Evaluation of ontology matching

Jérôme Euzenat (INRIA Rhône-Alpes & LIG)
+ work within Knowledge web 2.2
and esp. Malgorzata Mochol (FU Berlin)

April 19, 2007
Outline

1. Heterogeneity, matching and evaluation
2. Experimental work: OAEI 2006
3. Theoretical work: evaluation measures
4. Methodological work: matcher selection
5. General conclusion
1. Heterogeneity, matching and evaluation

2. Experimental work: OAEI 2006

3. Theoretical work: evaluation measures

4. Methodological work: matcher selection

5. General conclusion
Matching

- Employee
  - Marketing
  - Computer
  - Optics
  - Administration
    - Accounting
  - Production
    - Electronics

Heterogeneity, matching and evaluation
Experimental work: OAEI 2006
Theoretical work: evaluation measures
Methodological work: matcher selection
General conclusion

Evaluation of ontology matching
Matching

Employee
  - Marketing
    - Computer
    - Optics
  - Administration
    - Accounting
  - Production
    - Electronics

Worker
  - Spain
    - Salesforce
    - Headquarters
  - Japan

$R = OAEI 2006$
Matching

Employee

Marketing

Computer

Optics

Administration

Accounting

Production

Electronics

Worker

Spain

Saleforce

Headquarters

Japan
Alignments

Definition (Alignment, correspondence)

Given two ontologies $o$ and $o'$, an alignment between $o$ and $o'$ is a set of correspondences (i.e., 4-uples): $\langle e, e', r, n \rangle$ with

- $e \in o$ and $e' \in o'$ being the two matched entities,
- $r$ being a relationship holding between $e$ and $e'$, and
- $n$ expressing the level of confidence $[0..1]$ in this correspondence.
Goal of evaluation

- Improving the performances of the ontology matching field...
  through the comparison of algorithms...
  on various sets of tests:

We created the Ontology Alignment Evaluation Initiative (OAEI) which aims at evaluating the various matching systems available on the model of TREC; Campaigns have been run in 2004, 2005 and 2006; http://oaei.ontologymatching.org
Goal of evaluation

▶ Improving the performances of the ontology matching field ...

... through the comparison of algorithms ...

... on various sets of tests:

▶ We created the Ontology Alignment Evaluation Initiative (OAEI) which aims at evaluating the various matching systems available on the model of TREC;
▶ Campaigns have been run in 2004, 2005 and 2006;
▶ http://oaei.ontologymatching.org
General setting:

▶ From a set of pairs of ontologies (in OWL) $o$ and $o'$

... use one automatic system...

... with the same set of parameters...

... to output an alignment $A$ (in the ontology alignment format).

▶ All general purpose resources authorized.

$A$ will be compared with some reference alignment $R$. 
Precision and recall

\[ o \times o' \times Q \]
Precision and recall

\[ \text{Precision} = \frac{|R \cap A|}{|A|} \]

\[ \text{Recall} = \frac{|R \cap A|}{|R|} \]

\( A \times o \times o' \times Q \)
Precision and recall

\[
\text{precision} (A, R) = \frac{|R \cap A|}{|A|}
\]

\[
\text{recall} (A, R) = \frac{|R \cap A|}{|R|}
\]
Precision and recall

Definition (Precision, Recall)

Given a reference alignment \( R \), the **precision** of some alignment \( A \) is given by

\[
P(A, R) = \frac{|R \cap A|}{|A|}
\]

and **recall** is given by

\[
R(A, R) = \frac{|R \cap A|}{|R|}
\]
Outline

1. Heterogeneity, matching and evaluation

2. Experimental work: OAEI 2006

3. Theoretical work: evaluation measures

4. Methodological work: matcher selection

5. General conclusion
Heterogeneity, matching and evaluation
Experimental work: OAEI 2006
Theoretical work: evaluation measures
Methodological work: matcher selection
General conclusion

OAEI-2006

▶ Held at the Ontology matching workshop of ISWC.
▶ Featured 6 test cases for 10 participants.

Preparation  June 1st-July 1st;
Execution    July 1st-September 1st;
Evaluation   August 15th + September 1st-7th;
Many proposals for evaluation test beds, not that many for participating.
Test set

- The usual test set;
- based on a bibliography ontology in OWL-DL in RDF/XML featuring reference to outer ontologies;
- containing 33 named classes, 24 object properties, 40 data properties, 56 named individuals and 20 anonymous individuals;
- reference 1-1 alignments with “=” relation and 1. confidence;
- three group of tests: simple (4), systematic (46), and real-life (4);
- systematically altered by combining the transformations on names, comments, instances, classes, properties.
- this year we test new, more tolerant, measures. May be semantic measures next year.
### Results (precision and recall)

<table>
<thead>
<tr>
<th>algo</th>
<th>automs</th>
<th>coma</th>
<th>DSSim</th>
<th>falcon</th>
<th>hmatch</th>
<th>jhuapl</th>
<th>OCM</th>
<th>prior</th>
<th>RiMOM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1xx</td>
<td>.94</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
<td>.98</td>
<td>1.</td>
<td>1.</td>
<td>.91</td>
<td>1.</td>
</tr>
<tr>
<td>2xx</td>
<td>.94</td>
<td>.89</td>
<td>.99</td>
<td>.97</td>
<td>.99</td>
<td>.67</td>
<td>.97</td>
<td>.94</td>
<td>.70</td>
</tr>
<tr>
<td>3xx</td>
<td>.91</td>
<td>.70</td>
<td>.84</td>
<td>.69</td>
<td>.90</td>
<td>.78</td>
<td>.89</td>
<td>.78</td>
<td>.57</td>
</tr>
<tr>
<td>Total</td>
<td>.94</td>
<td>.67</td>
<td>.96</td>
<td>.83</td>
<td>.98</td>
<td>.55</td>
<td>.92</td>
<td>.86</td>
<td>.55</td>
</tr>
<tr>
<td>Sym.</td>
<td>.94</td>
<td>.68</td>
<td>.96</td>
<td>.83</td>
<td>.99</td>
<td>.55</td>
<td>.94</td>
<td>.89</td>
<td>.56</td>
</tr>
<tr>
<td>Eff.</td>
<td>.94</td>
<td>.68</td>
<td>.96</td>
<td>.83</td>
<td>.99</td>
<td>.55</td>
<td>.94</td>
<td>.89</td>
<td>.56</td>
</tr>
<tr>
<td>Orient.</td>
<td>.94</td>
<td>.68</td>
<td>.96</td>
<td>.83</td>
<td>.99</td>
<td>.55</td>
<td>.94</td>
<td>.89</td>
<td>.56</td>
</tr>
</tbody>
</table>
Food test set

- 31112 mappings, all exactMatch

![Diagram showing the comparison between AGROVOC and NALT with 28174 exactMatch mappings and 41577 mappings, respectively.]

- Submitted by 1 participant: 28174
- Submitted by 2 participants: 10028
- Submitted by 3 participants: 24525
- Submitted by 4 participants: 0
- Submitted by all participants: 0

Thanks: Willem Robert van Hage

Jérôme Euzenat (INRIA Rhône-Alpes & LIG) + work within Knowledge web 2.2 and esp. Malgorzata Mochol (FU Berlin)
Food example 1

**AGROVOC**

- Rodentia
  - Sigmodon
  - Dormice

**NALT**

- Rodentia
  - Myoxidae
  - Muridae
  - Sigmodon

- There is an exactMatch relationship between Sigmodon in AGROVOC and Sigmodon in NALT.
- There is a broader relationship between Rodentia in NALT and Rodentia in AGROVOC.
Food example 2

AGROVOC

British Isles

broader

broadMatch

Northern Ireland

broadMatch

United Kingdom

narrowMatch

exactMatch

Ireland

broader

broadMatch

Northern Ireland

broadMatch

Irish Republic

broader

British Isles

broader

broadMatch

exactMatch

exactMatch

NALT

Evaluation of ontology matching
Results

<table>
<thead>
<tr>
<th>test</th>
<th>falcon</th>
<th>hmatch</th>
<th>dssim</th>
<th>coma</th>
<th>automs</th>
<th>jhuapl</th>
<th>prior</th>
<th>RiMOM</th>
<th>OCM</th>
<th>nih</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>benchmark</td>
<td>P</td>
<td>.92</td>
<td>.84</td>
<td>.98</td>
<td>.96</td>
<td>.94</td>
<td>.22</td>
<td>.95</td>
<td>.96</td>
<td>.93</td>
<td>9</td>
</tr>
<tr>
<td>benchmark</td>
<td>R</td>
<td>.89</td>
<td>.55</td>
<td>.55</td>
<td>.83</td>
<td>.67</td>
<td>.85</td>
<td>.63</td>
<td>.88</td>
<td>.55</td>
<td>9</td>
</tr>
<tr>
<td>anatomy</td>
<td>P</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5</td>
</tr>
<tr>
<td>anatomy</td>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>5</td>
</tr>
<tr>
<td>jobs</td>
<td>P</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6</td>
</tr>
<tr>
<td>jobs</td>
<td>R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>6</td>
</tr>
<tr>
<td>directory</td>
<td>(P)</td>
<td>.40</td>
<td>.32</td>
<td>.31</td>
<td>.31</td>
<td>.34</td>
<td>.39</td>
<td>.33</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>directory</td>
<td>(R)</td>
<td>.45</td>
<td>.13</td>
<td>.27</td>
<td>.15</td>
<td>.24</td>
<td>.40</td>
<td>.16</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>food</td>
<td>P</td>
<td>.83</td>
<td>.61</td>
<td>.54</td>
<td>.71</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>food</td>
<td>(R)</td>
<td>.65</td>
<td>.65</td>
<td>.33</td>
<td>.64</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>conference</td>
<td>P</td>
<td>.68</td>
<td>.45</td>
<td>.34</td>
<td>.90</td>
<td>.38</td>
<td>.68</td>
<td>.68</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>conference</td>
<td>R</td>
<td>.50</td>
<td>.48</td>
<td>.54</td>
<td>.20</td>
<td>.62</td>
<td>.29</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

No validation has been done this year
Precision/recall profiles for benchmarks

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>refalign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edna</td>
<td></td>
<td></td>
</tr>
<tr>
<td>coma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>falcon2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>falcon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hmatch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>prior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RiMOM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comments

- Again, more participants than last year;
- More systems at the top, but the top did not increased its score;
- We took into account most of the remarks of last year;
- The time devoted to this experiment (3 month during summer) was, once again, too short;
- OAEI has acquired some reputation;
- The fact that it has been set independently of Knowledge web and its success are indices of its sustainability.
Outline

1. Heterogeneity, matching and evaluation
2. Experimental work: OAEI 2006
3. Theoretical work: evaluation measures
4. Methodological work: matcher selection
5. General conclusion
Example

- **Employee**
  - Marketing
  - Computer
  - Optics
  - Administration
  - Accounting
  - Production
  - Electronics

- **Worker**
  - Spain
  - Saleforce
  - Headquarters
  - Japan

- **Relationships**
  - Employee = Worker
  - Marketing ≤ Spain
  - Computer = Saleforce
  - Optics ≤ Headquarters
  - Administration ≤ Japan
  - Accounting
  - Production
  - Electronics
Heterogeneity, matching and evaluation
Experimental work: OAEI 2006
Theoretical work: evaluation measures
Methodological work: matcher selection
General conclusion

Example

**Employee**
- Marketing
- Computer
- Optics
- Administration
- Accounting
- Production
- Electronics

**Worker**
- Spain
- Saleforce
- Headquarters
- Japan

\[ P = \frac{2}{4} = 0.5 \]
\[ R = \frac{2}{4} = 0.5 \]

\[ A_1 \]
Example

Employee
  - Marketing
    - Computer
    - Optics
  - Administration
  - Accounting
  - Production
  - Electronics

Worker
  - Spain
  - Salesforce
  - Headquarters
  - Japan

Evaluation of ontology matching

\[ P = \frac{2}{6} = 0.33 \]
\[ R = \frac{2}{4} = 0.5 \]
Heterogeneity, matching and evaluation
Experimental work: OAEI 2006
Theoretical work: evaluation measures
Methodological work: matcher selection

General conclusion

Example

Example:

Employee
-->
Marketing
-->
Computer
-->
Optics
-->
Administration
-->
Accounting
-->
Production
-->
Electronics

Worker
-->
Spain
-->
Saleforce
-->
Headquarters
-->
Japan

Equality relationships:

Employee = Worker
Marketing ≤ Computer
Optics ≥ Administration
Accounting ≤ Production
Electronics = Headquarters

Inequality relationships:

Spain ≤ Saleforce

Evaluation:

\[ P = \frac{2}{4} = 0.5 \]
\[ R = \frac{2}{4} = 0.5 \]
Problems with precision and recall

- P and R do not make a difference between a nearly good alignment ($A_1$) and a bad one ($A_4$).

$\Rightarrow$ They lack tolerance.
Problems with precision and recall

- P and R do not make a difference between a nearly good alignment ($A_1$) and a bad one ($A_4$).

$\Rightarrow$ They lack tolerance.

- P and R do not recognise two equivalent alignments;
- P and R are not degrees of correctness and completeness.
- P and R do not make a difference between a correct alignment ($A_1$) and an incorrect one ($A_4$).

$\Rightarrow$ They lack semantics.
Heterogeneity, matching and evaluation  
Experimental work: OAEI 2006  
Theoretical work: evaluation measures  
Methodological work: matcher selection  
General conclusion

Tolerent precision and recall [Ehrig 2005]

- Generalises classical P/R by using a similarity function $\omega$ instead of $|A \cap R|$
- Three concrete such extensions proposed.

Definition (Generalised precision and recall)

Given a reference alignment $R$ and an overlap function $\omega$ between alignments, the precision of an alignment $A$ is given by

$$P_\omega(A, R) = \frac{\omega(A, R)}{|A|}$$

and recall is given by

$$R_\omega(A, R) = \frac{\omega(A, R)}{|R|}.$$ 

such that $|A \cap R| \leq \omega(A, R) \leq \min(|A|, |R|)$
Semantic precision and recall [Euzenat 2007]

- Generalises classical P/R by interpreting alignments with the semantics introduced in [Zimmermann 2006];
- Can be fully seen as degrees of correctness and completeness.

**Definition (Semantic precision and recall)**

Given a reference alignment $R$, the precision of some alignment $A$ is given by

$$P_s(A, R) = \frac{|A \cap Cn(R)|}{|A|}$$

and recall is given by

$$R_s(A, R) = \frac{|Cn(A) \cap R|}{|R|}$$
We used the tolerant versions in OAEI-2006 and they provided results as expected, but did not change the order between systems.
Outline

1. Heterogeneity, matching and evaluation
2. Experimental work: OAEI 2006
3. Theoretical work: evaluation measures
4. Methodological work: matcher selection
5. General conclusion
Problem

- How to provide potential users, with a particular application, advice on which matcher to use?
- Cross application needs and matcher characteristics.
- We followed two approaches reported in Deliverable 1.2.6 (a.k.a. 1.2.2.2.1).
- This is joint work with WP 1.2.
Shallow approach

Applications →

Assess type of application needs through literature review
Shallow approach

Applications → Assess type of application needs through literature review

Systems → Assess system capabilities through OAEI evaluation results (particularly benchmark tests)

Jerôme Euzenat (INRIA Rhône-Alpes & LIG) + work within Knowledge web 2.2 and esp. Malgorzata Mochol (FU Berlin)
Shallow approach

Applications → Assess type of application needs through literature review

Systems → Assess system capabilities through OAEI evaluation results (particularly benchmark tests)

Use the needs to weight the criteria and aggregate system performances → Selection
In-depth approach

Applications → Assess application requirements through questionnaires;
In-depth approach

Applications → Assess application requirements through questionnaires;

Systems → Assess system capabilities through questionnaires to developers.
In-depth approach

Applications

Assess application requirements through questionnaires;

Systems

Assess system capabilities through questionnaires to developers

Apply a multi-criteria decision making technique.

Selection
In-depth approach

Applications

Assess application requirements through questionnaires;

Systems

Assess system capabilities through questionnaires to developers

Apply a multi-criteria decision making technique. → Selection
Analytic Hierarchy Process (AHP)

- Methodology for supporting a (multi-criteria) decision making process
- Sets priorities
- Reduces complex decisions to a series of pair-wise comparisons
Analytic Hierarchy Process (AHP)

- Methodology for supporting a (multi-criteria) decision making process
- Sets priorities
- Reduces complex decisions to a series of pair-wise comparisons

AHP-steps

1. define the problem or the project objectives
2. build a hierarchy of decision
3. data collection
4. build a pairwise comparison
5. calculate the final result
Step 1: Problem definition

Which matching approach is suitable with respect to the given (application) requirements?
Step 2: Hierarchy of decision

0 Level:
Problem (Goal)

FIND A SUITABLE APPROACH
Step 2: Hierarchy of decision

0 Level: Problem (Goal)

1st Level: Dimensions

- APPROACH
- INPUT
- USAGE
- OUTPUT
- COSTS
- DO
Step 2: Hierarchy of decision

0 Level: Problem (Goal)

1st Level: Dimensions
- APPROACH
- INPUT
- USAGE
- OUTPUT
- COSTS
- DO

2nd Level: Factors

FIND A SUITABLE APPROACH
Step 2: Hierarchy of decision

0 Level: Problem (Goal)

1\textsuperscript{st} Level: Dimensions
- APPROACH
- INPUT
- USAGE
- OUTPUT
- COSTS
- DO

2\textsuperscript{nd} Level: Factors

3\textsuperscript{rd} Level: Attributes
Step 2: Hierarchy of decision

0 Level: Problem (Goal)

1st Level: Dimensions
- APPROACH
- INPUT
- USAGE
- OUTPUT
- COSTS
- DO

2nd Level: Factors

3rd Level: Attributes

4th Level: Alternatives
- Matcher 1
- Matcher 2
- ...
Multilevel Characteristic for Matching Approaches

1\textsuperscript{st} Level: DIMENSIONS

2\textsuperscript{nd} Level: Factors

3\textsuperscript{rd} Level: Attributes
Multilevel Characteristic for Matching Approaches

1st Level: DIMENSIONS

Input size

2nd Level: Factors

Formality level

Input category

3rd Level: Attributes

number of ontologies

number of primitives

formal

semiformal

informal

taxonomy

ontology
Step 3: Data collection

- **Collection of the relevant information** about the particular matching approaches (alternatives)

- **Development of an online questionnaire**
  - to be fill out by the matching experts / matcher developers
  - rating of the matching alternatives

- **All available alternatives automatically weighted against the new added approach (pairwise comparison of the alternatives)**
Step 4: Pairwise comparison

- **Definition of the application requirements**
  - weighting of the criteria
    → one criteria is more relevant than the other concerning the system specification
  - each level of criteria build a pairwise comparison between the sibling nodes
    → weighting of attributes against attributes, factors against factors, dimensions against dimensions

- **AHP-Tool for user-friendly pairwise comparison of the criteria**
Step 4: Pairwise comparison - AHP Tool

Multilevel characteristics for matching approaches

Matching Algorithm Evaluation
- INPUT CHARACTERISTIC (SIZE)
- USAGE CHARACTERISTIC
- APPROACH CHARACTERISTIC
- INPUT CHARACTERISTIC
  - Input category
    - Formality level
      - informal
      - semi-formal
      - formal
  - Model type
  - Input type
  - External sources
  - Input NL
  - Input structure
  - Input rep. langu
- COSTS CHARACTERISTIC
- OUTPUT CHARACTERISTIC
- DOCUMENTATION CHARACTERISTIC
Step 4: Pairwise comparison - AHP Tool

The slide shows a hierarchical structure with various characteristics for matching approaches. The characteristics include:

- Input Characteristic (Size)
- Usage Characteristic
- Approach Characteristic
- Input Characteristic

Each characteristic is further divided into sub-characteristics, such as Formality level, which is further divided into:

- Informal
- Semi-formal
- Formal

The pairwise comparison is indicated by arrows connecting the sub-characteristics. The multilevel characteristics for matching approaches are highlighted in the slide.
Step 5: Final results

- Decision regarding the determination of the suitable matching approach is based on the ranking of a matcher alternative.

- The ranking reflects the global importance of the approach according to the alternative weightings performed in step 3 and criteria weightings from step 4.

- The higher a matcher alternative is weighted for various criteria, the higher the priority of the particular approach in the entire ranking.
Matcher Selection - Review

Step 1 Problem definition: Which matching approach is currently relevant and suitable with respect to the given requirements?
Matcher Selection - Review

**Step 1 Problem definition:** Which matching approach is currently relevant and suitable with respect to the given requirements?

**Step 2 Hierarchy of decision**

(usage of the multilevel characteristic for matching approaches)

- **0 Level:** Problem (Goal)
- **1st Level:** Dimensions
  - INPUT
  - APPROACH
  - USAGE
  - OUTPUT
  - COSTS
  - DOC
- **2nd Level:** Factors
- **3rd Level:** Attributes
- **4th Level:** Alternatives
  - Matcher 1
  - Matcher 2
  - ...
  - Matcher n
Matcher Selection - Review

Step 1 Problem definition: Which matching approach is currently relevant and suitable with respect to the given requirements?

Step 2 Hierarchy of decision
(usage of the multilevel characteristic for matching approaches)

Step 3 Matcher-data collection
http://matching.ag-nbi.de
Matcher Selection - Review

Step 1 Problem definition: Which matching approach is currently relevant and suitable with respect to the given requirements?

Step 2 Hierarchy of decision (usage of the multilevel characteristic for matching approaches)

0 Level: Problem (Goal)

1st Level: Dimensions

INPUT

APPROACH

USAGE

OUTPUT

COSTS

DOC

FIND A SUITABLE APPROACH

1st Level: Factors

2nd Level: Attributes

3rd Level: Alternatives

Matcher 1

Matcher 2

... Matcher n

Step 3 Matcher-data collection

matching online questionnaire

Questionnaire Database

http://matching.ag-nbi.de

Domain experts

Users of the matching approaches

Jérôme Euzenat (INRIA Rhône-Alpes & LIG) + work within Knowledge web 2.2 and esp. Malgorzata Mochol (FU Berlin)

Evaluation of ontology matching
Matcher Selection - Review

Step 1 Problem definition: Which matching approach is currently relevant and suitable with respect to the given requirements?

Step 2 Hierarchy of decision

0 Level: Problem (Goal)

1st Level: Dimensions
- INPUT
- APPROACH
- USAGE
- OUTPUT
- COSTS
- DOC

2nd Level: Factors

3rd Level: Attributes

4th Level: Alternatives

MATCHING APPROACHES

Step 3 Matcher-data collection

Step 4 Pairwise comparison (definition of requirements w.r.t the desired matcher & pairwise comparison of the importance of the requirements)

Step 5 Calculation of the final result (ranked list of suitable matchers)

http://matching.ag-nbi.de
Findings

- We applied this to the Knowledge web use case #1: worldwidejobs recruitment portal.
- We found relatively different advices...
We applied this to the Knowledge web use case #1: worldwidejobs recruitment portal.

We found relatively different advices...

This remains to be analysed;

... but an be related to the difference in data (source, granularity) and approaches (weights).
**Towards a methodology**

**superficial** Use the table for identifying the profile of the application either to select the best suited matcher through weighted aggregation or to select a subset of matchers on which to perform a deep analysis.

**deep** Use the AHP tool, input the detailed criterion preferences in order to find the matching system the closest to the requirements.

**involved** Instrument the application in order to carry out application specific evaluation. This is a very costly approach however.
Outline

1. Heterogeneity, matching and evaluation
2. Experimental work: OAEI 2006
3. Theoretical work: evaluation measures
4. Methodological work: matcher selection
5. General conclusion
Lesson learned

- It is possible to evaluate matching systems;
- We can measure quality increase;
- More and more tools, and more robust;
- Not sure what a good test case is (a lot of opinions about what a bad one is);
- It remains difficult to use these results to find an adequate system.
http://exmo.inrialpes.fr
http://oaei.ontologymatching.org
http://book.ontologymatching.org