Computer-supported Interactive Assignment of Keywords for Literature Collections

Shivam Agarwal*
University of Duisburg-Essen, Germany

Jürgen Bernard†
TU Darmstadt, Germany

Fabian Beck‡
University of Duisburg-Essen, Germany

ABSTRACT

A curated literature collection on a specific topic helps researchers to find relevant articles quickly. Assigning multiple keywords to each article is one of the techniques to structure such a collection. But it is challenging to assign all the keywords consistently without any gaps or ambiguities. We propose to support the user with a machine learning technique that suggests keywords for articles in a literature collection browser. We provide visual explanations to make the keyword suggestions transparent. The suggestions are based on previous keyword assignments. The machine learning technique learns on the fly from the interactive assignments of the user. We seamlessly integrate the proposed technique in an existing literature collection browser and investigate various usage scenarios through an early prototype.

Index Terms: Human-centered computing—Visualization systems and tools; Computing methodologies—Supervised learning by classification

1 INTRODUCTION

Researchers often curate their personalized literature collections based on their research interests and area of work. Also, they create topic-specific collections for a project or literature survey, which they share with other researchers. A literature collection can be organized by assigning multiple keywords to each article. However, achieving a good quality of keyword assignment in a literature collection is not an easy task and requires much effort. Two specific problems have to be addressed to ensure good quality of keyword assignment. The first problem is to achieve *completeness* where each keyword must be assigned to all the relevant publications. The second problem is maintaining *consistency* where the concept behind a keyword needs to be represented by only one unique keyword.

Multi-label classification is a well-established area of research which deals with the problem of assigning multiple labels to individual data samples. Various techniques have been invented to solve this problem of classification. Although these techniques have high efficiency, they suffer from two common problems. First, it is difficult to understand the complex working details of these techniques, specially for non-expert users. This reduces the trust of users on the classification results. Second, it generally requires large amount of data for training. We adopt an interactive approach and use a multi-label classification technique to suggest keywords for publications in a literature collection. We hide the working details of the algorithm and make the system transparent. We do this by incorporating different visual and interactive methods to explain the output of the multi-label classification technique. By using the interactive approach in the limited context of keyword assignment, we get meaningful results with comparatively fewer data samples.

*e-mail: shivam.agarwal@paluno-uni-due.de
†e-mail: juergen.bernard@gris.tu-darmstadt.de
‡e-mail: fabian.beck@paluno-uni-due.de

SurVis [5] is a visualization system for literature collections which supports the manual assignment of user-defined keywords to publications in the collection. However, keyword assignment in the system faces the same problems discussed above. In this paper, we extend the system and address these problems through visual interactive keyword assignment with suggestions from machine learning models. In the traditional (model-centered) machine learning process, the strength of human involvement is not exploited. Even if a machine learning model performs well in general, human judgement for individual instances would be beneficial to improve its performance. Hence, we involve users in the keyword assignment process and use a supervised machine learning technique that learns from previous keyword assignments.

As shown in Figure 1, the workflow of the proposed technique starts with training machine learning models from previous keyword assignments of publications in a literature collection. The trained models are then used to suggest keywords for an existing/new publication in the collection. In addition to the trained models, a text processing technique is also used to suggest new keywords which are extracted from title and abstract of the publication.A visual explanation helps in justifying the keyword suggestions. To make them more transparent, we use different models to suggest keywords based on different features of input data. The major contributions are the following:

- We instantiate the visual-interactive labeling *(VIAL)* process [7] for a multi-labeling scenario with few modifications (Section 3).
- We propose consistent interactions and visual explanations for exploiting machine learning in *SurVis* [5], which extends the current workflow of its users (Section 5).
- We investigate four usage scenarios in the process of organizing a literature collection with the help of an early prototype implementation based on *SurVis* (Section 6).

2 RELATED WORK

The keyword assignment of publications in a literature collection can be formulated as a machine learning problem of multi-label classification. Many machine learning algorithms exist which performs the task with high efficiency [23, 25]. These algorithms have applications in various fields [11] such as in image/video annotation [18], assigning emotion tags to music [22], and classification of text [16].

Active learning [21] is a machine learning technique that involves users for training the machine learning model. In this approach, unlabeled data instances are sampled and presented to a user for assigning appropriate labels. Various studies have shown the usage of active learning to perform multi-labeling task [8, 14, 15, 24]. In our problem scenario, we also need to assign multiple labels to data samples. But we also want users to always stay in control of the assignment process. This is possible because we work with lesser amount of data samples and we want to always ensure good quality of keyword assignment. Hence, we refrain from using any fully automated approach of keyword assignment. We do not make use

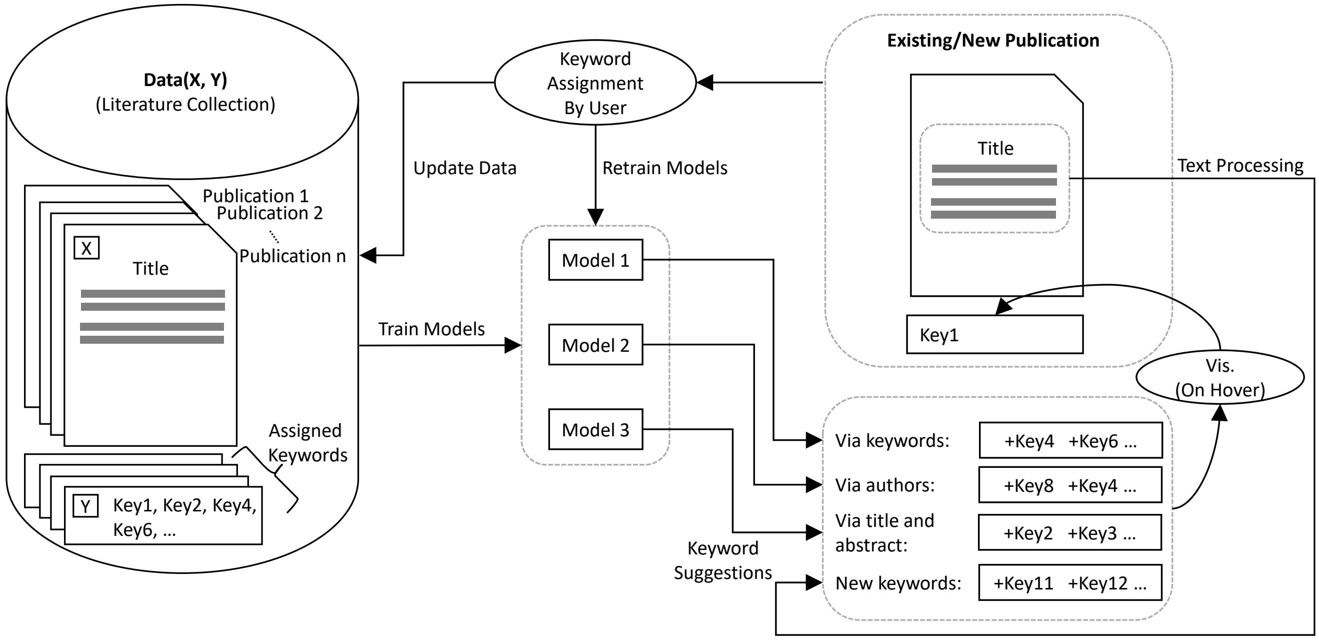


Figure 1: The suggestion process of the proposed computer-supported multiple keyword assignment technique for literature collections.

of a traditional active learning approach but conflate the principles with visual-interactive labeling (*VIAL*) [7].

Lack of explanation of results in machine learning techniques affects the trustworthiness of the systems built using them. Some studies have provided guidelines on designing interfaces for recommender systems [17] and presence of explanation components [12] to increase trust of users on these intelligent systems. Some visual interfaces help users to make decisions, for instance, by showing the state of the machine learning model and the feature space through scatter plots with dimensionality reduction [6, 13, 20]. Such interactive learning systems result in better experience, increased trust, and higher effectiveness [2]. We adopt the principle of involving users for interactive keyword assignment supported by suggestions from machine learning.

3 VIAL FOR MULTI-LABELING

Visual-interactive labeling *(VIAL)* [7] is a generalized process that unifies machine learning and visual interactive approaches for the task of labeling. The process focuses on achieving three goals: labeled data, trained models, and knowledge of labeling process. The process fits our scenario of assigning keywords to publications in a literature collection. We instantiate the *VIAL* process with few modifications in the process and implement them in a prototype based on *SurVis* [5] system.

In our case, publications in a literature collection are the data instances where multiple keywords can be assigned to each of them. The operations performed in *preprocessing and feature extraction* step, as proposed in the baseline VIAL, should ensure compatibility of the input data with models. We use title, abstract, authors, assigned keywords, year, and venue of publications. The *learning model* step in VIAL process involves training a model based on the extracted data. The process also include a feedback loop to retrain the models with every keyword assignment in a publication. The details of learning algorithm is presented in Section 5. To reflect transparency of suggestions in our implementation, we use different models based on different features of the publication data. We instantiate the *result visualization* step by light-weight visualization to explain the suggestions with a customized metric (Section 5).

We do not explain working details of the learning algorithm, rather provide visual explanation of the suggested keywords. A *labeling interface* is implemented and integrated into *SurVis* with the ability to assign multiple keywords to each publication in the collection, as shown in Figure 4. *Feedback interpretation* involves updating the literature collection and machine learning models with every keyword assignment.

4 INTRODUCTION TO *SurVis*

SurVis [5] is an interactive visual analytics system for browsing literature collections. Its interface is divided vertically into two parts, as shown in Figure 2. The left area includes visual components to show temporal development, word clouds to show assigned keywords, authors, and publication series. The right part shows a list of publications, which is filtered and sorted according to the selected parameters called *selectors*. Each record in the system is a publication and displayed with title, abstract, authors, and assigned keywords. The header shows current selectors, which are color-encoded. The footer shows advanced features including *add new entries*, *download BibTex*, *rename keyword*, etc.

Every entry in the word clouds in the left region is clickable. Every click creates a selector which is used to sort the list of publications. The selectors can be applied on any keyword, author, year, publication series, and search query (it is also treated as a selector). Each selector has a different color, which helps in keeping track of applied selectors on filtered results. The colored selectors provide an easy and powerful interaction to search, explore, and analyze the literature collection. Small vertical bars ( ) are attached to a selected entry in the word clouds, temporal bar, and publication on the right. They encode the strength of agreement with applied selectors. Advantages of the system include better dissemination of the collection and reproducible literature analysis.

The system has two types of users: A *curator*, who organizes and analyzes the literature collection and a *reader*, who browses and substructures the collection, and tries to find publications of interest. The workflow of both users involve interactions and visual feedback from the system. A *curator* is responsible for curating the collection, which includes maintaining *completeness* and *consistency* of

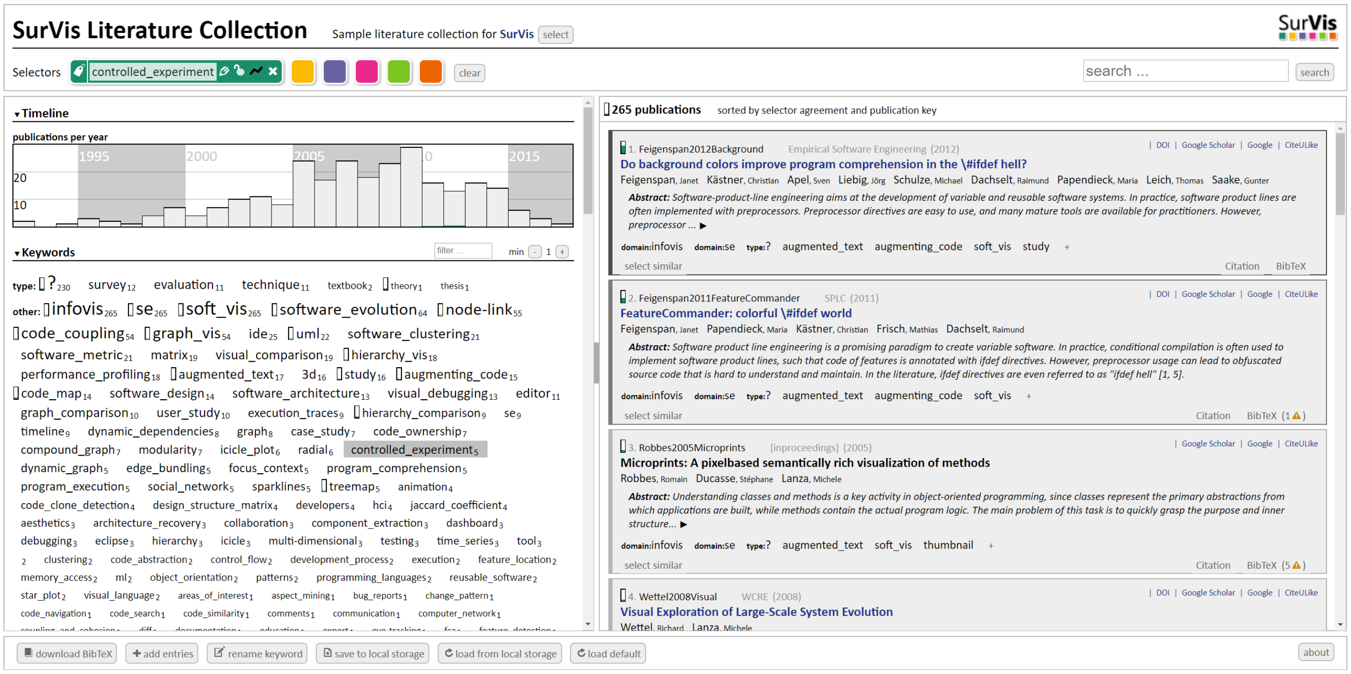


Figure 2: Screenshot of the *SurVis* system with machine learning extension of keyword selector. The keyword *controlled_experiment* is selected and its prediction functionality is switched on ( ).

keyword assignments. To maintain these criteria a *curator* has to regularly update the keywords of existing and newly added publications, which may also involve the introduction of new keywords to the collection. With our proposed technique we target the *curator* of a literature collection.

5 Interactive Labeling Approach

Keyword assignment is an important step while curating a literature collection in *SurVis*. The assigned keywords give the collection a structure, which is crucial during its exploration. A curated literature collection helps *readers* by returning relevant publications quickly. But a good quality of keyword assignment is a difficult goal to achieve while curating a collection. The magnitude and impact of the problem increase with an increasing number of publications and keywords. It becomes difficult to keep track of all keywords.

5.1 Machine Learning Approach

As shown in Figure 1, we train machine learning models, which help the *curator* by suggesting keywords for publications. We do not assume that the *curator* has previous knowledge of machine learning. To make the suggestion trustworthy, we need suggestions to be transparent. The visual interface of *SurVis* is already dense and leaves very little room for additional contents. Hence, we present the keyword suggestions and the visual explanations for every publication concisely and show them only on demand.

$K = \{k_1, k_2, ..., k_n\}$ represents the set of keywords, $X = \{x_1, x_2, ..., x_m\}$ the set of publications in a literature collection. Each publication has several features which are modeled as a sequence, i.e. $x_i =< Title, Abstract, Author(s) >$. The set $Y = \{y_1, y_2, ..., y_m\}$ denotes the assigned keywords for every publication, where y_j is a set of keywords from set K, assigned to publication x_j. For example, $y_2 = \{k_1, k_4\}$ should be inferred as two keywords, k_1 and k_4, were assigned to the publication x_2.

We use a multiclass multi-label algorithm implementation[1] as the machine learning technique to suggest keywords. The strategy involves training a single classifier per keyword. Publications that have the assigned keyword are treated as positive samples for training the keyword's classifier while the rest of the publications forms the negative samples. Every classifier uses a support vector machine (SVM) and produces a real-valued confidence score for its decision. The approach is also known as *one-vs-all* or *one-vs-the-rest* (OvR).

To make the suggestions transparent, we train three different machine learning models, as shown in Figure 1 and implement a customized metric for visual explanation of the suggested keywords. Each publication x_i contains a sequence of title, abstract, and authors. It also has a set of assigned keywords, represented by y_i. The three models are trained with different features in their training data as shown in Table 1. In addition to the three machine learning models, we include a simple text processing technique to extract keywords from title and abstract of the publication. It suggests only those extracted keywords which are not present in the literature collection. The technique is useful to introduce new keywords in a collection.

Table 1: Machine learning models used for suggesting keywords.

Model	Name	Features in Training Data
1	*Via keywords*	Previously assigned keywords of publications
2	*Via authors*	Authors of publications
3	*Via title & abstract*	Text from title and abstract of publications

We define a function to calculate the relation between a *suggested keyword* and an *assigned keyword* of the publication in Equation 1. Using this function, metric values are calculated between every keyword in the suggestions and every keyword already assigned to the publication.

$$f(suggested\ keyword, assigned\ keyword) = |S \cap A| / |A| \qquad (1)$$

S and A are the sets of those publications which were previously

[1] from scikit-learn: sklearn.multiclass.OneVsRestClassifier()

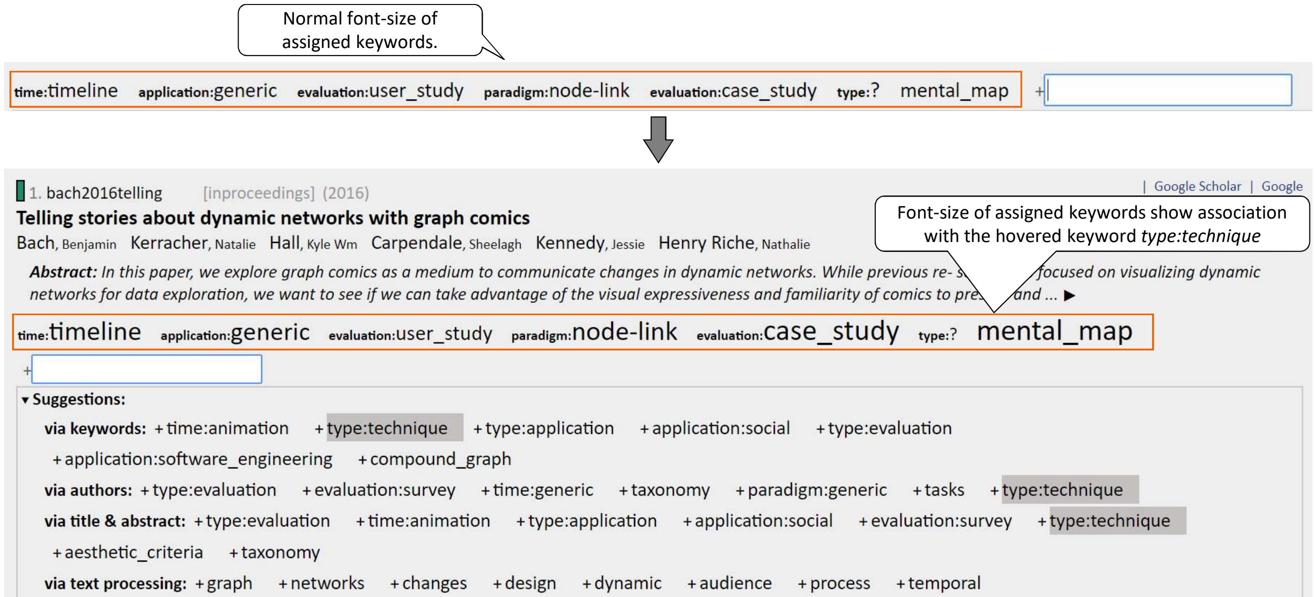


Figure 3: Visual explanation of the relation between the suggested keyword *type:technique* and the assigned keywords of a publication. Metric values are encoded in font size of the assigned keywords and shown on hovering over a suggested keyword. The hovered keyword *type:technique* is suggested by three different models.

assigned with the *suggested keyword* and *assigned keyword* respectively. The intuition behind the metric is that two keywords have high association with each other when they are assigned to the same publications. The metric is independent of the machine learning technique and acts as a good indicator of keyword suggestions.

5.2 Publication-centric Keyword Assignment

Publication-centric perspective of keyword assignment includes those usage scenarios in which the focus of a *curator* is on assigning keywords to an individual publication. The selection of the publication depends on the curator. It could be a new or an existing publication in the literature collection.

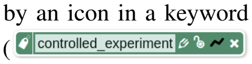
We show suggested keywords only when the *curator* wants to assign or update keywords of a publication. A button with a plus symbol under the abstract of each publication provides the on demand keyword assignment functionality. As depicted in Figure 1, the suggested keywords from different models are shown separately.

The computed metric value is visually encoded in the font size of all *assigned keywords* of the publication. The font size changes when a pointer is hovered over a *suggested keyword*, as depicted by *Vis. (On Hover)* step in Figure 1. Hovering also highlights the presence of the same suggested keyword in the list of suggestions from other models. The result of this interaction is shown by an example in Figure 3. Suggested keyword can be assigned to the publication by clicking on its plus symbol. Every keyword assignment retrains the machine learning models.

The combination of keyword suggestion through different models and the visual explanation supports transparency and increases the trust of users on these suggestions. This perspective of keyword assignment helps users in assigning fitting keywords to a publication and maintains the *consistency* criterion of the keyword.

5.3 Keyword-centric Keyword Assignment

The keyword-centric perspective is focused on maintaining the *completeness* criterion of keyword assignment. This means that *curator* should be able to find those publications to which an existing keyword should be also assigned. The focus of the *curator* is on an existing keyword rather than the publications.

The requirements of this perspective demand a sorted list of publications of potential candidates for assignment of selected keyword. We implement and integrate this functionality of prediction in keyword selector of the *SurVis* system. It is indicated by an icon in a keyword selector and highlighted in black color (

). The sorting of publications is done on prediction values which are also visualized as colored vertical bars along with the list of publications, as shown in Figure 2.

Interactions used for enabling this functionality integrate well with the existing interactions. An example of this perspective is shown in Figure 2, where the *controlled_experiment* keyword was selected and then the prediction activated.

6 Usage Scenarios

We investigate four usage scenarios in the process of curating a literature collection. These usage scenarios address *curator* of a literature collection. Two authors of the paper were involved in the investigation. The literature collections used are centered on two different themes and have different quality levels of assigned keywords. The first collection has dynamic graph visualization as a central theme (LC1) and is in an already well-curated state, while the second collection is about visualization and software engineering (LC2). The keyword assignment in the second collection is of lower quality as it has gaps in maintaining *completeness* and *consistency* criteria.

6.1 Scenario 1: Adding a Publication

This is a publication-centric scenario where we use the LC1 collection, which was built by Beck *et al.* [4] for a state-of-the-art report on dynamic graph visualization. We add a paper by Bach *et al.* [3] to the collection and document the process step by step in Figure 4. Initially, there are no keywords assigned to the added publication, hence, the list of suggestions *via keywords* is empty, as shown in Figure 4. The paper is comparatively new and that is why it has not already been present in the collection. The publication is about showing temporal changes through graph comics. The

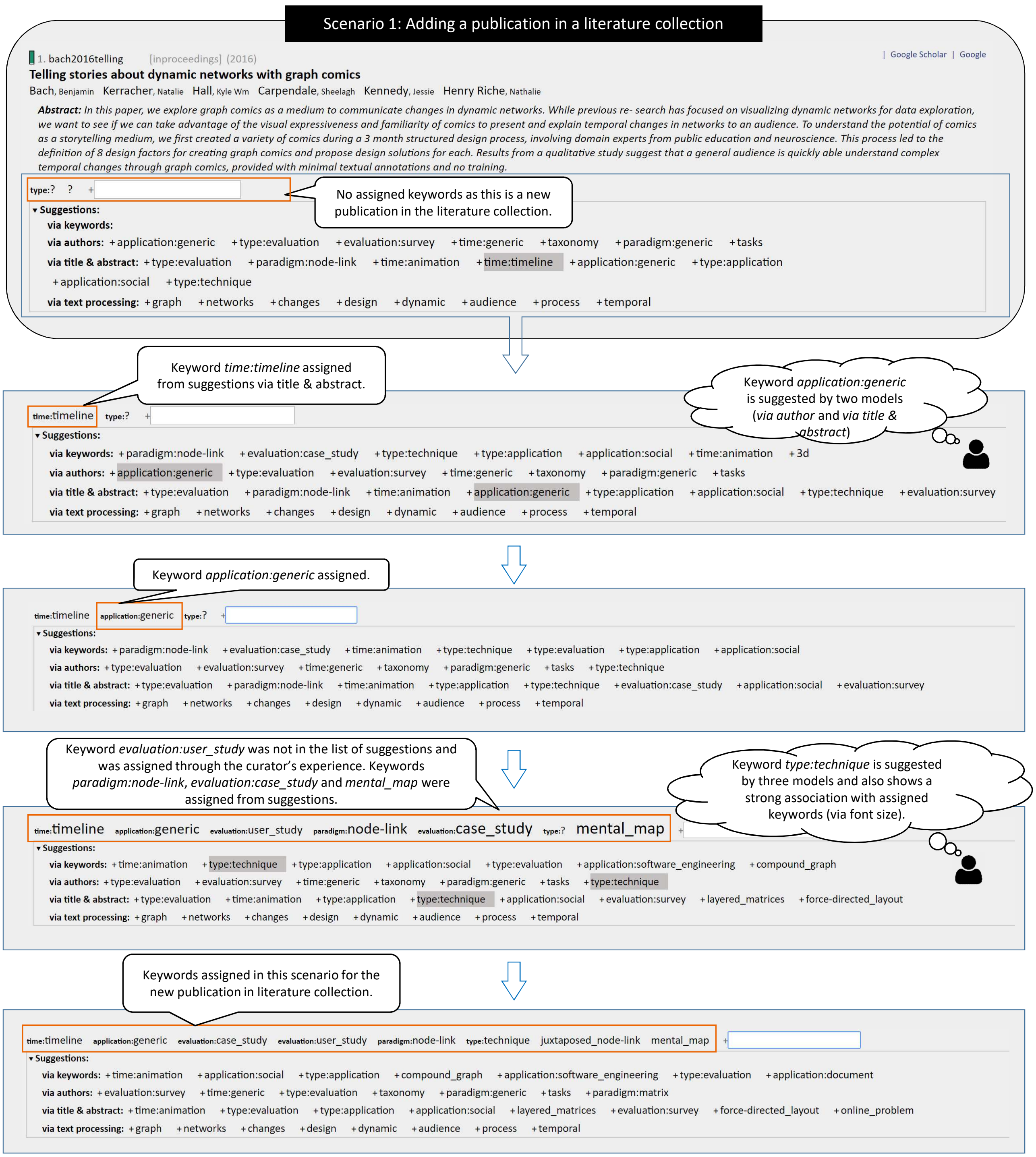


Figure 4: An example of adding a publication to a literature collection. It shows the process of interactive assignment of multiple keywords (Scenario 1).

keyword *time:timeline* describes a concept related to comic strip and is shown in the list of suggestions *via title & abstract*. Hence, we assign it to the publication. Keywords *application:generic* and *evaluation:survey* are suggested by two approaches (*via authors* and *via title & abstract*). The keyword *paradigm:node-link* is also suggested by two approaches (*via keywords* and *via title & abstract*), which acts as good indicators and we assign them to the publication.

The new publication reports on a user study to evaluate the approach of graph comics. At first, the suggested keyword *type:evaluation* seems to be a good candidate for assignment. But we realize that the keyword is only used in the literature collection if the evaluation is the focus of a publication. Another keyword (*evaluation:user_study*) exists in the collection, which is a better candidate for the assignment. The keyword *time:animation* is also not a fitting suggestion as the added publication does not contribute in the area of animation. Keywords *evaluation:user_study* and *juxtaposed_node-link* were not present in the list of suggestions and were assigned to the publication from the experience of the *curator*. On hovering over a suggested keyword we see our customized metric encoding through font size of assigned keywords. Figure 4 shows this encoding for suggested keyword *type:technique*. On hovering over suggested keyword *type:technique*, we understand that it is strongly associated with keywords *mental_map* and *evaluation:case_study* and least with *evaluation:user_study*. This makes the suggestion more transparent.

We add another publication [9] to the collection, which was published recently. The publication is a typical example of the theme of the collection and hence the keyword suggestions helped a lot in the assignment. We went through the abstract and skimmed through the publication, which proved the relevance of suggested keywords. We thought of assigning *software_evolution* and *software_execution* keywords to the publication while reading the abstract but forgot to do so. Later, we got reminded while going through the list of keyword suggestions and then assigned them to the publication.

6.2 Scenario 2: Updating Keywords of a Publication

This is a publication-centric scenario where the focus is on updating keywords of a particular publication in a collection as illustrated in Figure 5. We use the LC2 collection for this scenario which focuses on software visualization. Some keywords like *graph_vis* and *modularity* were added to the collection in later stages of its curation. The publications are not fully updated with such keywords. This introduced gaps in keyword assignment of this collection. We discuss two examples in this scenario.

For the first example, we select the year 2008 and choose to update keywords of a publication by Abdeen *et al.* [1], which has been added long time ago to this literature collection. On hovering over suggested keyword *node-link*, as shown in Figure 5, we see a strong association with the assigned keyword *code_coupling* denoted by its large font size. After going through the abstract and images of the publication, we observe that it models package references as directed node-link graphs. It shows that the keyword *node-link* is a good suggestion and we assign it to the publication.

Keywords *modularity* and *software_architecture* are not present in the list of suggestions. We assign them to the publication based on the experience of the *curator* of the literature collection. We observe that it is hard for *curator* to remember all the keywords and their concepts present in a literature collection. To find these keywords, the *curator* has to go through the word cloud of keywords.

Keyword *graph_vis* shows strong association with *node-link* and *software_architecture* keywords. The publication employs graph visualization in terms of matrices, hence we assigned the keyword to the publication. Also, the *software_metric* keyword is assigned to the publication from suggestions.

For the second example, we select a publication by Cheng *et al.* [10] and update it by assigning *graph_vis* and *node-link* keywords from the list of suggested keywords, as shown in Figure 5. The figure also shows a strong association of keyword *node-link* to already assigned keywords *3d* and *soft_vis*. The keyword is proposed by two different models which supported the decision of assigning it to the publication.

Keyword *visual_comparison* is suggested by more than one technique, due to strong association with already assigned keyword *visual_debugging*, which indicate that usually both of them are assigned together to the same publications. But the publication has no concept related to the keyword *visual_comparison*. Hence, we do not assign it to the publication.

6.3 Scenario 3: Updating Publications with a Keyword

This is a keyword-centric scenario where focus is on a new or an existing keyword. In this scenario, we discuss two situations. First, where a new keyword is introduced in a literature collection and second where an existing keyword of a collection is used to find publications which are potential candidates for assignment of the keyword. This scenario is most useful to ensure the *completeness* criterion of selected keyword.

We add a new keyword *controlled_experiment* in LC2 by assigning it to a publication by Ricca *et al.* [19]. We manually search for a few publications that are good candidates for assigning the new keyword and update them. Then, we choose the keyword as a selector and switch on the prediction functionality (controlled_experiment). It sorts the list of publications using prediction value, as shown in Figure 2. We assigned the keyword to a few other publications using this functionality.

For the second situation, we pick an existing keyword, *performance_profiling*, in the same literature collection. We select the keyword and switch on the prediction feature. Browsing through the sorted list, we update a few publications by assigning the selected keyword. All the publications in the sorted list were not updated with the selected keyword and we had to go through title and abstract to make the final decision. This situation helped in maintaining the *completeness* criterion. The situation is very common while curating a literature collection. Keywords get introduced at different stages of the curation process. It becomes very difficult for a curator to remember previous keywords and publications. This introduces gaps in the assignment of such keywords. The keyword *performance_profiling* was one such example.

6.4 Scenario 4: Building a New Literature Collection

This scenario assumes that there is no existing literature collection of publication with already assigned keywords to start with. This is useful when a user starts collecting publications, which could happen in various ways. Examples includes PhD students building a literature collection in the early stages of their research career and researchers building a literature collection for writing state-of-the-art reports.

We started building a literature collection with a central theme of visualization and deep learning. There were no keyword suggestions at the start. The introduction of keywords for the first few publications was easy because there were not many keywords to remember. The keywords suggested *via text processing* were new keywords which were also useful.

We observed that introduction of new keywords for the assignment was easy at first, but soon became ambiguous and difficult. The ambiguity originated in the overlap of the concepts associated with the keywords. Closely related keywords often represent the same concept but at different levels of granularity. This required human intervention to maintain the quality of keyword assignments in the collection. We also observed that after adding and updating more publications with keywords, the suggestions begin to be more meaningful.

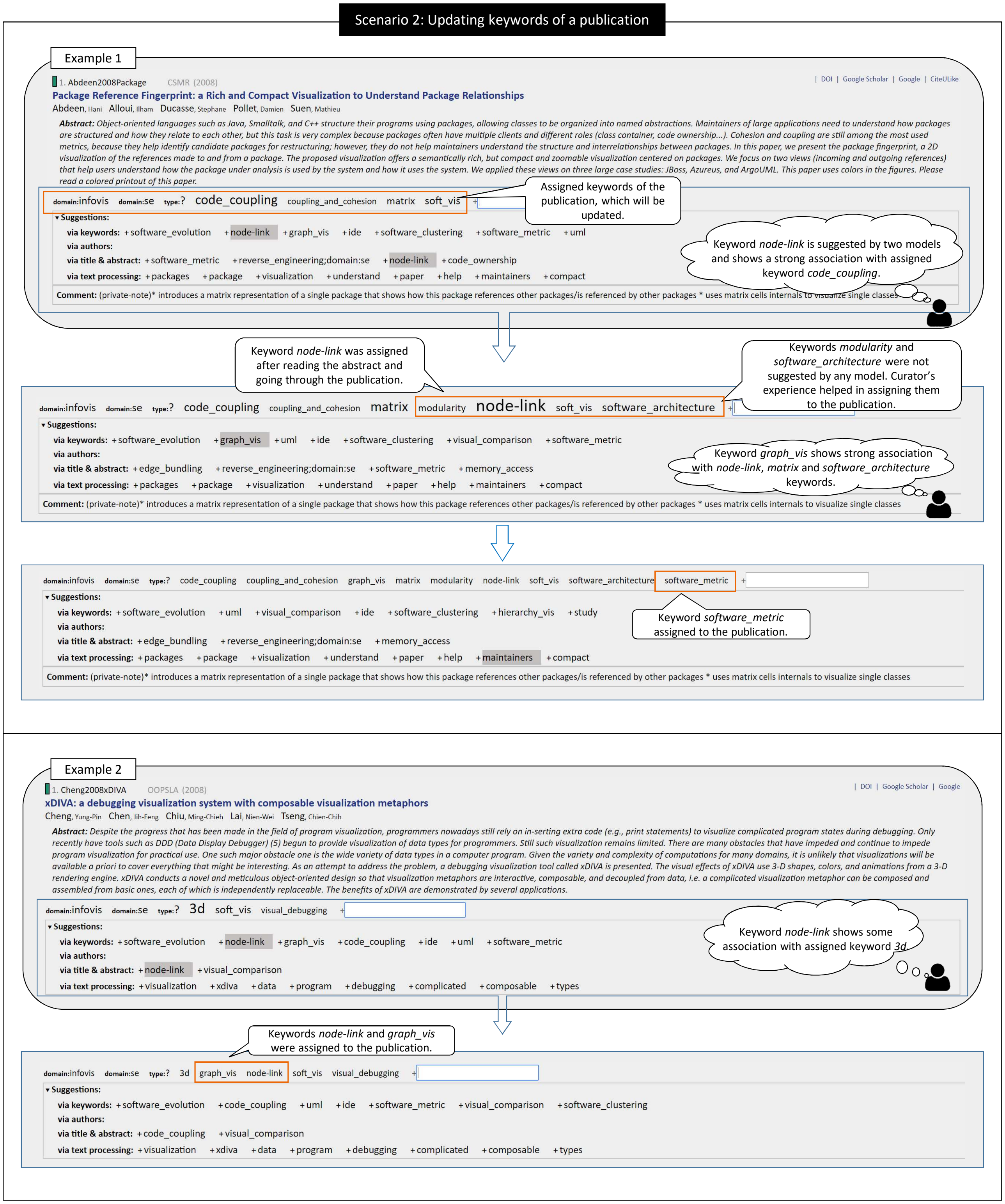


Figure 5: The two examples show publications in literature collection LC2 being updated with keywords (Scenario 2).

We were able to build a new literature collection with 25 publications in it. We faced problems in introduction of accurate and relevant keywords which could be assigned to the publications. The suggestions *via text processing* helped in building the initial set of keywords for the collection.

7 Discussion and Future Work

The proposed computer-supported process helped in assigning keywords to publications in a literature collection. The usage scenarios indicate its usefulness and drawbacks.

7.1 Lessons Learned

The suggestions helped in updating publications by reducing the need to remember all the keywords in a literature collection. They also helped us to remember the keywords that we decided to assign but forgot while going through the publication in Scenario 2. We observed that the suggestions become less meaningful if the publication introduces novel ideas.

The introduction of a new keyword was easy but maintaining its *completeness* criterion was difficult to achieve. The prediction functionality in keyword selectors helped in finding publications that are good candidates for assigning the selected keyword.

We observed that the quality of suggestions improves with an increase in the number of assignments for every keyword. The behavior can be explained by the nature of machine learning.

Visual explanations of suggested keywords helped in understanding why they were being suggested. The presence of a keyword in the suggestions of more than one model indicated good chances for its assignment to a publication. It guided us towards further investigation for assigning it to a publication. It also helped in increasing transparency and gaining trust on the suggestions.

User involvement was crucial for correct assignments. We saw few difficult instances where the assignment of keywords was ambiguous and had to rely on the user for the final decision. We learned that the suggestions were most helpful when the literature collection was curated and the publication to be updated was similar to others. User interactions demonstrated their usefulness to investigate the suggested keywords, which lead to an increase in the quality of suggestions after some assignments. We did not witness any noticeable delay while re-training models on every keyword assignment. It may be due to the small size of the literature collections used, but SurVis is designed specifically for such small collections.

7.2 Limitations and Future Work

Although the suggestions helped by providing relevant keywords, they were initially not very meaningful for building a new literature collection. We observed that few relevant keywords were missing from the list of suggested keywords and the curator had to assign them manually. It showcases a general limitation of machine learning.

With more publications in a literature collection, the re-training of models could introduce a noticeable delay which could impact the workflow of the user. One possible solution could be to re-train the models in batches of keyword assignments.

We felt the need to cross out the suggested keywords that are not fit for assignment to a publication. This interaction could also train the machine learning model with negative instance and would increase the quality of suggestions.

We implemented a metric to explain the keyword suggestions. However, the metric is simple and could face difficulty in explaining suggestions from complex machine learning techniques. Future work could include improvement of the metric.

8 Conclusion

We instantiated VIAL process for assigning multiple keywords to publications in a literature collection. We used a machine learning technique to suggest keywords during the assignment and make the suggestions transparent. We investigated the usefulness of the proposed techniques in a working prototype though different usage scenarios. The scenarios were realistic and the suggestions helped in each one of them. Few drawbacks were also discovered and mentioned with possible solutions as future work.

Acknowledgments

Fabian Beck is indebted to the Baden-Württemberg Stiftung for the financial support of this research project within the Postdoctoral Fellowship for Leading Early Career Researchers.

References

[1] H. Abdeen, I. Alloui, S. Ducasse, D. Pollet, and M. Suen. Package Reference Fingerprint: a rich and compact visualization to understand package relationships. In *Proceedings of the 12th European Conference on Software Maintenance and Reengineering*, pp. 213–222. IEEE Computer Society, 2008.

[2] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[3] B. Bach, N. Kerracher, K. W. Hall, S. Carpendale, J. Kennedy, and N. Henry Riche. Telling stories about dynamic networks with graph comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3670–3682. ACM, 2016.

[4] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. *EuroVis STAR*, 2:1–21, 2014.

[5] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with SurVis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, 2016.

[6] M. Behrisch, F. Korkmaz, L. Shao, and T. Schreck. Feedback-driven interactive exploration of large multidimensional data supported by visual classifier. In *Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology*, pp. 43–52. IEEE, 2014.

[7] J. Bernard, M. Zeppelzauer, M. Sedlmair, and W. Aigner. VIAL: a unified process for visual interactive labeling. *The Visual Computer*, 34(9):1189–1207, 2017.

[8] K. Brinker. On active learning in multi-label classification. In M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nürnberger, and W. Gaul, eds., *From Data and Information Analysis to Knowledge Engineering*, pp. 206–213. Springer, 2006.

[9] M. Burch. Property-driven dynamic call graph exploration. In *Proceedings of the 11th International Symposium on Visual Information Communication and Interaction*, pp. 72–79. ACM, 2018.

[10] Y.-P. Cheng, J.-F. Chen, M.-C. Chiu, N.-W. Lai, and C.-C. Tseng. xDIVA: a debugging visualization system with composable visualization metaphors. In *Companion to the 23rd ACM SIGPLAN Conference on Object-Oriented Programming Systems Languages and Applications*, pp. 807–810. ACM, 2008.

[11] E. Gibaja and S. Ventura. Multi-label learning: a review of the state of the art and ongoing research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(6):411–444, 2014.

[12] A. Glass, D. L. McGuinness, and M. Wolverton. Toward establishing trust in adaptive agents. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, pp. 227–236. ACM, 2008.

[13] L. Huang, S. Matwin, E. J. de Carvalho, and R. Minghim. Active learning with visualization for text data. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, pp. 69–74. ACM, 2017.

[14] A. J. Joshi, F. Porikli, and N. Papanikolopoulos. Multi-class active learning for image classification. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2372–2379, 2009.

[15] X. Li and Y. Guo. Active learning with multi-label SVM classification. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 1479–1485. AAAI Press, 2013.

[16] J. Nam, J. Kim, E. L. Mencía, I. Gurevych, and J. Fürnkranz. Large-scale multi-label text classification—revisiting neural networks. In

Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 437–452. Springer, 2014.

[17] P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems*, 20(6):542–556, 2007.

[18] G.-J. Qi, X.-S. Hua, Y. Rui, J. Tang, T. Mei, and H.-J. Zhang. Correlative multi-label video annotation. In *Proceedings of the 15th ACM International Conference on Multimedia*, pp. 17–26. ACM, 2007.

[19] F. Ricca, M. Di Penta, M. Torchiano, P. Tonella, and M. Ceccato. The role of experience and ability in comprehension tasks supported by UML stereotypes. In *Proceedings of the 29th International Conference on Software Engineering*, pp. 375–384. IEEE Computer Society, 2007.

[20] M. Sedlmair and M. Aupetit. Data-driven evaluation of visual quality measures. *Computer Graphics Forum*, 34(3):201–210, 2015.

[21] B. Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.

[22] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas. Multi-label classification of music into emotions. In *Proceedings of the 9th International Conference of Music Information Retrieval*, pp. 325–330, 2008.

[23] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3(3):1–13, 2007.

[24] B. Yang, J.-T. Sun, T. Wang, and Z. Chen. Effective multi-label active learning for text classification. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 917–926. ACM, 2009.

[25] M.-L. Zhang and Z.-H. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2014.