.csie.ntu.edu.tw/ cjlin/libsvmtools/multilabel/



A.	**B.**

Fig. 2. (A.) Comparison of single-label and multi-label F-Measure results for the facet Genre. and (B.)Comparison of single-label and multi-label F-Measure results for the facet Objective.

Table 7. Multi-label classification results based on the combination Genre-Objective for the facet Objective.

	50%/50%		
Objective	Precision	Recall	F-Measure
Action	0.6762	0.6420	0.6587
Resource	0.9019	0.9145	0.9082
Average	0.7890	0.7783	0.7834

on average). With the multi-label classification the results for *reference* are more balanced, the recall value improves around 0.5 and the precision is still good (up of 0.55). In the case of *news*, the overall performance is not so good, however it is notable that the multi-label classifier provides predictions for this category, given that the single-label classifier did not report results for this category. The category *news* was ignored for the single-label classifier because it is too small. When we use multi-label classification, the additional information provided with the multi-label training set allows the classifier to predict small categories. For the large categories the multi-label classification also improved the results. In the Figure 2-A we can see the comparison of the performance (f-measure values) of single-label and multi-label classification for the facet genre.

The results of the multi-label-classification for the facet objective are in the Table 7. The results for objective are even better than for the facet genre. With respect to the single-label classification results, the major improvements are for the category *Action* (see Figure 2-B). The recall value for *Action* changes from 0.18 to 0.64, which is a considerable improvement. For the category *Resource* the multi-label classifier maintain the high recall and precision obtained with the single-label classification (f-measure=0.908).

Overall, the combination genre-objective is positive for multi-label classification. From the results, we can see that genre-objective multi-label classification has better generalization performance than the traditional single-label classification.

5.3 Genre-Task-Topic Combination

Three of the most important facets we are evaluating are genre, task and topic. The combination of these facets might influence the prediction results of the queries. In the query log analysis reported in [9], they show that there are some topics that are oriented to specific genre categories. For instance, the topic *Cars & Transportation* is related with the genre-category *Business* and the topic *Politic & Government* is related with the genre-category *Community*. These relations suggest that knowing that a query belongs to a particular topic, could also indicate that the query belongs to a particular genre-category. Following, we test this hypothesis through performing multi-label classification with these three facets that according to the log analysis of queries, are related between them: genre, task and topic.

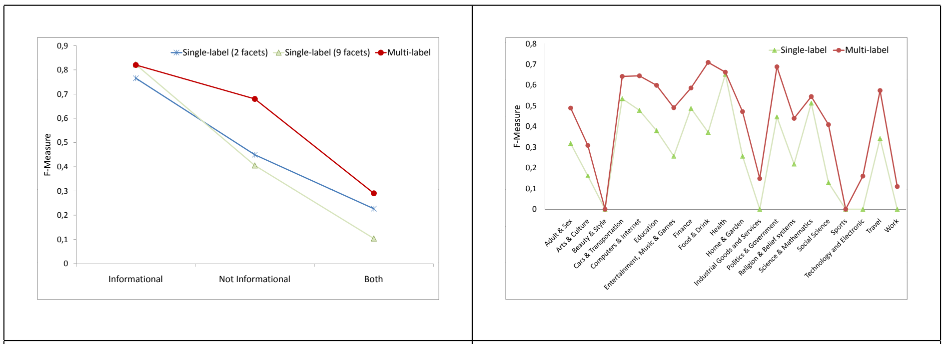
Table 8. Multi-label classification results based on the combination Genre-Task-Topic for the facet Genre.

	50%/50%		
Genre	Precision	Recall	F-Measure
Business	0.7170	0.7068	0.7119
Community	0.7116	0.7807	0.7446
News	0.1538	0.3214	0.2081
Reference	0.5908	0.3792	0.4619
Average	0.5433	0.5471	0.5316

The Table 8 shows the performance evaluation of the multi-label classification for the facet genre. In general, we obtain very similar results to the multi-label classification based on the combination genre-objective (see Tabele 6). The difference between the results of the two multi-label classifications is not high, the combination genre-objective has slightly better results, but in general the improvements are the same in both cases. Since genre is the facet that has more positive correlations with other facets, it could be combined in different ways with several facets.

Table 9. Multi-label classification results based on the combination Genre-Task-Topic for the facet Task.

	50%/50%		
Task	Precision	Recall	F-Measure
Informational	0.8235	0.8170	0.8202
Not Informational	0.6923	0.6682	0.6801
Both	0.2692	0.3150	0.2903
Average	0.5950	0.6001	0.5969



A.	B.

Fig. 3. (A.) Comparison of single-label and multi-label F-Measure results for the facet Task. and (B.)Comparison of single-label and multi-label F-Measure results for the facet Topic.

The results of the multi-label classification for the facet task are shown in the Table 9.The multi-label classification maintain the good results of the large categories and improve the results of the small categories. In this case, the large category is *Informational*. For this category, the multi-label classifier yields similar results than the single label classifier (f-measure=0.82). For the other two categories, *Not Informational* and *Both*, the multi-label classifier yields much better results than the single-label classifier. In the Figure 3-A we can see a comparison between the overall results (f-measure values) of the multi-label classification and the single-label classification for the facet task.

Finally, we show the results for the multi-label classification for the facet topic (see Table 10). The multi-label classifier provides predictions for eighteen of the twenty considered topics. With the single-label classification, five topics obtained f-measure values of zero. When we use combinations of facets to train automatic classifiers, the probability to predict small categories increases.We observe this effect of multi-label classification in facets like topic and genre. In general, the coverage (recall) of topics is substantially increased. The results of the multi-label classification are more balanced (precision-recall) than the results of the single-label classification. The Figure 3-B shows the f-measures values for the two types of classification. As we can observe, multi-label outperform single classification in all the topics. Some of them with remarkable improvements, like *Food & Drink* and *Politics & Government*.

In summary, the combination of the facets genre, task and topic is good for the automatic prediction of each facet. Specially, we show that the coverage of the facets increases considerably and the precision is balanced.

6 Conclusions

Some previous works have analyzed several facets, but they have not shown how these facets can be combined to automatically identify the intent of the query.

Table 10. Multi-label classification results based on the combination Genre-Task-Topic for the facet Topic.

Topic	50%/50%		
	Precision	Recall	F-Measure
Adult & Sex	0.5500	0.4400	0.4889
Arts & Culture	0.2313	0.4625	0.3083
Cars & Transportation	0.7333	0.5704	0.6417
Computers & Internet	0.6214	0.6692	0.6444
Education	0.6560	0.5503	0.5985
Entertainment. Music & Games	0.4944	0.4863	0.4904
Finance	0.5927	0.5787	0.5857
Food & Drink	0.7742	0.6545	0.7094
Health	0.6194	0.7111	0.6621
Home & Garden	0.5455	0.4154	0.4716
Industrial Goods and Services	0.5000	0.0870	0.1481
Politics & Government	0.7245	0.6553	0.6882
Religion & Belief Systems	0.5455	0.3673	0.4390
Science & Mathematics	0.6293	0.4796	0.5443
Social Science	0.3431	0.5054	0.4087
Technology and Electronic	0.2000	0.1333	0.1600
Travel	0.6074	0.5430	0.5734
Work	0.0606	0.5714	0.1096
Average	0.5238	0.4934	0.4818

As we can see from the results of our experiments, the best results are obtained with multi-label classification (multi-faceted classification).

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying search results. In: WSDM '09. pp. 5–14. ACM (2009)
2. Baeza-Yates, R., Calderón-Benavides, L., González-Caro, C.: The intention behind web queries. In: SPIRE'06. LNCS, vol. 4209, pp. 98–109. Springer (2006)
3. Beitzel, S.M., Jensen, E.C., Frieder, O., Lewis, D.D., Chowdhury, A., Kolcz, A.: Improving automatic query classification via semi-supervised learning. In: ICDM '05. pp. 42–49. IEEE Computer Society (2005)
4. Broder, A.: A taxonomy of web search. SIGIR Forum 36, 3–10 (2002)
5. Calderón-Benavides, L., González-Caro, C., Baeza-Yates, R.: Towards a deeper understanding of the user's query intent. In: Croft, W.B., Bendersky, M., Li, H., Xu, G. (eds.) Query representation and understanding. A workshop at SIGIR'2010
6. chung Chang, C., Lin, C.J.: Libsvm: a library for support vector machines (2001)
7. Cool, C., Belkin, N.J.: A classification of interactions with information. In: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science. pp. 1–15. Greenwood Village, CO: Libraries Unlimited (2002)
8. Dou, Z., Hu, S., Chen, K., Song, R., Wen, J.R.: Multi-dimensional search result diversification. In: Proceedings of the fourth ACM international conference on Web search and data mining. pp. 475–484. WSDM '11, ACM, New York, NY, USA (2011)
9. González-Caro, C., Calderón-Benavides, L., Baeza-Yates, R.: Web queries: The tip of the iceberg of user's intent. In: Carmel, D., Josifovski, V., Maarek, Y. (eds.) User modeling for web applications. A workshop at WSDM'2011

10. Gravano, L., Hatzivassiloglou, V., Lichtenstein, R.: Categorizing web queries according to geographical locality. In: CIKM'03. pp. 325–333. ACM (2003)
11. Herrera, M.R., de Moura, E.S., Cristo, M., Silva, T.P., da Silva, A.S.: Exploring features for the automatic identification of user goals in web search. Inf. Process. Manage. 46, 131–142 (2010)
12. Jansen, B.J., Booth, D.L., Spink, A.: Determining the informational, navigational, and transactional intent of web queries. Inf. Process. Manage. 44(3), 1251–1266 (2008)
13. Jones, R., Diaz, F.: Temporal profiles of queries. ACM Trans. Inf. Syst. 25 (July 2007)
14. Jones, R., Zhang, W.V., Rey, B., Jhala, P., Stipp, E.: Geographic intention and modification in web search. Int. J. Geogr. Inf. Sci. 22, 229–246 (2008)
15. Lee, U., Liu, Z., Cho, J.: Automatic identification of user goals in web search. In: WWW'05. pp. 391–400. ACM (2005)
16. Li, X., Wang, Y.Y., Shen, D., Acero, A.: Learning with click graph for query intent classification. ACM Trans. Inf. Syst. 28, 12:1–12:20 (2010)
17. Liu, Y., Zhang, M., Ru, L., Ma, S.: Automatic query type identification based on click through information. In: AIRS, pages 593–600, 2006.
18. Metzler, D., Jones, R., Peng, F., Zhang, R.: Improving search relevance for implicitly temporal queries. In: SIGIR'09. pp. 700–701. ACM (2009)
19. Nguyen, V.B., Kan, M.Y.: Functional faceted web query analysis. In: Amitay, E., Murray, C.G., Teevan, J. (eds.) Query Log Analysis: Social And Technological Challenges. A workshop at WWW'2007 (2007)
20. Qin, Y.p., Wang, X.k.: Study on multi-label text classification based on svm. In: FSKD '09 - Volume 01. pp. 300–304. IEEE Computer Society (2009)
21. Rafiei, D., Bharat, K., Shukla, A.: Diversifying web search results. In: WWW'10. pp. 781–790. ACM (2010)
22. Rose, D.E., Levinson, D.: Understanding user goals in web search. In: WWW '04. pp. 13–19. ACM, New York, NY, USA (2004)
23. Shen, D., Sun, J.T., Yang, Q., Chen, Z.: Building bridges for web query classification. In: SIGIR'06. pp. 131–138. ACM (2006)
24. Teevan, J., Dumais, S.T., Liebling, D.J.: To personalize or not to personalize: modeling queries with variation in user intent. In: SIGIR'08. pp. 163–170. ACM (2008)
25. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. Int J Data Warehousing and Mining 2007, 1–13 (2007)
26. Veilumuthu, A., Ramachandran, P.: Intent based clustering of search engine query log. In: CASE'09. pp. 647–652. IEEE Press (2009)
27. Yuan, X., Dou, Z., Zhang, L., Liu, F.: Automatic user goals identification based on anchor text and click-through data. Wuhan University Journal of Natural Sciences 13, 495–500 (2008)