

# DCU and UTA at ImageCLEFPhoto 2007

Anni Järvelin<sup>1</sup>, Peter Wilkins<sup>2</sup>, Tomasz Adamek<sup>2</sup>, Eija Airio<sup>1</sup>, Gareth J. F. Jones<sup>2</sup>,  
Alan F. Smeaton<sup>2</sup> & Eero Sormunen<sup>1</sup>

<sup>1</sup> University of Tampere, <sup>2</sup> Dublin City University

Anni.Jarvelin@uta.fi, pwilkins@computing.dcu.ie

## Abstract

Dublin City University (DCU) and University of Tampere (UTA) participated in the ImageCLEF 2007 photographic ad-hoc retrieval task with several monolingual and bilingual runs. Our approach was language independent: text retrieval based on fuzzy *s*-gram query translation was combined with visual retrieval. Data fusion between text and image content was performed using unsupervised query-time weight generation approaches. Our baseline was a combination of dictionary-based query translation and visual retrieval, which achieved the best result. The best mixed modality runs using fuzzy *s*-gram translation achieved on average around 83% of the performance of the baseline. Performance was more similar when only top rank precision levels of P10 and P20 were considered. This suggests that fuzzy *s*-gram query translation combined with visual retrieval is a cheap alternative for cross-lingual image retrieval where only a small number of relevant items are required. Both sets of results emphasize the merit of our query-time weight generation schemes for data fusion, with the fused runs exhibiting marked performance increases over single modalities, this is achieved without the use of any prior training data.

## Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval

## General Terms

Measurement, Performance, Experimentation

## Keywords

Fuzzy matching, Content-Based Image Retrieval, Data Fusion

## 1 Introduction

In earlier ImageCLEF campaigns combined text and visual retrieval approaches have performed better than text or image retrieval alone. Also, text retrieval alone has clearly outperformed visual-only retrieval approaches [15]. In this year's campaign, text retrieval was faced with a new challenge of retrieval of lightly annotated photographs [6]. A negative impact on the performance of the text retrieval techniques was to be expected. When the textual context available for image retrieval is sparse, visual retrieval gains more significance and thus effective fusion of text and visual features becomes important. Therefore one of our main research interests in the ImageCLEF photo 2007 task was to study different techniques for fusion of text and visual retrieval results.

Retrieving images by their associated text is a common approach in image retrieval [15]. When cross-language image retrieval is considered, this approach requires language dependent linguistic resources for query translation. Machine-readable dictionaries, machine translation tools or

corpus-based translation tools are expensive, and are not available for all the language pairs. However, there are alternative approaches which may be used to compensate for poor or absent linguistic translation tools, for example the approximate string matching techniques of n-gramming and its generalization *s*-gramming. An aspect that speaks for the use of approximate string matching in image retrieval is that proper names are very common query terms when searching image databases [13]. Machine readable bilingual dictionaries typically do not cover proper names and thus they often remain untranslatable in queries. Proper names in related languages are nevertheless often spelling variants of each other and can thus be recognized and translated using approximate string matching techniques. Generally, when approximate string matching techniques are used, the query words need to have somewhat close cognates in the target language for the translation to succeed.

We participated in ImageCLEF 2007 photographic ad-hoc retrieval task in order to test a language independent image retrieval approach. The *s*-gram-based query translation was fused with visual retrieval. To study the relatedness of the source and target languages on the translation quality we selected language pairs where source/target languages were related to each other at different levels. The Scandinavian languages Danish, Norwegian and Swedish are quite closely related to German, while French was the language closest to English available. German and English instead are not very closely related languages and therefore translation between them should be hard. Thus, we had six language pairs: French-English, German-English, Danish-German, Norwegian-German, Swedish-German and English-German. To have a strong baseline, the performance of the language independent approach was compared to a state-of-the-art technique, a combination of dictionary-based query translation and visual retrieval. A total of 138 runs were submitted. Reporting the results for all of these would be impractical and therefore only the results for the best and most interesting runs are presented here.

Our exploration of data fusion continued within our work in query-time coefficient generation for retrieval expert combination. Within our ImageCLEF work we experimented primarily with altering the stages at which we fuse various experts together. For instance we experimented with fusing all the visual experts into a single expert, then fusing with text, as opposed to treating all experts equally.

The structure of the paper is following. The text and visual retrieval approaches are presented in more detail in Section 2. In Section 3 our retrieval systems and runs are presented in detail. The results are reported in section 4 and Section 5 finally concludes with a short discussion.

## 2 Background

### 2.1 *s*-gram-based query translation

Approximate string matching techniques enable target database index words most similar to a query word to be recognized as its translations, and can be used for fuzzy query translation. *s*-gram matching is a technique like this. In *s*-gram matching the text strings to be compared are decomposed into substrings (*s*-grams) and the similarity between the strings is calculated as the overlap of their common substrings. *s*-gram matching is a generalization of the well known n-gram technique, which is quite commonly used for matching cognates in CLIR. While the n-gram substrings consist of adjacent characters of the original strings, skipping some characters is allowed when forming the *s*-grams. In classified *s*-gram matching [18] different types of *s*-grams are formed by skipping different number of characters. The *s*-grams are then classified into sets based on the number of characters skipped, and only the *s*-grams belonging to the same set are compared to each other when calculating the similarity. Character combination index (CCI) indicates the set of all the *s*-gram types to be formed from a string. CCI  $\{\{0\}, \{1, 2\}\}$  for example means that three types of *s*-grams are formed and classified into two sets: one set of conventional n-grams formed of adjacent characters ( $\{0\}$ ) and one of *s*-grams formed both by skipping one and two characters ( $\{1, 2\}$ ). For an extensive description of the *s*-gram matching technique see [18, 10].

n-gram matching has been used for fuzzy translation in a few studies, e.g., by [7] and [8] as a complement to the dictionary-based translation and by [14] for the whole query translation. Järvelin et al. 2007 [9] used *s*-gram matching in fuzzy query translation between the closely related languages Swedish and Norwegian. Other approaches to fuzzy query translation have been developed by [19], who translated out-of-vocabulary words based on statistical transformation rules for character changes between the source and target languages, and by [4] who treated English query words as misspelled French and attempted to translate them using a spelling correction program.

## 2.2 Visual Retrieval

To facilitate visual retrieval we made use of six ‘low-level’ global visual experts. Our visual features are MPEG7 features and were extracted using the aceToolBox, developed as part DCU’s collaboration in the aceMedia project [1]. These six features included: Colour Layout, Colour Structure, Colour Moments, Scalable Colour, Edge Histogram and Homogenous Texture. Further details on these descriptors can be found in [16] and [11]. Distance metrics for each of these features were implementations of those specified in the MPEG7 specification [11].

## 2.3 Query-Time Fusion

The combination of retrieval experts for a given information need can be expressed as a *data fusion* problem [3]. Given that for any given information need different retrieval experts perform differently, we require some form of weighting scheme in order to combine experts. Typical approaches to weight generation include the use of query-independent weights, learnt through experimentation on a training collection or query-class experts weights [22]. A query is categorized into some pre-determined class, and the weights for each class are again learnt through experimentation on a training collection. These approaches however require that both a training collection, and a representative set of training queries for that collection so as to avoid problems of over-fitting expert weights.

Our approach to weight generation differs in that it is a query-dependant weighting scheme for expert combination which requires no training data [21]. This work was based upon an observation, that if we plot the normalized scores of an expert, against that of scores of other experts used for a particular query, the expert who’s scores exhibited the greatest initial change in scores correlated with that expert being the best performer for this query. An example of this observation is presented in Figure 1 and Table 1.

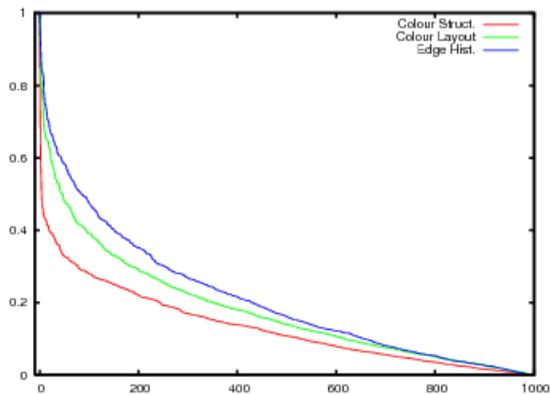


Figure 1: ImageCLEF Topic 35 Single Image Query

Name	AP	P@5	P@10	Recall
Edge Hist.	0.0130	0.2	0.1	10%
Colour Layout	0.0260	0.4	0.3	16%
<b>Colour Struct.</b>	<b>0.0643</b>	<b>0.6</b>	<b>0.4</b>	<b>43%</b>

Table 1: ImageCLEF Topic 35 Single Image Results

From the graph it can be observed that the greatest initial change in score is exhibited by the Colour Structure expert, which correlates to the table of performance where Colour Structure is the best performing of the experts. However, this observation is informing us that the Colour Structure expert is likely to be the better performing expert relative to the other experts used for this query. This observation is not giving us any absolute indication of expert performance, which would be the case in techniques such as those employed by Manmathma et al. [12], or Arampatzis et al. [2].

At this stage we should note that this observation is not universal, and that there are failure cases where this observation will not hold. If we assume though that in a majority of queries this observation does hold, then we can employ techniques that leverage this approach to create query-time expert coefficients for data fusion.

Our initial method of exploiting these observations is designed to assign a higher weight to the expert which undergoes the greatest initial change in score. For a given expert, we can measure the average change in score through Equation 1, which we refer to as Mean Average Distance (MAD):

$$MAD = \frac{\sum_{n=1}^{N-1} (score(n) - score(n+1))}{N-1} \quad (1)$$

This value alone though does not provide sufficient information for deriving the rate of change in score of an expert. To achieve this, we define a ratio which measures MAD for a top subset of an expert’s document scores over a larger set of document scores, given in Equation 2:

$$SC = \frac{MAD(subset)}{MAD(largerset)} \quad (2)$$

This ratio provides us with a single measure of the degree of change in an expert’s document scores. With this we can simply divide each ratio value by the sum of all expert’s ratio values to arrive at the coefficient for the given expert, formally given in Equation 3:

$$Expert\ Weight = \frac{Feature\ SC\ Score}{\sum All\ SC\ Scores} \quad (3)$$

The crux of this approach lies in the determination of the values to be used for ‘subset’ and ‘larger set’ in Equation 2. We have previously experimented with fixed sizes for these variables, but found that using a percentage of the size of the expert’s returned result list works best, which in practise is the top 5% of an expert’s results over the top 95% of the expert’s results.

As this data fusion technique only requires the results that an expert provides, it can be used to combine multi-modal experts together. Furthermore, it allows us to combine single experts for a given query image into a single result list for that given query image. Once this is performed, we are left with a single result list per query image, meaning we can reapply the same technique to generate weights that can be applied to the query image. This allows us to weight query images differently, as for various queries, each query image may perform better or worse than other query images for a given query. A more complete explanation of this process can be found in [21].

## 3 Resources, Methods and Runs

### 3.1 Text Retrieval and Indexing

We utilized the Lemur toolkit (Indri engine) for indexing and retrieval ([www.lemurproject.org](http://www.lemurproject.org), [20]). Indri combines language modeling to inference nets and allows structured queries to be evaluated using language modeling estimates rather than tf\*idf estimates. The word tokenization rules used in indexing were the following. First, punctuation marks were converted into spaces, and capitals were converted into lower case. Next, strings broken down by the space character were decoded to be individual words. Different indices were created for the *s*-gram-based and dictionary-based query translation. For the dictionary-based translation, lemmatized English and German indices were created. The words of the image annotations were lemmatized with English and German lemmatizers and lemmas were stored into the indices. Words not recognized by the lemmatizers were indexed as such. Compound words were split and both the original compound and the constituents were indexed. With the *s*-gram-based translation we used inflected indices, where the words were stored in the inflected word forms in which they appeared in the image annotations. Stop words were removed.

Topic words were not morphologically processed prior to the *s*-gram-based query translation. Stop words were removed and the remaining words were translated into the target language with *s*-gram matching. The CCI was set to be 0,1,2 and the Jaccard coefficient [10] was used for calculating the *s*-gram proximity. Three best matching index words were selected as translations for each topic word. A similarity threshold value of 0.3 was set to discard bad translations: only the index words exceeding this threshold with respect to a source word were accepted as its translation equivalents. As a consequence, some of the query words could not be translated. Queries were structured utilizing a synonym operator: target words derived from the same source word were grouped into the same synonym group. This kind of structure balances the queries and down-weights the ambiguous words with many possible translations (*the Pirkola method*, [17]). We used Indri's Pseudo-Relevance Feedback (PRF) with 10 keys taken from 10 highest ranked documents in the original result list.

The UTAQLIR query translation tool developed at UTA was used for the dictionary-based query translation. UTAQLIR was originally developed for the CLEF 2000 and 2001 campaigns [7]. The system utilizes external language resources, such as translation dictionaries, stemmers and lemmatizers. Word processing in UTAQLIR proceeds in the following way. First, topic words are lemmatized in order to match them with dictionary words (our translation dictionaries include source words only in their basic forms). The lemmatizer produces one or more basic forms for a token. After normalization, stop words are removed, and non-stop words are translated. Untranslatable compound words are split and the constituents are translated. Translation equivalents are normalized utilizing a target language lemmatizer. Untranslatable words are matched against the database index with the *s*-gram matching technique and three best matches are chosen as translations. Queries were structured according to the Pirkola method utilizing a synonym operator. Since there was no French morphological analyzer available to us, the French topics were analyzed manually. This might result in a slightly better quality of lemmatization than automatic analysis, even though we strived for equal quality. In this case we performed PRF with 20 expansion keys from the 15 top ranked documents.

### 3.2 Data Fusion

The query-time data fusion approach described in Section 2.3 was adopted as our basic approach to expert combination. However, one set of parameters that was not specified was the order in which experts will be combined. This is the focus of our experimental work in this section.

One commonality between all the combination approaches we try in this work is the fusion of the low-level visual experts. For each query image we fuse the six low-level visual experts into a single result for each image, where their combination uses the aforementioned technique.

Therefore for each query, the visual component was then represented by three result sets, one for each query image. Additionally for a subset of our runs we introduce a seventh visual expert, the FIRE baseline [5]. In cases where FIRE was used, because it was a single result for the three visual query images, we first combined our MPEG7 visual features into a single result for each image, then these were combined into an overall image result, which was then combined with the FIRE baseline. This combination made use of the query-time weighting scheme.

We explored four variants in our combination work as follows:

- dyn-equal: Query-time weighting method with text and individual image results combined at the same level, i.e. we have three image results and one text result which are to be combined.
- dyn-top: As above, except the results for each query image were fused into a single image result, which was then combined with the text result, i.e. image results combined into a single result, which is then combined with the single text result.
- stat-eventop: Query-time weighting to produce single image result list, image and text fused together with equal static weighting (0.5 coefficient).
- stat-imgHigh: As above, except with the image result assigned a static weight of 0.8 and text a static weight of 0.2.

Additionally, any of our runs which ended in ‘fire’ incorporated the FIRE baseline into the set of visual experts used for combination.

## 4 Results

Our tables of results are organized as follows. Table 2 presents our baseline runs, both text-only monolingual and visual-only. Table 3 presents our baseline fusion results, mixing monolingual text with visual experts. Table 4 presents our central cross-lingual results with mixed modalities. In all tables where data fusion is utilized, we present only the best performing data fusion approach. With the exception of Table 2, all visual results used in data fusion presented here incorporated the FIRE baseline. Visual data which included the FIRE baseline with our global MPEG7 features consistently outperformed global MPEG7 by themselves.

Language Pair	Modality	Text	Fusion	FB	MAP	P@10	P@20
EN-EN	Text	dict	n/a	no	0.1305	0.1550	0.1408
EN-EN	Text	<i>s</i> -gram	n/a	yes	0.1245	0.1133	0.1242
DE-DE	Text	dict	n/a	yes	0.1269	0.1717	0.1533
DE-DE	Text	<i>s</i> -gram	n/a	yes	0.1067	0.1233	0.1125
MPEG7 With FIRE	Visual	na	dyn-equal	no	0.1340	0.3600	0.2658
MPEG7 Without FIRE	Visual	na	dyn-equal	no	0.1000	0.2700	0.1958

Table 2: ImageCLEF Baseline Results

Two different indexing approaches were used for the monolingual text runs: The runs marked dict in Table 2 used lemmatized indices and topic words. The *s*-gram runs used inflected indices and the topic words were matched against the index words with *s*-gram matching. The runs with morphological analysis performed slightly better than the *s*-gram runs, but for the English runs the difference is very small. For German runs the difference is greater, which is understandable as German has much more complex inflectional morphology than English. Our text-only and

visual-only retrieval techniques were almost equal, which is notable in the context of ImageCLEF results in previous years. Our best visual-only run performed well being the overall second best visual approach in terms of Mean Average Precision (MAP). Its MAP value 0.1340 is comparable to our best English-English text-only run scoring 0.1305. We believe that the comparative low performance of the text expert (when compared to the dominance of text in previous years) was due to the reduced length of the annotations for 2007.

Language Pair	Modality	Text	Fusion	FB	MAP	P@10	P@20
EN-EN	Mixed	dict	dyn-equal	yes	0.1951	0.3967	0.3150
EN-EN	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1833	0.3833	0.3092
DE-DE	Mixed	dict	dyn-equal	yes	0.1940	0.4033	0.3300
DE-DE	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1628	0.3350	0.2792

Table 3: ImageCLEF Monolingual Fusion Results

Table 3 presents our fused text and visual retrieval runs, which performed best among our submissions, being clearly better than any of the text or visual runs alone. Of interest is that in each data fusion case, the increase in MAP was 65%-67%. Whilst this is too small a sample to infer any conclusions, it is nonetheless interesting that such a consistent gain was achieved and is worthy of further investigation.

From a data fusion perspective, no single approach of the four we tried consistently performed the best. Whilst our results presented here show the “dyn-equal” fusion as being superior, this is because it was the only fusion type attempted with visual data which incorporated the FIRE baseline. For runs where FIRE was not used, the best performing fusion type varied depending on the text type (dictionary or *s*gram) or language pair used. In the majority of cases all fusion types performed similarly, deeper investigation with significance testing will be required in order to infer any meaningful interpretations. However, we can conclude that as all four fusion types made use of our query-time weight generation method at some level, that this technique is capable of producing weights which lead to performance improvements when combining results. What is unknown is how far from the optimal query-dependant combination we achieved, and knowing this would give one of the ultimate measures of the success of this approach.

Table 4 presents our central cross-lingual results. Dictionary-based query translation was the best query translation approach. The best mixed modality runs using the *s*-gram-based query translation reached on average around 84% of the MAP of the best mixed modality runs using dictionary-based translation. These approaches were more similar when the top rank precision values at P10 and P20 were considered: the best *s*-gram runs reached on average around 91% of the best dictionary-based runs performance at P10 and around 89% at P20. If the high ranks of the result list are considered to be important from the user perspective, the *s*-gram translation could be seen as almost equal with the dictionary-based translation in mixed modality runs. The results for the two text query translation techniques varied depending on the language pair. Generally the *s*-gram-based translation and the dictionary-based translation were quite equal for the more closely related language pairs, while the dictionary-based translation was clearly better for the more distant language pairs. The *s*-gram translation reached its best results with Norwegian and Danish topics and German annotations - over 90% of the dictionary translation’s MAP, and the worst ones between German and English - less than 80% of the dictionary translation’s MAP.

Language Pair	Modality	Text	Fusion	FB	MAP	P@10	P@20
FR-EN	Mixed	dict	dyn-equal	yes	0.1819	0.3583	0.2967
FR-EN	Mixed	<i>s</i> -gram	dyn-equal	no	0.1468	0.3483	0.2667
DE-EN	Mixed	dict	dyn-equal	yes	0.1910	0.3483	0.3042
DE-EN	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1468	0.3233	0.2533
DA-DE	Mixed	dict	dyn-equal	yes	0.1730	0.3467	0.2942
DA-DE	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1572	0.3350	0.2717
NO-DE	Mixed	dict	dyn-equal	yes	0.1667	0.3517	0.2700
NO-DE	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1536	0.3167	0.2667
SV-DE	Mixed	dict	dyn-equal	yes	0.1788	0.3817	0.2942
SV-DE	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1472	0.3050	0.2500
EN-DE	Mixed	dict	dyn-equal	yes	0.1828	0.3633	0.3008
EN-DE	Mixed	<i>s</i> -gram	dyn-equal	yes	0.1446	0.3350	0.2667

Table 4: ImageCLEF CLIR Fusion Results

## 5 Conclusions

In this paper we have presented results for the joint DCU and UTA participation in the ImageCLEF 2007 Photo task. In our work we experimented with two major variables, that of cross-lingual text retrieval utilizing minimal translation resources, and query-time weight generation for expert combination. Our results are encouraging and support further investigation into both approaches. Further work is now required to conduct a more thorough analysis of contributing factors to performance emphasized by each approach. Of particular interest will be the degree to which each of these approaches introduced new information, or re-ordered existing information presented by the systems. For instance, we do not know yet if the *s*-gram retrieval found relevant documents that were missed by the dictionary based approach. Likewise for data fusion, we do not know yet if we promoted into the final result set relevant results which were only present in some and not all of the experts used.

## 6 Acknowledgments

We are grateful to the AceMedia project (FP6-001765) which provided us with output from the AceToolbox image analysis toolkit.

The research leading to this paper was supported by the European Commission under contract FP6-027026 (K-Space).

The work of the first author is funded by Tampere Graduate School of Information Science and Engineering (TISE).

## References

- [1] The AceMedia Project, available at <http://www.acemedia.org>.
- [2] Avi Arampatzis and André van Hameran. The score-distributional threshold optimization for adaptive binary classification tasks. In *SIGIR '01: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 285–293, New York, NY, USA, 2001. ACM Press.
- [3] N. J. Belkin, P. Kantor, E. A. Fox, and J. A. Shaw. Combining the evidence of multiple query representations for information retrieval. *Information Processing and Management*, 31(3):431–448, 1995.

- [4] Chris Buckley, Mandar Mitra, Janet A. Walz, and Claire Cardie. Using clustering and super-concepts within SMART: TREC 6. *Information Processing and Management*, 36(1):109–131, 2000.
- [5] Thomas Deselaers, Tobias Weyand, Daniel Keysers, Wolfgang Macherey, and Hermann Ney. FIRE in ImageCLEF 2005: Combining content-based image retrieval with textual information retrieval. In *CLEF*, pages 652–661, 2005.
- [6] Michael Grubinger, Paul Clough, Allan Hanbury, and Henning Müller. Overview of the ImageCLEFphoto 2007 photographic retrieval task. In *Working Notes of the 2007 CLEF Workshop*, Budapest, Hungary, September 2007.
- [7] Turid Hedlund, Heikki Keskustalo, Ari Pirkola, Eija Airio, and Kalervo Järvelin. Utaclir @ CLEF 2001 - effects of compound splitting and n-gram techniques. In *CLEF '01: Revised Papers from the Second Workshop of the Cross-Language Evaluation Forum on Evaluation of Cross-Language Information Retrieval Systems*, pages 118–136, London, UK, 2002. Springer-Verlag.
- [8] Djoerd Hiemstra and Wessel Kraaij. Twenty-One at TREC7: Ad-hoc and cross-language track. In *TREC*, pages 174–185, 1998.
- [9] Anni Järvelin, Antti Järvelin, and Kalervo Järvelin. s-grams: Defining generalized n-grams for information retrieval. *Information Processing and Management*, 43(4):1005–1019, 2007.
- [10] Heikki Keskustalo, Ari Pirkola, Kari Visala, Erkkä Leppänen, and Kalervo Järvelin. Non-adjacent digrams improve matching of cross-lingual spelling variants. In *SPIRE*, pages 252–265, 2003.
- [11] B.S. Manjunath, P. Salembier, and T. Sikora, editors. *Introduction to MPEG-7: Multimedia Content Description Language*. Wiley, 2002.
- [12] R. Manmatha, T. Rath, and F. Feng. Modeling score distributions for combining the outputs of search engines. In *SIGIR '01: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 267–275, New York, NY, USA, 2001. ACM Press.
- [13] Marjo Markkula and Eero Sormunen. End-user searching challenges indexing practices in the digital newspaper photo archive. *Information Retrieval*, 1(4):259–285, 2000.
- [14] Paul McNamee and James Mayfield. Character n-gram tokenization for european language text retrieval. *Information Retrieval*, 7(1-2):73–97, 2004.
- [15] Henning Müller, Thomas Deselaers, Thomas Lehmann, Paul Clough, and William Hersh. Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. In C. Peters, P. Clough, F. Gey, J. Karlgren, B. Magnini, D.W. Oard, M. de Rijke, and M. Stempfhuber, editors, *Evaluation of Multilingual and Multi-modal Information Retrieval – Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, page to appear, Alicante, Spain, September 2007.
- [16] Noel O’Connor, Edward Cooke, Herve le Borgne, Michael Blighe, and Tomasz Adamek. The AceToolbox: Low-Level Audiovisual Feature Extraction for Retrieval and Classification. In *2nd IEE European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, 2005.
- [17] Ari Pirkola. The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval. In *SIGIR '98: Proceedings of the 21st Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 55–63, 1998.

- [18] Ari Pirkola, Heikki Keskustalo, Erkkä Leppänen, Antti-Pekka Känsälä, and Kalervo Järvelin. Targeted s-gram matching: a novel n-gram matching technique for cross- and mono-lingual word form variants. *Information Research*, 7(2), 2002.
- [19] Ari Pirkola, Jarmo Toivonen, Heikki Keskustalo, Kari Visala, and Kalervo Järvelin. Fuzzy translation of cross-lingual spelling variants. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 345–352, New York, NY, USA, 2003. ACM Press.
- [20] Trevor Strohman, Donald Metzler, Howard Turtle, and W. Bruce Croft. Indri: A language-model based search engine for complex queries (extended version). 2005-02-14 2005.
- [21] Peter Wilkins, Paul Ferguson, and Alan F. Smeaton. Using score distributions for querytime fusion in multimedia retrieval. In *MIR 2006 - 8th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2006.
- [22] Rong Yan, Jun Yang, and Alexander G. Hauptmann. Learning query-class dependent weights in automatic video retrieval. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 548–555, New York, NY, USA, 2004. ACM Press.