Multimodal information approaches for the Wikipedia collection at ImageCLEF 2011

1 Universidad Nacional de Educación a Distancia, UNED
2 Universidad de Valencia, UV

R. Granados¹, J. Benavent², X. Benavent², E. de Ves², Ana García-Serrano¹
Index

1. Objectives UNED-UV @ ImageCLEF2011
2. System overview
3. TBIR: Textual approach
4. CBIR: Visual approach
5. Fusion: merge and reordering
6. Runs and Results
7. Conclusions and future work
2. System Overview
3TBIR

- **IDRA Tool (InDexing and Retrieving Automatically)**
  Our own implemented tool (under GPL 3.0)

- **Architecture in several modules**

  - **Preprocessing** (language-dependent):
    - Stopwords, Stemming (Snowball)
    - NER (Stilus - DAEDALUS)

  - **Indexing**: Vector Space Model approach (IDRA) and Basic Lucene executed from IDRA

  - **Merging Module to Multilinguality and textual enrichment**:
    - Join (features)
    - Enrich (merge of two results lists)
    - MAXmerge (merge of two results lists)
## 6. TXT Runs and Results - Baseline

<table>
<thead>
<tr>
<th>ID</th>
<th>Lang</th>
<th>Details</th>
<th>Sub-System</th>
<th>Fusion</th>
<th>MAP</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run1</td>
<td>EN</td>
<td></td>
<td>IDRA</td>
<td>-</td>
<td>0.1727</td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td>run2</td>
<td>EN</td>
<td></td>
<td>IDRA + stemming</td>
<td>-</td>
<td>0.2056</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>run3</td>
<td>EN</td>
<td></td>
<td>IDRA + stemming + Cat</td>
<td>-</td>
<td>0.2243</td>
<td>21</td>
<td>6</td>
</tr>
</tbody>
</table>

- Stemming (EN, DE, and FR)
- WP Cat and TITLE are joined to DE, EN, FR
<table>
<thead>
<tr>
<th>ID</th>
<th>Lang</th>
<th>Sub-System</th>
<th>Fusion</th>
<th>MAP</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run3</td>
<td>EN</td>
<td>IDRA</td>
<td>-</td>
<td>0.2243</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>run4</td>
<td>DE+FR +EN</td>
<td>IDRA</td>
<td>MAXmerge (3en, 3fr, 3de)</td>
<td>0.2489</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>run5</td>
<td>EN</td>
<td>IDRA+Lucene</td>
<td>-</td>
<td>0.2601</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>run8</td>
<td>DE+FR +EN</td>
<td>IDRA+Lucene</td>
<td>MAXmerge (5en, 5fr, 5de)</td>
<td>0.3044</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

- Differences between **IDRA** indexing and ranking and **Basic Lucene**
### 3TBIR

- **NEs** are indexed separately or joined with EN metadata

<table>
<thead>
<tr>
<th>ID</th>
<th>Lang</th>
<th>Sub-System</th>
<th>Fusion</th>
<th>MAP</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run5</td>
<td>EN</td>
<td>IDRA+Lucene</td>
<td></td>
<td>0.2601</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>run7</td>
<td>EN</td>
<td>IDRA+ Lucene+NER</td>
<td>Enrich (5en,NEs)</td>
<td>0.2515</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>run10</td>
<td>EN</td>
<td>IDRA+ Lucene+NER</td>
<td>JOIN</td>
<td>0.2403</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>
2. Multilingual Search

- TBIR
  - WP Cat
  - NER
  - DE
  - FR
  - EN

- CBIR
  - TXT Results
  - IMG Results

- Merging
  - MAXmerge
  - ENRICH

- FUSION
  - (Pt)
  - (Pi)
  - (Pi * Pi)

- Topics

- Wikipedia Collection
  - metadata
  - articles
  - images
6. Runs and Results - Multilinguality

<table>
<thead>
<tr>
<th>ID</th>
<th>Lang</th>
<th>Sub-System</th>
<th>Details</th>
<th>MAP</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run5</td>
<td>EN</td>
<td>IDRA+Lucene</td>
<td>-</td>
<td>0.2601</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>run8</td>
<td>DE+FR</td>
<td>IDRA+Lucene</td>
<td>MAXmerge (5en, 5fr, 5de)</td>
<td>0.3044</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>run 9</td>
<td>ALL</td>
<td>IDRA+Lucene</td>
<td>JOIN</td>
<td>0.2758</td>
<td>9</td>
<td>3</td>
</tr>
</tbody>
</table>
4. CBIR at 2009

**Low level Features:**
- Color
- Texture

**Algorithms:**
- Automatic
- Query expansion: N first TXT Results List
- **Relevance feedback:** Logistic Regression algorithm

**Best results at 2009**
4. CBIR (Relevance Feedback)

- The logistic regression model is used to estimate the relevance probabilities of each image in the collection (10,000).

\[
\begin{align*}
x &= (x_1, \ldots, x_p) \\
\pi(x) &= P(Y = 1 \mid x) \\
\text{logit}[\pi(x)] &= \alpha + \beta_1 x_1 + \ldots + \beta_p x_p = \ln \frac{\pi(x)}{1-\pi(x)} \\
\pi(x) &= \frac{\exp(\alpha + \beta_1 x_1 + \ldots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \ldots + \beta_p x_p)}
\end{align*}
\]

- **2011**: Manual relevance feedback into automatic one
  - The relevant images set: query images plus the first 3 ranked images of the text filter.
  
  - The non-relevant images set: first we select # random images out of the text filter, then it is ordered by the Euclidean distance, and finally the negative examples are de M latest ones.
4. CBIR (ImageClef@Wikipedia 2011)

- Testing the behavior of our low-level features (UV) with the CEDD ones

- UV Low level features:
  - Color information (four fragments of same size): A feature vector of 222 components
  - Texture information (granulometric distribution and Spatial size distribution): A feature vector pf 71 components

<table>
<thead>
<tr>
<th>ID</th>
<th>TXT run</th>
<th>Details</th>
<th>MAP</th>
<th>Fusion</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run11</td>
<td>run9+UV</td>
<td>0.0553</td>
<td>Pi</td>
<td>56</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>run12</td>
<td>run9+CEDD</td>
<td>0.0516</td>
<td>Pi</td>
<td>57</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>run13</td>
<td>run9+CEDD</td>
<td>0.2869</td>
<td>Automatic</td>
<td>18</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>run14</td>
<td>run9 +CEDD</td>
<td>0.3006</td>
<td>Pt*Pi</td>
<td>12</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>run17</td>
<td>run9+UV</td>
<td>0.3006</td>
<td>Pt*Pi</td>
<td>13</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
5. Late Fusion

Main process:
- First TBIR works over the whole Image Collection
- CBIR works from the TXT results list (10,000)
  TBIR acts as a filter based on the assumption that the textual module better captures the meaning of a topic.

- Approaches (Pt, Pi ranges are not normalized):
  - Pt*Pi, Pt = 1 (only visual ranking)
  - Pt*Pi
  - OWA: average(Pt, Pi), orness(Pt, Pi) with 0.3
## 6. Runs and Results

<table>
<thead>
<tr>
<th>ID</th>
<th>TXT run</th>
<th>Details</th>
<th>MAP</th>
<th>Fusion</th>
<th>Run Pos.</th>
<th>Group Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>run14</td>
<td>run9 +CEDD</td>
<td>0.3006</td>
<td>Pt*Pi</td>
<td>12</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>run15</td>
<td>run9 +CEDD</td>
<td>0.2869</td>
<td>OWA(Med)</td>
<td>17</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>run16</td>
<td>run9 +CEDD</td>
<td>0.2980</td>
<td>OWA(Orness0.3)</td>
<td>14</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>run18</td>
<td>run8 +CEDD</td>
<td><strong>0.3405</strong></td>
<td>Pt*Pi</td>
<td><strong>8</strong></td>
<td><strong>2</strong></td>
<td></td>
</tr>
<tr>
<td>run19</td>
<td>run8 +CEDD</td>
<td>0.3233</td>
<td>OWA(Med)</td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>run20</td>
<td>run8 +CEDD</td>
<td>0.3367</td>
<td>OWA(Orness0.3)</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

*Note: run18's MAP is highlighted due to having the highest value.*
7. Future Work

- Good way of working fusing textual and visual list
  - Refine the identified differences between the three ways of calculating the score of the final list
- CBIR: Non Automatic Relevance feedback algorithm based on logistic regression
- TBIR: Why NEs does not improve previous results?
- Textual enrichment (WP articles summaries)