BTU DBIS' Plant Identification Runs at ImageCLEF 2012

Thomas Böttcher, Christoph Schmidt, David Zellhöfer, and Ingo Schmitt

Brandenburg Technical University, Database and Information Systems Group,
Walther-Pauer-Str. 1, 03046 Cottbus
`tboettcher|christoph.schmidt|david.zellhoefer|schmitt@tu-cottbus.de`

Abstract. In this work, we summarize the results of our first participation in the plant identification task.
Unlike other contributors, we present a rather untypical approach, which does not rely on classification techniques. In contrast, logical combinations of low-level features expressed in a query language are used to assess a document's similarity to a species. Similar to ImageCLEF 2011, DBIS' approach is based on the commuting quantum query language (CQQL). CQQL was proposed by the workgroup to combine similarity predicates as found in information retrieval and relational predicates common in databases. In order to combine both predicate types, CQQL utilizes the mathematical formalisms of quantum mechanics and logic eventually forming a probabilistic logic.
To test the utility of our query language, three different automatic approaches are discussed. First, a query by example approach towards plant identification is presented. Second, the approach is combined with a k-medoid technique to exploit relationships within the top-k results. To conclude with, the aforementioned techniques are compared with the utilization of the k-medoid method alone.
With respect to the non-existent experience with the task, the results of the discussed approach are fairly decent but leave room for improvement being outlined as future work.

Keywords: Content-Based Image Retrieval, Clustering, Experiments

1 Introduction

In this paper we present the results of the Database and Information Systems Group's (DBIS) participation in the plant identification task that was organized within ImageCLEF 2012. Because it is our first try with the Pl@ntLeaves data set, our main objective was to gain experience with the data set and to investigate future directions of research. Unlike other contributors, we present a rather untypical approach, which does not rely on classification techniques. In contrast, logical combinations of low-level features expressed in a query language are used to assess a document's similarity to a species.

Similar to ImageCLEF 2011 [15], DBIS' approach is based on the commuting quantum query language (CQQL) [8]. CQQL was proposed by the workgroup

to combine similarity predicates as found in information retrieval (IR) and relational predicates common in databases (DB). In order to combine both predicate types, CQQL utilizes the mathematical formalisms of quantum mechanics and logic eventually forming a probabilistic logic [5]. For the sake of brevity, the theory of CQQL is not covered in this paper. Instead, its relation to probabilistic and other quantum mechanics-derived IR models is covered in [16], while [9] discriminates it from fuzzy logic [11].

The core idea of the presented approach can be summarized as follows. Based on different document representations, e.g. color-based low-level features or accompanying metadata such as GPS information, a logical CQQL query is formulated, e.g. to define the species' characteristics. Based on the CQQL evaluation rules describes in prior work [8] and summarized in Section 2, the query is transformed into an arithmetic formula which is then used to calculate the similarity between the documents in the data set.

1.1 Plant identification Task

The plant identification task is part of ImageCLEF for the second time. It is focused on tree species identification based on leaf images. In comparison to the last year's challenge there are some novelties. The number of species has been increased from 70 to 126. Furthermore the main objective changed from pure classification to a plant species retrieval task. In the following section we describe the basic characteristics of the data and their identification as far as it is needed for the understanding of this paper. The complete description of the plant identification task 2012 be found in [2].

Training and Test Data The plant identification task is based on the Pl@ntLeaves data set [6], which is divided into training and test data. The training subset was built by including the training and test subsets of last year's Pl@ntLeaves data set, and by randomly selecting 2/3 of the individual plants [3]. The complete data set contains 11,572 pictures, 126 tree species mainly from the French Mediterranean area, subdivided into 3 different kinds of pictures: scans, scan-like photos and natural photos [3]. The distribution of training and test data of the Pl@ntLeaves data set is shown in Table 1.

Table 1. Distribution of the Pl@ntLeaves data set

	Scan	Scan-like	Photograph	$\sum$
Train data	4,870	1,819	1,733	8,422
Test data	1,760	907	483	3,150
Complete data set	6,630	2,726	2,216	11,572

Identification and evaluation The goal of this task is to identify the tree species, whose leaf is depicted on a given test image. In consequence, for each

test image, a prediction should be made for each of the 126 plant species. For the scope of the task, a prediction is a score between 0 and 1 expressing the confidence that a given sample images belongs to a given species.

To evaluate the prediction quality, the task organizers calculated a score which is related to the rank of the correct species in the result list. Thereby a mean value is built per author and per plant which is in the collection. An author is a person which helped to built up the Pl@ntLeaves collection. The score is defined in the following formula [3]:

$$S = \frac{1}{U} \sum_{u=1}^{U} \frac{1}{P} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} s_{u,p,n} \tag{1}$$

U : number of users (who have at least one image in the test data)
P_u : number of individual plants observed by the u-th user
$N_{u,p}$: number of pictures taken from the p-th plant observed by the u-th user
$s_{u,p,n}$: classification score (1 or 0) for the n-th picture taken from the p-th plant observed by the u-th user

2 Retrieval Model

Repeating our summary found in prior work [15], CQQL can be considered a query language dealing with probabilities that is consistent with the laws of the Boolean algebra. The probabilities denote how "similar" a document is to a query regarding a given condition in the query, e.g. a color histogram. In the next section, we will sketch the arithmetic evaluation of CQQL as it is necessary for the understanding of this paper.

2.1 Evaluation of CQQL

Given that $f_\varphi(d)$ is the evaluation of a document d w.r.t. a CQQL query q. To form q, various conditions φ can be linked in an arbitrary manner using the conjunction (Equation 2), disjunction (Equation 3), or negation (Equation 4). If φ is atomic, $f_\varphi(d)$ can be directly evaluated yielding a value out of the interval $[0; 1]$ As stated before, the actual value of a representation can be calculated by a similarity measure or a Boolean evaluation carried out by a DB system or the like.

After a necessary syntactical normalization step [18], the evaluation of a CQQL query is performed by recursively applying the succeeding formulas until the atomic base case is reached:

$$f_{\varphi_1 \wedge \varphi_2}(d) = f_{\varphi_1}(d) * f_{\varphi_2}(d) \tag{2}$$

$$f_{\varphi_1 \vee \varphi_2}(d) = f_{\varphi_1}(d) + f_{\varphi_2}(d) - (f_{\varphi_1}(d) \wedge f_{\varphi_2}(d)) \tag{3}$$

$$f_{\neg\varphi}(d) = 1 - f_\varphi(d) \tag{4}$$

The result of an evaluation of a document d yields the probability of relevance of d w.r.t. the given query. This probability value is then used for the ranking of the result list of documents.

2.2 Weighting in CQQL

In order to steer the influence of certain conditions onto the query evaluation, CQQL has been extended with a weighting scheme [7]. This weighting scheme can be used for relevance feedback (RF) during the retrieval process. Weighting is a crucial part of our machine-based learning supported user interaction model discussed in [18]. Although an extensive evaluation of RF for multimodal retrieval is not in the scope of this paper, we will outline how weights are embedded in a CQQL query because the weights are later used for the optimization of the discussed queries (see Table 2).

Equation 5 denotes a weighted conjunction, whereas Equation 6 states a weighted disjunction. A weight θ_i is directly associated with a logical connector and steers the influence of a representation φ_i on the evaluation. To evaluate a weighted CQQL query, the weights are syntactically replaced by constant values according to the following rules:

$$\varphi_1 \wedge_{\theta_1, \theta_2} \varphi_2 \rightsquigarrow (\varphi_1 \vee \neg\theta_1) \wedge (\varphi_2 \vee \neg\theta_2) \tag{5}$$

$$\varphi_1 \vee_{\theta_1, \theta_2} \varphi_2 \rightsquigarrow (\varphi_1 \wedge \theta_1) \vee (\varphi_2 \wedge \theta_2) \tag{6}$$

3 Experimental Description

For the scope of this paper, experiments have been conducted on low-level features combined with some of the provided metadata. The discussed approach aims at improving the performance on the plant identification task by using combinations of different features. Hence, the three main objectives of the experiments are as follows:

First, we will investigate and optimize the efficiency of a combination of visual low-level features alone.

Second, the performance improvement using low-level features as well as metadata in a CQQL query will be examined.

Third, the discussed CQQL approach will be compared with other state-of-the-art systems using classification systems like support vector machines (SVM).

In order to investigate these points, three different approaches are used in combination with a preparatory study (see Section 3.1).

Section 3.2 describes the results of a query by example (QBE) approach, while Section 3.4 presents a solution of the plant identification task using a k-medoid clustering technique. Section 3.3 combines both approaches.

3.1 Preparatory Study

In order to conduct the experiments, a preparatory study has been carried out. All assumptions used later are based on this study.

To conduct the preparatory study, we used our own developed multimodal retrieval system [14, 12]. The retrieval systems allows the extraction and combination of several different document representations, e.g. low-level features, metadata, textual information or database attributes. Additionally, the system allows a preference-based relevance feedback approach for learning weights inside a CQQL formula. A detailed description of the approach is discussed in prior work [17, 18].

To get an overview of the performance of the individual low-level features we measured the accuracy of them with the Pl@ntLeaves data set. For our first participation we decided not to evaluate each image type (scan, scan like, photographs) separately. An excerpt of the results of our initial evaluation runs is shown in Figure 1 displaying the values of precision at 5, 10, 20, 30, and mean average precision (MAP). Overall, we tested 15 low-level features in addition to the given GPS information. To calculate the similarity between the GPS data of two images we used the following similarity measure:

$$GPS_{sim} = 1 - \frac{\sqrt{(71.5 \cdot (long_x - long_y))^2 + (111.3 \cdot (lat_x - lat_y))^2}}{6378.388} \quad (7)$$

whereas $long$ stands for longitude and lat for latitude.

The MPEG-7 Color Structure Descriptor (CSD) [4], which defines a color distribution and the spatial structure of an image, was the best performing visual feature whereas GPS performed surprisingly poor. The bad performance is due to the distribution of the plant species over several territories. As a single feature it is nearly worthless for this task. Although, it can improve performance when used in combination (see Figure 2; UCC_{08}).
Based on the results of Figure 1, we tried to find different CQQL combinations that would exceed the performance of the low-level feature runs. Our first combinations were based on our former evaluations with general purpose image collections like Caltech 101 [1], Pythia [13], or Wang [10]. Unfortunately, the results could not be transferred successfully because of the specialized nature of the plant identification task.

Consequently, new combinations using the best performing low-level visual features were examined. An excerpt of the used CQQL combinations can be found in Table 2.

Test Design We evaluated the performance using the given training data but using MAP and Precision at n evaluation metrics instead of the special plant identification metric. In total, 14 CQQL formulas containing visual low-level features and two combinations of visual and GPS data were tested. To evaluate the CQQL combinations, we used a Z-Score normalization for each feature similarity
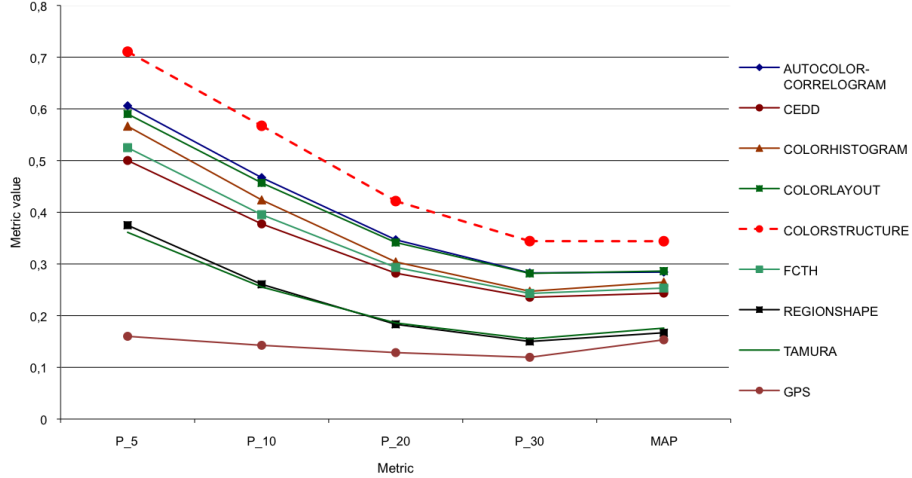
Fig. 1. Low-level feature performance on the Pl@ntLeaves data set (excerpt)

Table 2. Analyzed CQQL combinations

Name	CQQL combination
Q_{10}	$(CEDD_{sim} \vee_{\theta_1,\theta_2} FCTH_{sim}) \wedge (COLORLAYOUT_{sim} \vee (TAMURA_{sim} \wedge EDGEHISTOGRAM_{sim}))$
UCC_{00}	$\bigwedge_{\theta_i}(COLORSTRUCTURE_{sim}, CEDD_{sim}, FCTH_{sim})$
UCC_{03}	$(CEDD_{sim} \vee_{\theta_1,\theta_2} FCTH_{sim}) \wedge_{\theta_5,\theta_6}$ $(COLORSTRUCTURE_{sim} \vee_{\theta_3,\theta_4} AUTOCOLORCORRELOGRAM_{sim})$
UCC_{04}	$(REGIONSHAPE_{sim} \vee_{\theta_1,\theta_2} COLORSTRUCTURE_{sim}) \wedge_{\theta_5,\theta_6}$ $(TAMURA_{sim} \vee_{\theta_3,\theta_4} COLCORHISTOGRAM_{sim})$
UCC_{06}	$(REGIONSHAPE_{sim} \vee_{\theta_1,\theta_2} COLORSTRUCTURE_{sim}) \wedge_{\theta_7,\theta_8}$ $(TAMURA_{sim} \vee_{\theta_3,\theta_4} COLCORHISTOGRAM_{sim}) \wedge_{\theta_9,\theta_{10}}$ $(EDGEHISTOGRAM_{sim} \vee_{\theta_5,\theta_6} COLCORLAYOUT_{sim})$
UCC_{08}	$GPS_{sim} \wedge ((REGIONSHAPE_{sim} \vee_{\theta_1,\theta_2} COLORSTRUCTURE_{sim}) \wedge_{\theta_7,\theta_8}$ $(TAMURA_{sim} \vee_{\theta_3,\theta_4} COLCORHISTOGRAM_{sim}) \wedge_{\theta_9,\theta_{10}}$ $(EDGEHISTOGRAM_{sim} \vee_{\theta_5,\theta_6} COLCORLAYOUT_{sim}))$
	Weights (θ_i) are initially set to 1.0. $Feature_{sim}$ denotes similarity of a representation to the QBE document.

and the CQQL evaluation rules (see Formulae 2, 3, and 4). Together with the ground truth of the training data and our preference based approach we tried to optimize the weights used in the CQQL formulas in order to improve the retrieval metrics. Initially all weights (θ_i) are set to 1. After a learning run, the weights are set to values between 0 and 1 that fulfill the most preferences given by the ground truth. An excerpt from the final evaluation results can be found in Figure 2.

The best visual-only run gives us a small performance boost of about 14% with a MAP value of 0.39 and a P@5 of 0.75 in comparison to the best single feature ColorStructure. In contrast, the best multimodal CQQL combination (UCC08) gives a clearly greater performance boost of about 39% with a MAP value of 0.47 and P@5 of 0.8. It should be noted that we use no optimization for each image type (scan, scan like or photograph) which should improve the values even further.
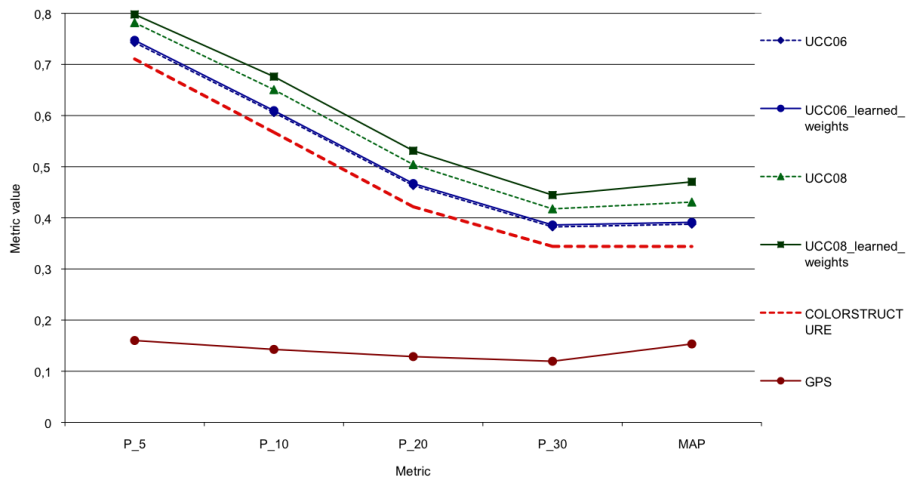

Fig. 2. Evaluation of CQQL combinations (unweighted and with learned weights)

3.2 QBE-based Approach

We used a query by example (QBE) based approach for run 1 and 2 (see Figure 3). For this approach, test images are considered QBE documents and the training images form the collection to be used for retrieval.

The only difference between both runs of the QBE approach is that the first run (our main run) uses a multimodal CQQL combination (UCC08) whereas the second run is based on a CQQL combination consisting only of visual features (UCC06).

Based on the preparatory study, weighted CQQL queries were defined using the best performing features. To learn the actual weight values, we picked a small random sample of images of all image types (scans, scan-like and photographs). With this training data, we evaluated the initial retrieval performance. To learn the best weights for a query, three steps were carried out:

1. A set of QBE images was chosen randomly from the training data. The QBE set includes all image types and some species which have the highest frequency in the training data.
2. For each of the QBE images, a ranking is calculated using a distinct set of weight values.
3. The results were interpreted using precision at n and MAP to reveal the best weight setting.

In order to find out the optimal weight setting, each QBE image is used to retrieve similar images from the collection eventually generating a ranking. As we know the species for each QBE document the ranking can then be compared with the optimal ranking of a species. This comparison is needed for the automatic preference input in order to learn weights with the presented approach. In order to provide preferences, the first 500 (at most) documents of the generated ranking are checked for irrelevant documents preceding relevant ones regarding the examined species defined by the current QBE document. If such an order $d_{irrelevant} > d_{relevant}$ is detected, and inverted preference $d_{irrelevant} < d_{relevant}$ is defined. After the first 500 (at most) documents have been tested, these preferences serve as input for the weight learning algorithm. Using these weights, a new ranking is generated. This ranking is then evaluated for the aforementioned retrieval metrics. To conclude with, all runs are averaged to determine an average set of weight values for a given query.

3.3   QBE-based Approach Combined with Top-k Clustering

For our second test (run 3), we used an approach combining a QBE-based approach with image clustering. The first part of this approach uses the same techniques as described in the last section. To reveal the relationship of all images found in the top-k result, we applied an image clustering method. We expected to find a homogeneous group of images which could then be used to determine the species of the query document. Therefore we used a distance based clustering approach, a k-medoid clustering. It was necessary to use such a method because we want use the CQQL similarities between all top-k images.

The result of the k-medoid clustering is a set of clusters. A cluster contains a mixture of training and test data (cluster members). To make a prediction which species should be associated with a given test image, we inspect the cluster containing the test image. Analyzing the rest of the cluster members (considering only training images of which the species is known), we check to which species the training data belongs. Taking the distance between the test image and the nearest training image, a ranking is created. The ordered list gives us the probability of a membership for all species within the cluster. To get a prediction

for each type of species, we take the other clusters and continue following the same principle. To optimize our approach, we evaluate which values for top-k and the number of clusters reached the best quality. The idea of the used quality measure is straightforward: a correct classification on the first predicted species will define a score value of 1. If the correct species is on the n-th position we define a score value of n. Then a minimized summed score value defines the best quality.

The usage of this combined approach gives us the opportunity to use the frequency of occurrence of the plant species. In the next approach, multiple occurrences of a species are ignored. Instead, only the first appearance is regarded important. Considering all top-k documents, we expect some species to occur repeatedly. A clustering on these images (with multiple occurrences of a species) yields a (good) probability that images with the same species belong to the same cluster (because of their high similarity). For a given test image which belongs to the same cluster there is a high probability that this test image belongs to the species too. The handling of images which occur very rarely or never in the training data set is complicated because our prediction is undefined.

3.4 Cluster-based approach

For our last try (run 4), we used a pure image clustering approach. Unlike our first methods we do not calculate an individual species prediction for each test image. Instead, we cluster the complete data set including the train and test images. We used a k-medoid clustering approach, which works on distances, calculated by a logical combination of global visual features combined with GPS data (UCC08). In this run, we used no special optimization techniques, so we apply initial weights to the CQQL formula. The species prediction of each test image is similar to the method used in the top-k clustering. For a given test image we analyzed the result list of the cluster which holds these object. To make a prediction for all categories we analyzed the nearest clusters and took all unused categories like we did in top-k clustering approach.

With this approach we are able to exploit the relationship of the test images and training images as additional information to identify the species. But without an equal distribution of all species we get some problems with categories which occur rarely as we already discussed in last section.

4   Results

According to the official results, our best run achieved place 20 in the overall ranking. Considering only automatic runs, we achieved rank 11. At a closer look, the score values of our results for scans and scan-like photographs are very poor. We cannot fully explain the failure with these image types as there were no indications for the weak performance during the training stage. One reasonable interpretation of the results is that the conducted learning runs led to an over-fitting. Another reason could be the gap between the final metric used in the

task and the metrics (precision at n, MAP) used during our preparatory study. A detailed analysis is not yet possible until the exact score calculation method and the ground truth of the test data is released. To conclude with, further research has to be carried out to find satisfying answers.

Regarding our general approach and the results in the photograph category we obtained a decent rank 6 (3rd best group) considering only automatic runs as illustrated in Figure 3. The fact that our run based on visual features alone
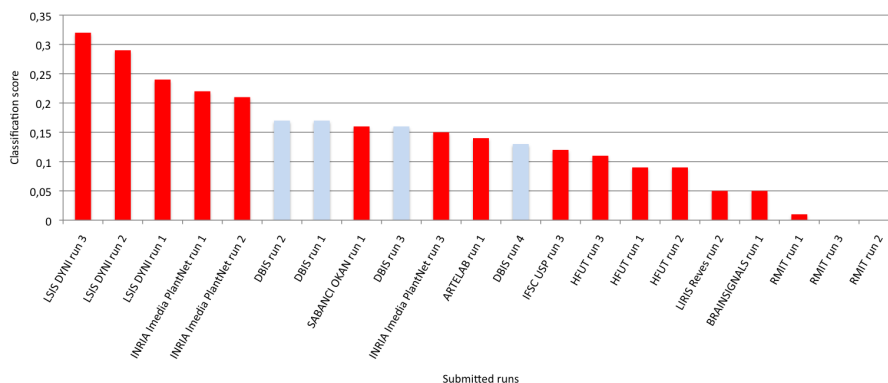

Fig. 3. Ranking score for fully automatic runs considering photographs

(run 2) performed better or as well as the combined run (run 1) for all image types is surprising. This is contradictory to the results we had observed during the training runs.

In comparison to the best submission, our results are far off. One reason for the large distance to the best group can be the training which was realized for all three image types in the same run, while other teams tried to differentiate.

5 Conclusions and Future Work

Our participation on the ImageCLEF plant identification task poses a lot of questions. The results are not devastating for our first participation. Anyhow, we are not satisfied because the results during the training were auspiciously. Furthermore, our run relying on visual features alone performed best and the differences between scan, scan-like and photographs were very small. Further research might reveal the reasons. For our next participation, we have to consider some general points that were neglected. First, we learned that concentrating on each individual image type definitely improves performance and should be incorporated in our approach. Additionally, we acknowledge that a reasonably working image segmentation is important for natural photographs in order to extract shape features which could be included into our retrieval model.

Acknowledgments

References

1. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples an incremental Bayesian approach tested on 101 object categories. In: Proceedings of the Workshop on Generative-Model Based Vision (2004)
2. Hervé Goëau, Pierre Bonnet, Alexis Joly, Itheri Yahiaoui, ans Nozha Boujemaa, D., Jean-François Molino: The ImageCLEF 2012 Plant Identification Task, CLEF 2012 Working Notes, Rome, Italy (2012)
3. ImageClef: Plant Identification 2012. http://imageclef.org/2012/plant, 12. August 2012.
4. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley & Sons, Inc., New York, NY, USA (2002)
5. Nilsson, J.N.: Probabilistic logic. Artif. Intell. 28, 71–88 (1986), http://dl.acm.org/citation.cfm?id=5447.5450
6. Pl@ntNet: Interactive plant identification and collaborative information system. http://www.plantnet-project.org/papyrus.php?langue=en, 12. August 2012.
7. Schmitt, I.: Weighting in CQQL. Cottbus (2007)
8. Schmitt, I.: QQL: A DB&IR Query Language. The VLDB Journal 17(1), 39–56 (2008)
9. Schmitt, I., Zellhöfer, D., Nürnberger, A.: Towards quantum logic based multimedia retrieval. In: IEEE (ed.) Proceedings of the Fuzzy Information Processing Society (NAFIPS). pp. 1–6. IEEE (2008), 10.1109/NAFIPS.2008.4531329
10. Wang, Z.J., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries. In: Proceedings of the 4th International Conference on Advances in Visual Information Systems. pp. 360–371. VISUAL '00, Springer-Verlag (2000), http://portal.acm.org/citation.cfm?id=647061.714442
11. Zadeh, A.L.: Fuzzy Logic. IEEE Computer 21(4), 83–93 (1988)
12. Zellhöfer, D.: A Permeable Expert Search Strategy Approach to Multimodal Retrieval. In: Proceedings of the 4th international information interaction in context symposium. p. to appear. IIiX, ACM (2012)
13. Zellhöfer, D.: An Extensible Personal Photograph Collection for Graded Relevance Assessments and User Simulation. In: Proceedings of the ACM International Conference on Multimedia Retrieval. ICMR '12, ACM (2012)
14. Zellhöfer, D., Bertram, M., Böttcher, T., Schmidt, C., Tillmann, C., Schmitt, I.: PythiaSearch – A Multiple Search Strategy-supportive Multimedia Retrieval System. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval. p. to appear. ICMR '12, ACM (2012)
15. Zellhöfer, D., Böttcher, T.: BTU DBIS' Multimodal Wikipedia Retrieval Runs at ImageCLEF 2011. In: Vivien Petras, Pamela Forner and Paul D. Clough (eds.) CLEF 2011 Labs and Workshop, Notebook Papers, 19-22 September 2011, Amsterdam, The Netherlands (2011)

16. Zellhöfer, D., Frommholz, I., Schmitt, I., Lalmas, M., van Rijsbergen, K.: Towards Quantum-Based DB+IR Processing Based on the Principle of Polyrepresentation. In: Clough, P., Foley, C., Gurrin, C., Jones, G., Kraaij, W., Lee, H., Murdoch, V. (eds.) Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings, Lecture Notes in Computer Science, vol. 6611, pp. 729–732. Springer (2011)
17. Zellhöfer, D., Schmitt, I.: A Poset Based Approach for Condition Weighting. In: 6th International Workshop on Adaptive Multimedia Retrieval (2008)
18. Zellhöfer, D., Schmitt, I.: A Preference-based Approach for Interactive Weight Learning: Learning Weights within a Logic-Based Query Language. Distributed and Parallel Databases (2009), doi:10.1007/s10619-009-7049-4