Jérôme Euzenat and Pavel Shvaiko

Ontology matching

Solutions to problems

(Fist edition)

Springer
We provide sketches of the solutions to the exercises of Appendix C. Most computations are made with the help of Alignment API (version 4.1). The book web site\(^1\) provides scripts and Java classes for using the Alignment API in order to solve some of the questions.

### D.1 Applications

**Solution D.1 (Application definition).**

1. A possible architecture for the application is provided by the following drawing:

   ![Diagram]

   - The requirements of the application for ontology matching appears to be:
     - instances: YES, there are instances available.

\(^1\) [http://book.ontologymatching.org](http://book.ontologymatching.org)
Appendix D: Solutions

- run-time: No, matching is possible at design time. However, since it is said that the database keeps evolving, there is still come alignment maintenance to be carried out.
- automatic: NO, automatic matching is not strictly required and it may be wise to have results checked before putting them in production.
- correct: YES, the alignment must be correct. It would be particularly embarrassing, that it provides incorrect information.
- complete: Yes, the more complete the results, the best. However, having incomplete matching is not critical, this means that some information will be missing in the unified view.
- operation: the operation to be carried out is query mediation because the databases have to remain independent.

D.2 The matching problem

Solution D.2 (Ontology representation). Let o be a first data source to be integrated. o is described in English as follows:

- People are divided among Students, Faculty and Staff. Faculty is further divided depending on its departments and sub-departments, e.g., Philosophy, Science.
- People are characterised by their firstname and lastname which are strings, id which is a uri and birthdate which is a date.
- Students will attend Courses which are taught by Faculty and Faculty people have an Office as a room.
- Pr. Carla Cipolla is a visiting professor in Computer science and Stefano Zucchini is a PhD Student.

1. Provide a description for o as a folksonomy:

```
Course   People
Science

Student   Staff

Pr. Carla Cipolla

Office

Stefano Zucchini

Philosophy
```

2. Provide a description for o as a directory:
3. Here is a description of $\sigma$ as an XML schema:

```xml
<schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
        xmlns="http://www.w3.org/2001/XMLSchema">
  <complexType name="People">
    <sequence>
      <element name="firstname" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <element name="lastname" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <element name="birthdate" type="xsd:date" minOccurs="1" maxOccurs="1"/>
      <element name="id" type="xsd:anyURI" minOccurs="1" maxOccurs="1"/>
      <element name="type" type="xsd:string"/>
    </sequence>
  </complexType>

  <complexType name="Student">
    <complexContent>
      <extension base="People">
        <sequence>
          <element name="attends" type="Course"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>

  <complexType name="Faculty">
    <complexContent>
      <extension base="People">
        <sequence>
          <element name="office" type="Room" maxOccurs="1"/>
          <element name="department" type="Department" minOccurs="1" maxOccurs="1"/>
          <element name="teaches" type="Course"/>
        </sequence>
      </extension>
    </complexContent>
  </complexType>

  <complexType name="Philosophy">
    <complexContent>
      <extension base="Faculty"/>
    </complexContent>
  </complexType>

  <complexType name="Science">
    <complexContent>
      <extension base="Faculty"/>
    </complexContent>
  </complexType>

  <complexType name="Course">
    <sequence>
      <element name="title" type="xsd:string" minOccurs="1" maxOccurs="1"/>
      <element name="room" type="Room" minOccurs="1" maxOccurs="1"/>
    </sequence>
  </complexType>
</schema>
```
The two instances would be described by the following XML file:

```xml
<?xml version="1.0" ?>
<Student>
  <firstname>Stefano</firstname>
  <lastname>Zucchini</lastname>
  <birthdate>16/04/1987</birthdate>
  <id>1178</id>
  <type>PhD student</type>
</Student>

<Science>
  <firstname>Carla</firstname>
  <lastname>Cipolla</lastname>
  <birthdate>13/06/1977</birthdate>
  <id>1998</id>
  <type>Visiting professor</type>
  <departement>
    <Department><name>Computer science</name></Department>
  </departement>
</Science>
</xml>

4. Here is a description of o as an SQL relational database schema:

```sql
CREATE TABLE People (  
type VARCHAR(20),
department VARCHAR(40),
firstname VARCHAR(30),
lastname VARCHAR(30),
birthdate DATE,
id URI,
office INT,
PRIMARY KEY (firstname lastname birthdate id)  
);

CREATE TABLE Department (  
name VARCHAR(40),
PRIMARY KEY (name)  
);

CREATE TABLE Room (  
number INT,
PRIMARY KEY (number)  
);

CREATE TABLE Course (  
title VARCHAR(100),
id VARCHAR(10),
```
D.2 The matching problem

```sql
CREATE TABLE teaches (
  course INT,
  firstname VARCHAR(30),
  lastname VARCHAR(30),
  birthdate DATE,
  id URI
);

CREATE TABLE attends (
  course INT,
  firstname VARCHAR(30),
  lastname VARCHAR(30),
  birthdate DATE,
  id URI
);

INSERT INTO People (firstname, lastname, birthdate, id, type, department) VALUES ('Carla', 'Cipolla', '13/06/1977', 1998, 'Visiting professor', 'Computer science');

INSERT INTO People (firstname, lastname, birthdate, id, type) VALUES ('Stefano', 'Zucchini', '16/04/1987', 1178, 'PhD student');
```

5. Figure D.1 is a description for o as a UML class diagram. Instances do not appear there.

![UML class diagram](image-url)

**Fig. D.1.** Representation of o as a UML class diagram.

6. Fig. D.2 is a description of o as an ontology.

**Solution D.3 (Ontology semantics).** Assume that the left hand side ontology of Fig. C.1 is denoted as o.
Appendix D: Solutions

1. $o$ can be expressed in OWL as follows:

```xml
<owl:Class rdf:about="#People">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#firstname"/>
      <owl:allValuesFrom rdf:resource="xsd:string"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#lastname"/>
      <owl:allValuesFrom rdf:resource="xsd:string"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#id"/>
      <owl:allValuesFrom rdf:resource="xsd:uri"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#birthdate"/>
      <owl:allValuesFrom rdf:resource="xsd:date"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#Student">
  <rdfs:subClassOf rdf:resource="#People"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#attends"/>
      <owl:allValuesFrom rdf:resource="#Course"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:about="#Faculty">
  <owl:Class rdf:about="#Staff">
  </owl:Class>
</owl:Class>
```

---

**Fig. D.2.** Representation of $o$ as an ontology.
2. \( o \) can be expressed in the simplified description logic format as follows:

\[
\begin{align*}
\text{People} & \leq \forall \text{firstname}.\text{string} & \text{People} & \leq \forall \text{lastname}.\text{string} \\
\text{People} & \leq \forall \text{id}.\text{uri} & \text{People} & \leq \forall \text{birthdate}.\text{date} \\
\text{Student} & \leq \text{People} & \text{Student} & \leq \forall \text{attends}.\text{Course} \\
\text{Faculty} & \leq \text{People} & \text{Faculty} & \leq \forall \text{teaches}.\text{Course} & \text{Faculty} & \leq \forall \text{room}.\text{Office} \\
\text{Science} & \leq \text{Faculty} & \text{Computer science} & \leq \text{Science} \\
\text{Biology} & \leq \text{Science} & \text{Philosophy} & \leq \text{Faculty} \\
\text{Boxology} & \leq \text{Faculty} & \text{Staff} & \leq \text{People} \\
\text{Stefano Zucchini} & \in \text{Student} & \text{Pr. Carla Cipolla} & \in \text{Computer science}
\end{align*}
\]

3. The models of \( o \) are those interpretations \( m \) over a domain \( D \) such that:
Appendix D: Solutions

\[ m(\text{teaches}) \subseteq D \times D \]
\[ m(\text{attends}) \subseteq D \times D \]
\[ m(\text{People}) \subseteq D \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{firstname}), y \in S \} \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{lastname}), y \in S \} \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{id}), y \in U \} \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{birthdate}), y \in D \} \]
\[ m(\text{Student}) \subseteq m(\text{People}) \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{attends}), y \in m(\text{Course}) \} \]
\[ m(\text{Faculty}) \subseteq m(\text{People}) \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{teaches}), y \in m(\text{Course}) \} \]
\[ \cap \{ x \in D; \forall y; (x, y) \in m(\text{room}), y \in m(\text{Office}) \} \]
\[ m(\text{Science}) \subseteq m(\text{Faculty}) \]
\[ m(\text{Computer science}) \subseteq m(\text{Science}) \]
\[ m(\text{Biology}) \subseteq m(\text{Science}) \]
\[ m(\text{Philosophy}) \subseteq m(\text{Faculty}) \]
\[ m(\text{Boxology}) \subseteq m(\text{Faculty}) \]
\[ m(\text{Staff}) \subseteq m(\text{People}) \]
\[ m(\text{Stefano Zucchini}) \in m(\text{Student}) \]
\[ m(\text{Pr. Carla Cipolla}) \in m(\text{Computer science}) \]

4. \( o \models \text{Philosophy} \leq \text{People} \) because for any model \( m \) of \( o \), \( m(\text{Philosophy}) \subseteq m(\text{Faculty}) \) and \( m(\text{Faculty}) \subseteq m(\text{People}) \), hence \( m(\text{Philosophy}) \subseteq m(\text{People}) \).

5. \( o \not\models \text{teaches} \leq \text{attends} \) because the interpretation \( I \) with one course \( c \), one student \( s \) and one faculty \( f \) such that \( I(\text{People}) = \{s, t\} \), \( I(\text{Course}) = \{c\} \), \( I(\text{Student}) = \{s\} \), \( I(\text{Computer science}) = I(\text{Science}) = I(\text{Faculty}) = \{f\} \), \( I(\text{Stefano Zucchini}) = s \), \( I(\text{Pr. Carla Cipolla}) = f \), \( I(\text{attends}) = \{(s, c)\} \), \( I(\text{teaches}) = \{(f, c)\} \) all other terms having an empty interpretation, is a model of the ontology and it does not satisfy the assertion.

D.3 Classification

Solution D.4 (Kinds of techniques). It seems that all the techniques considered in our classifications can be used:

- terminological: because the two ontologies are expressed in the same language. However, the poor overlap between the two sets of terms can be the sign of a poor overlap between the ontologies or require lexicon-based methods;
- structural: because there are constraints on properties, but we need anchors;
D.4 Basic techniques

Solution D.5 (Name-based distance computation). Given the ontologies \( o \) and \( o' \) of Problem C.1 as illustrated in Fig. C.1.

1. Here are the distances or similarity between all the labels occurring in Fig. C.1 obtained with the OntoSim library (version 2.1)\(^2\)

<table>
<thead>
<tr>
<th></th>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Botany</th>
<th>People</th>
<th>Biology</th>
<th>Psychology</th>
<th>Computer/Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>0.87</td>
<td>0.88</td>
<td>0.75</td>
<td>0.57</td>
<td>0.87</td>
<td>1.00</td>
<td>1.00</td>
<td>0.88</td>
<td>0.89</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>0.87</td>
<td>0.50</td>
<td>0.88</td>
<td>0.75</td>
<td>0.86</td>
<td>0.73</td>
<td>1.00</td>
<td>0.87</td>
<td>1.00</td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.84</td>
<td>0.74</td>
<td>0.83</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.86</td>
<td>0.94</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.91</td>
<td>0.83</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.85</td>
<td>0.94</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>0.83</td>
<td>0.92</td>
<td>0.92</td>
<td>0.83</td>
<td>0.73</td>
<td>0.91</td>
<td>1.00</td>
<td>0.91</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>0.91</td>
<td>0.92</td>
<td>0.83</td>
<td>0.83</td>
<td>0.73</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>Professor</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
<td>0.86</td>
<td>0.73</td>
<td>0.88</td>
<td>0.87</td>
<td>0.88</td>
<td>0.79</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Room</td>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
<td>0.82</td>
<td>0.86</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>PhDStudent</td>
<td>0.88</td>
<td>0.76</td>
<td>0.88</td>
<td>0.18</td>
<td>0.73</td>
<td>0.88</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
<td>0.80</td>
<td>0.84</td>
</tr>
<tr>
<td>Lecture</td>
<td>0.69</td>
<td>0.86</td>
<td>0.86</td>
<td>0.71</td>
<td>0.83</td>
<td>0.85</td>
<td>0.87</td>
<td>0.69</td>
<td>0.86</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.89</td>
<td>0.90</td>
<td>0.80</td>
<td>0.90</td>
<td>0.89</td>
<td>0.79</td>
<td>0.90</td>
<td>0.89</td>
<td>0.90</td>
<td>0.83</td>
<td>0.93</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.92</td>
<td>0.84</td>
<td>0.92</td>
<td>0.91</td>
<td>0.83</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.86</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>Staff</td>
<td>0.82</td>
<td>0.83</td>
<td>0.83</td>
<td>0.67</td>
<td>0.00</td>
<td>0.64</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.87</td>
<td>0.90</td>
</tr>
</tbody>
</table>

\(^2\) \url{http://ontosim.gforge.inria.fr}
### Appendix D: Solutions

- **3-gram distance:**

<table>
<thead>
<tr>
<th></th>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Reference</td>
<td>1.00</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Professor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Room</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PhDStudent</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.23</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Lecture</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

There is not much in common between the trigrams of these labels: trigrams are more effective on longer texts.

- **edit distance:**

<table>
<thead>
<tr>
<th></th>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>0.89</td>
<td>0.78</td>
<td>0.89</td>
<td>0.67</td>
<td>0.67</td>
<td>0.89</td>
<td>1.00</td>
<td>0.89</td>
<td>0.90</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>Reference</td>
<td>0.78</td>
<td>0.56</td>
<td>0.89</td>
<td>0.78</td>
<td>1.00</td>
<td>0.67</td>
<td>1.00</td>
<td>0.78</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>0.89</td>
<td>0.78</td>
<td>0.94</td>
<td>0.83</td>
<td>0.78</td>
<td>0.89</td>
<td>0.94</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>0.88</td>
<td>0.82</td>
<td>1.00</td>
<td>0.82</td>
<td>0.82</td>
<td>0.88</td>
<td>0.94</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.94</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>0.88</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.84</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>0.88</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.88</td>
</tr>
<tr>
<td>Professor</td>
<td>0.78</td>
<td>0.89</td>
<td>1.00</td>
<td>0.89</td>
<td>0.89</td>
<td>0.78</td>
<td>0.89</td>
<td>0.78</td>
<td>0.89</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>Room</td>
<td>0.83</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.83</td>
<td>0.71</td>
<td>0.80</td>
<td>0.93</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>PhDStudent</td>
<td>0.90</td>
<td>0.80</td>
<td>0.90</td>
<td>0.30</td>
<td>0.80</td>
<td>0.90</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Lecture</td>
<td>0.57</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>0.86</td>
<td>1.00</td>
<td>0.71</td>
<td>1.00</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.85</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.83</td>
<td>0.72</td>
<td>0.89</td>
<td>0.83</td>
<td>0.83</td>
<td>0.89</td>
<td>0.89</td>
<td>0.83</td>
<td>0.89</td>
<td>0.89</td>
<td>0.78</td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
<td>0.71</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>0.93</td>
</tr>
</tbody>
</table>
- Jaro–Winckler measure:

<table>
<thead>
<tr>
<th></th>
<th>Assistant</th>
<th>Reference</th>
<th>AssistantProfessor</th>
<th>VisitingProfessor</th>
<th>ResearchAssistant</th>
<th>TeachingAssistant</th>
<th>Professor</th>
<th>Room</th>
<th>PhDStudent</th>
<th>Lecture</th>
<th>FullProfessor</th>
<th>AssociateProfessor</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>0.72</td>
<td>0.75</td>
<td>0.68</td>
<td>0.67</td>
<td>0.78</td>
<td>0.74</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.89</td>
<td>1.00</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>0.84</td>
<td>0.75</td>
<td>0.78</td>
<td>0.80</td>
<td>0.66</td>
<td>1.00</td>
<td>0.64</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.74</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Faculty</td>
<td>0.74</td>
<td>0.80</td>
<td>0.71</td>
<td>0.73</td>
<td>0.80</td>
<td>0.76</td>
<td>1.00</td>
<td>0.76</td>
<td>0.76</td>
<td>0.81</td>
<td>1.00</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>0.75</td>
<td>0.86</td>
<td>0.76</td>
<td>0.81</td>
<td>1.00</td>
<td>0.73</td>
<td>0.76</td>
<td>1.00</td>
<td>0.81</td>
<td>0.75</td>
<td>1.00</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Office</td>
<td>0.86</td>
<td>0.84</td>
<td>1.00</td>
<td>0.72</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
<td>0.71</td>
<td>1.00</td>
<td>0.71</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Botany</td>
<td>1.00</td>
<td>0.84</td>
<td>0.72</td>
<td>0.82</td>
<td>0.70</td>
<td>0.91</td>
<td>0.76</td>
<td>0.75</td>
<td>0.76</td>
<td>0.72</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td>0.76</td>
<td>0.82</td>
<td>1.00</td>
<td>0.75</td>
<td>0.80</td>
<td>0.70</td>
<td>0.75</td>
<td>0.70</td>
<td>0.66</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biology</td>
<td>0.67</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.67</td>
<td>0.65</td>
<td>0.67</td>
<td>0.64</td>
<td>0.98</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philosophy</td>
<td>0.77</td>
<td>0.93</td>
<td>0.90</td>
<td>0.68</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td>0.79</td>
<td>1.00</td>
<td>0.68</td>
<td>0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ComputerScience</td>
<td>0.55</td>
<td>0.80</td>
<td>0.76</td>
<td>0.65</td>
<td>1.00</td>
<td>0.68</td>
<td>1.00</td>
<td>0.68</td>
<td>0.97</td>
<td>0.98</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.69</td>
<td>0.76</td>
<td>0.65</td>
<td>0.76</td>
<td>1.00</td>
<td>0.80</td>
<td>0.75</td>
<td>0.77</td>
<td>0.80</td>
<td>0.74</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.90</td>
<td>0.79</td>
<td>0.68</td>
<td>0.81</td>
<td>0.80</td>
<td>0.79</td>
<td>0.85</td>
<td>0.74</td>
<td>0.83</td>
<td>0.79</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>0.73</td>
<td>0.73</td>
<td>0.66</td>
<td>0.40</td>
<td>0.69</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. For computing linguistic measures, we have used WordNet 3.0, JWN 1.4, Lucene 3.0.2 through the OntoSim library and the Alignment API 4.1:

- cosynonymy distance:

<table>
<thead>
<tr>
<th></th>
<th>Assistant</th>
<th>Reference</th>
<th>AssistantProfessor</th>
<th>VisitingProfessor</th>
<th>ResearchAssistant</th>
<th>TeachingAssistant</th>
<th>Professor</th>
<th>Room</th>
<th>PhDStudent</th>
<th>Lecture</th>
<th>FullProfessor</th>
<th>AssociateProfessor</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Faculty</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Office</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Botany</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>People</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Biology</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Philosophy</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>ComputerScience</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>FullProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

This distance is not very useful because few synonyms are in the classes.
Solution D.6 (Extensional distance computation). Consider the two ontologies $o$ and $o'$ of Problem C.1 as illustrated in Fig. C.1; assume that the two following tables of Problem C.6 specify data instances for $o$ and $o'$, respectively.

1. In order to use the extensions of classes for matching $o$ and $o'$, it is first necessary to produce record linkage because there are no “common extension” or easy way to identify objects.
2. It is clear that firstname and lastname in the first table have to be compared with name in the second one. An easy approximation would be to check if the lastname is a suffix for name, this joined with the test that firstname is also a prefix (of length 1) for name should provide a good guess. Unfortunately, it does not help to distinguish between Paola and Pierluigi Pomodoro. It is also not enough for identifying Federico di Guava since the name has not been written in the same way. So, further normalisation of the names should certainly be performed (identifying first and last names in the second table and comparing them to the first one). So, one good record linkage technique for these tables would be:

a) name2 is name with the first term (alternatively all terms but the last) acronymed, i.e., replaced by their first letter followed with a dot ".";
b) name3 is name2 with all capital letter but the first in each names preceded by a space " ";
c) name4 is the concatenation of firstname and lastname;
d) compute edit distance between name3 and name4.

3. This distance will provide the following results:

<table>
<thead>
<tr>
<th></th>
<th>128</th>
<th>1998</th>
<th>1678</th>
<th>1178</th>
<th>88</th>
<th>445</th>
<th>1664</th>
<th>unknown</th>
<th>178</th>
</tr>
</thead>
<tbody>
<tr>
<td>fdg</td>
<td>0.90</td>
<td>0.70</td>
<td>0.90</td>
<td>0.70</td>
<td>0.90</td>
<td>0.80</td>
<td>0.00</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>gc</td>
<td>0.50</td>
<td>0.70</td>
<td>0.90</td>
<td>0.70</td>
<td>0.00</td>
<td>0.70</td>
<td>0.90</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>sz</td>
<td>0.80</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.80</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>plp</td>
<td>0.60</td>
<td>0.80</td>
<td>0.00</td>
<td>0.90</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>cc</td>
<td>0.80</td>
<td>0.00</td>
<td>0.80</td>
<td>0.90</td>
<td>0.78</td>
<td>0.70</td>
<td>0.56</td>
<td>0.70</td>
<td>0.73</td>
</tr>
<tr>
<td>pc</td>
<td>0.00</td>
<td>0.80</td>
<td>0.60</td>
<td>0.80</td>
<td>0.80</td>
<td>0.50</td>
<td>0.70</td>
<td>0.90</td>
<td>0.73</td>
</tr>
<tr>
<td>ms</td>
<td>0.64</td>
<td>0.91</td>
<td>0.82</td>
<td>0.64</td>
<td>0.91</td>
<td>0.73</td>
<td>0.64</td>
<td>0.91</td>
<td>0.82</td>
</tr>
<tr>
<td>pp</td>
<td>0.60</td>
<td>0.80</td>
<td>0.00</td>
<td>0.90</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>dm</td>
<td>0.91</td>
<td>0.82</td>
<td>0.91</td>
<td>0.82</td>
<td>0.73</td>
<td>0.82</td>
<td>0.73</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>

4. The substring similarity yields:

<table>
<thead>
<tr>
<th></th>
<th>128</th>
<th>1998</th>
<th>1678</th>
<th>1178</th>
<th>88</th>
<th>445</th>
<th>1664</th>
<th>unknown</th>
<th>178</th>
</tr>
</thead>
<tbody>
<tr>
<td>fdg</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
<td>0.83</td>
<td>0.83</td>
<td>0.91</td>
<td>0.58</td>
<td>0.84</td>
</tr>
<tr>
<td>gc</td>
<td>0.85</td>
<td>0.84</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.38</td>
<td>0.84</td>
<td>0.92</td>
<td>0.85</td>
</tr>
<tr>
<td>sz</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.33</td>
<td>0.91</td>
<td>0.92</td>
<td>0.91</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>plp</td>
<td>0.92</td>
<td>0.92</td>
<td>0.38</td>
<td>0.92</td>
<td>0.84</td>
<td>0.92</td>
<td>0.92</td>
<td>0.85</td>
<td>0.93</td>
</tr>
<tr>
<td>cc</td>
<td>0.71</td>
<td>0.30</td>
<td>0.90</td>
<td>0.90</td>
<td>0.81</td>
<td>0.80</td>
<td>0.90</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>pc</td>
<td>0.27</td>
<td>0.81</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.73</td>
<td>0.81</td>
<td>0.91</td>
<td>0.83</td>
</tr>
<tr>
<td>ms</td>
<td>0.83</td>
<td>0.91</td>
<td>0.91</td>
<td>0.83</td>
<td>0.91</td>
<td>0.74</td>
<td>0.82</td>
<td>0.91</td>
<td>0.83</td>
</tr>
<tr>
<td>pp</td>
<td>0.91</td>
<td>0.81</td>
<td>0.27</td>
<td>1.00</td>
<td>0.90</td>
<td>0.82</td>
<td>0.71</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>dm</td>
<td>0.92</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.92</td>
<td>0.92</td>
<td>0.84</td>
<td>0.92</td>
<td>0.78</td>
</tr>
</tbody>
</table>
5. Single linkage applied to the results of the first method comparing individuals (Answer 3) provides:

| Assistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| Reference | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| AssistantProfessor | 1.00 0.00 0.00 0.70 0.64 1.00 1.00 0.64 0.60 0.00 0.00 |
| VisitingProfessor | 1.00 0.60 0.56 0.64 1.00 1.00 0.56 0.00 0.60 0.00 0.00 |
| ResearchAssistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| TeachingAssistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| Professor | 1.00 0.00 0.00 0.56 0.64 1.00 1.00 0.56 0.00 0.00 0.00 |
| Room | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| PhDStudent | 1.00 0.73 0.64 0.00 0.64 1.00 1.00 0.00 0.73 0.64 0.73 |
| Lecture | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| FullProfessor | 1.00 0.00 0.90 0.80 0.91 1.00 1.00 0.80 0.70 0.90 0.00 |
| AssociateProfessor | 1.00 0.00 0.60 0.80 0.64 1.00 1.00 0.64 0.00 0.60 0.70 |
| Staff | 1.00 0.73 0.64 0.00 0.64 1.00 1.00 0.00 0.73 0.64 0.73 |

and the substring comparison yields:

| Assistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| Reference | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| AssistantProfessor | 1.00 0.38 0.27 0.81 0.83 1.00 1.00 0.81 0.90 0.27 0.38 |
| VisitingProfessor | 1.00 0.30 0.71 0.80 0.91 1.00 1.00 0.80 0.38 0.71 0.30 |
| ResearchAssistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| TeachingAssistant | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| Professor | 1.00 0.27 0.27 0.71 0.83 1.00 1.00 0.71 0.27 0.27 0.30 |
| Room | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| PhDStudent | 1.00 0.74 0.83 0.33 0.78 1.00 1.00 0.33 0.85 0.83 0.74 |
| Lecture | 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 |
| FullProfessor | 1.00 0.58 0.92 0.91 0.84 1.00 1.00 0.84 0.83 0.92 0.58 |
| AssociateProfessor | 1.00 0.27 0.91 0.71 0.91 1.00 1.00 0.71 0.27 0.91 0.81 |
| Staff | 1.00 0.74 0.83 0.33 0.78 1.00 1.00 0.33 0.85 0.83 0.74 |

D.5 Strategies

**Solution D.7 (Measure aggregation).** Consider the distances between ontology entities in \( o \) and \( o' \) given by \( (i) \) edit distance computed on their names (see Problem D.5,
3rd item), and (ii) the distance computed with the single linkage measure applied to the substring distance (see Problem D.6, 3).

1. The aggregation with the \( \max(x + y - 1, 0) \) triangular norm is:

<table>
<thead>
<tr>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Reference</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>1.00</td>
<td>0.78</td>
<td>0.94</td>
<td>0.83</td>
<td>0.83</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>1.00</td>
<td>0.82</td>
<td>1.00</td>
<td>0.82</td>
<td>0.91</td>
<td>1.00</td>
<td>0.88</td>
<td>0.88</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Professor</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>0.89</td>
<td>0.89</td>
<td>0.78</td>
<td>0.89</td>
<td>0.80</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>Room</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>PhDStudent</td>
<td>1.00</td>
<td>0.80</td>
<td>0.90</td>
<td>0.33</td>
<td>0.80</td>
<td>1.00</td>
<td>0.80</td>
<td>0.83</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>Lecture</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>1.00</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>1.00</td>
<td>0.85</td>
<td>0.85</td>
<td>0.92</td>
<td>0.93</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>1.00</td>
<td>0.72</td>
<td>0.91</td>
<td>0.83</td>
<td>0.91</td>
<td>1.00</td>
<td>0.83</td>
<td>0.89</td>
<td>0.91</td>
<td>0.81</td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
<td>0.71</td>
<td>0.78</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.90</td>
<td>0.93</td>
</tr>
</tbody>
</table>

2. The aggregation with the weighted product, such that \( 2/3 \) is the weight for the former distance and \( 1/3 \) is the weight for the latter one, is:

<table>
<thead>
<tr>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>0.20</td>
<td>0.17</td>
<td>0.20</td>
<td>0.15</td>
<td>0.15</td>
<td>0.20</td>
<td>0.22</td>
<td>0.22</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Reference</td>
<td>0.17</td>
<td>0.12</td>
<td>0.20</td>
<td>0.17</td>
<td>0.22</td>
<td>0.15</td>
<td>0.22</td>
<td>0.17</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>0.20</td>
<td>0.07</td>
<td>0.06</td>
<td>0.15</td>
<td>0.14</td>
<td>0.20</td>
<td>0.21</td>
<td>0.16</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>0.20</td>
<td>0.05</td>
<td>0.16</td>
<td>0.15</td>
<td>0.17</td>
<td>0.20</td>
<td>0.21</td>
<td>0.16</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>0.20</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.21</td>
<td>0.22</td>
<td>0.21</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>0.20</td>
<td>0.20</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.18</td>
<td>0.20</td>
</tr>
<tr>
<td>Professor</td>
<td>0.17</td>
<td>0.05</td>
<td>0.06</td>
<td>0.14</td>
<td>0.16</td>
<td>0.17</td>
<td>0.20</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Room</td>
<td>0.19</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.17</td>
<td>0.19</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td>PhDStudent</td>
<td>0.20</td>
<td>0.13</td>
<td>0.17</td>
<td>0.02</td>
<td>0.14</td>
<td>0.20</td>
<td>0.22</td>
<td>0.06</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>Lecture</td>
<td>0.13</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.22</td>
<td>0.16</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.19</td>
<td>0.12</td>
<td>0.19</td>
<td>0.17</td>
<td>0.19</td>
<td>0.21</td>
<td>0.16</td>
<td>0.16</td>
<td>0.17</td>
<td>0.12</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.19</td>
<td>0.04</td>
<td>0.18</td>
<td>0.13</td>
<td>0.17</td>
<td>0.20</td>
<td>0.20</td>
<td>0.13</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>Staff</td>
<td>0.22</td>
<td>0.14</td>
<td>0.18</td>
<td>0.05</td>
<td>0.00</td>
<td>0.22</td>
<td>0.22</td>
<td>0.07</td>
<td>0.19</td>
<td>0.17</td>
</tr>
</tbody>
</table>
3. The aggregation with the weighted sum, such that $2/3$ is the weight for the former distance and $1/3$ is the weight for the latter one, is:

<table>
<thead>
<tr>
<th>Assistant</th>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.78</td>
<td>0.78</td>
<td>0.93</td>
<td>1.00</td>
<td>1.00</td>
<td>0.93</td>
<td>0.93</td>
<td>0.93</td>
<td>0.91</td>
</tr>
<tr>
<td>Reference</td>
<td>0.85</td>
<td>0.70</td>
<td>0.93</td>
<td>0.85</td>
<td>1.00</td>
<td>0.78</td>
<td>1.00</td>
<td>0.85</td>
<td>1.00</td>
<td>0.93</td>
<td>0.73</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>0.93</td>
<td>0.65</td>
<td>0.72</td>
<td>0.83</td>
<td>0.79</td>
<td>0.93</td>
<td>0.96</td>
<td>0.86</td>
<td>0.89</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>0.92</td>
<td>0.65</td>
<td>0.90</td>
<td>0.82</td>
<td>0.85</td>
<td>0.92</td>
<td>0.96</td>
<td>0.85</td>
<td>0.72</td>
<td>0.83</td>
<td>0.73</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>0.92</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.96</td>
<td>1.00</td>
<td>0.96</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>0.92</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.88</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Professor</td>
<td>0.85</td>
<td>0.68</td>
<td>0.76</td>
<td>0.83</td>
<td>0.87</td>
<td>0.85</td>
<td>0.93</td>
<td>0.76</td>
<td>0.68</td>
<td>0.62</td>
<td>0.68</td>
</tr>
<tr>
<td>Room</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.83</td>
<td>0.89</td>
<td>0.81</td>
<td>0.87</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>PhDStudent</td>
<td>0.93</td>
<td>0.78</td>
<td>0.88</td>
<td>0.31</td>
<td>0.79</td>
<td>0.93</td>
<td>1.00</td>
<td>0.64</td>
<td>0.95</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>Lecture</td>
<td>0.71</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>1.00</td>
<td>0.81</td>
<td>1.00</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.90</td>
<td>0.81</td>
<td>0.92</td>
<td>0.92</td>
<td>0.90</td>
<td>0.90</td>
<td>0.95</td>
<td>0.84</td>
<td>0.84</td>
<td>0.87</td>
<td>0.82</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.89</td>
<td>0.57</td>
<td>0.90</td>
<td>0.79</td>
<td>0.86</td>
<td>0.93</td>
<td>0.93</td>
<td>0.79</td>
<td>0.68</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>0.82</td>
<td>0.94</td>
<td>0.59</td>
<td>0.26</td>
<td>1.00</td>
<td>1.00</td>
<td>0.78</td>
<td>0.95</td>
<td>0.88</td>
<td>0.87</td>
</tr>
</tbody>
</table>

4. The aggregation with the ordered weighted average, such that $2/3$ is the weight for the former distance and $1/3$ is the weight for the latter one, is:

<table>
<thead>
<tr>
<th>Assistant</th>
<th>Course</th>
<th>Science</th>
<th>Faculty</th>
<th>Student</th>
<th>Staff</th>
<th>Office</th>
<th>Zoology</th>
<th>People</th>
<th>Biology</th>
<th>Philosophy</th>
<th>ComputerScience</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.96</td>
<td>0.93</td>
<td>0.96</td>
<td>0.89</td>
<td>0.89</td>
<td>0.96</td>
<td>1.00</td>
<td>1.00</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>Reference</td>
<td>0.93</td>
<td>0.85</td>
<td>0.96</td>
<td>0.93</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
<td>0.93</td>
<td>1.00</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td>AssistantProfessor</td>
<td>0.96</td>
<td>0.65</td>
<td>0.72</td>
<td>0.83</td>
<td>0.81</td>
<td>0.96</td>
<td>0.98</td>
<td>0.86</td>
<td>0.90</td>
<td>0.68</td>
<td>0.72</td>
</tr>
<tr>
<td>VisitingProfessor</td>
<td>0.96</td>
<td>0.65</td>
<td>0.90</td>
<td>0.82</td>
<td>0.88</td>
<td>0.96</td>
<td>0.98</td>
<td>0.85</td>
<td>0.72</td>
<td>0.83</td>
<td>0.73</td>
</tr>
<tr>
<td>ResearchAssistant</td>
<td>0.96</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.98</td>
<td>1.00</td>
<td>0.98</td>
<td>0.98</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>TeachingAssistant</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Professor</td>
<td>0.93</td>
<td>0.68</td>
<td>0.76</td>
<td>0.83</td>
<td>0.87</td>
<td>0.93</td>
<td>0.96</td>
<td>0.76</td>
<td>0.68</td>
<td>0.62</td>
<td>0.68</td>
</tr>
<tr>
<td>Room</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.92</td>
<td>0.94</td>
<td>0.90</td>
<td>0.93</td>
<td>0.98</td>
</tr>
<tr>
<td>PhDStudent</td>
<td>0.97</td>
<td>0.78</td>
<td>0.88</td>
<td>0.32</td>
<td>0.79</td>
<td>0.97</td>
<td>1.00</td>
<td>0.64</td>
<td>0.95</td>
<td>0.82</td>
<td>0.78</td>
</tr>
<tr>
<td>Lecture</td>
<td>0.86</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>0.97</td>
<td>0.93</td>
</tr>
<tr>
<td>FullProfessor</td>
<td>0.95</td>
<td>0.81</td>
<td>0.92</td>
<td>0.92</td>
<td>0.90</td>
<td>0.95</td>
<td>0.97</td>
<td>0.84</td>
<td>0.84</td>
<td>0.89</td>
<td>0.82</td>
</tr>
<tr>
<td>AssociateProfessor</td>
<td>0.94</td>
<td>0.57</td>
<td>0.90</td>
<td>0.79</td>
<td>0.89</td>
<td>0.96</td>
<td>0.96</td>
<td>0.79</td>
<td>0.68</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>Staff</td>
<td>1.00</td>
<td>0.82</td>
<td>0.94</td>
<td>0.59</td>
<td>0.52</td>
<td>1.00</td>
<td>1.00</td>
<td>0.78</td>
<td>0.95</td>
<td>0.88</td>
<td>0.87</td>
</tr>
</tbody>
</table>

In summary, it seems that the two measures either complement each other or agree. The two former aggregations heavily depend on the higher or the lower aggregated value, while the two latter aggregation are rather an average between these values. In spite of the different techniques used, they provide similar results.
**Solution D.8 (Thresholds).** The correspondences resulting from the application of a .6 threshold to the similarity of Problem C.8 are:

1. for a hard threshold of .6 are (11):

   \[
   \begin{align*}
   \text{PhDStudent} & = .92 \text{ Student} & \text{Staff} & = .65 \text{ Student} \\
   \text{PhDStudent} & = .65 \text{ Philosophy} & \text{Staff} & = .64 \text{ Science} \\
   \text{PhDStudent} & = .64 \text{ People} & \text{Staff} & = .63 \text{ Staff} \\
   \text{Professor} & = .02 \text{ People} & \text{Assistant} & = .62 \text{ Science} \\
   \text{Room} & = .02 \text{ Office} & \text{Professor} & = .60 \text{ Faculty} \\
   \text{PhDStudent} & = .60 \text{ Science}
   \end{align*}
   \]

3. for a proportional threshold of .6 are those over the best value (.92) multiplied by .6, i.e., those over .552, that is the previous ones plus 2 correspondences (+2=13):

   \[
   \begin{align*}
   \text{Assistant} & = .58 \text{ Faculty} & \text{Staff} & = .56 \text{ People}
   \end{align*}
   \]

2. for a delta threshold of .6 are those over the best value (.92) minus .6, i.e., those over .32, that is the previous ones plus 17 correspondences (+17=30):

   \[
   \begin{align*}
   \text{Professor} & = .55 \text{ Staff} & \text{PhDStudent} & = .55 \text{ Staff} \\
   \text{PhDStudent} & = .52 \text{ Boxology} & \text{Assistant} & = .46 \text{ Philosophy} \\
   \text{PhDStudent} & = .45 \text{ Faculty} & \text{Professor} & = .44 \text{ Philosophy} \\
   \text{Assistant} & = .44 \text{ Student} & \text{Assistant} & = .43 \text{ Staff} \\
   \text{Professor} & = .40 \text{ Science} & \text{Assistant} & = .40 \text{ People} \\
   \text{Professor} & = .36 \text{ Student} & \text{Professor} & = .36 \text{ Office} \\
   \text{Lecture} & = .34 \text{ Philosophy} & \text{PhDStudent} & = .34 \text{ Office} \\
   \text{Assistant} & = .33 \text{ Boxology} & \text{Staff} & = .33 \text{ Faculty} \\
   \text{PhDStudent} & = .33 \text{ Course}
   \end{align*}
   \]

4. for a percentage threshold of .6 are the 60% best correspondences, i.e., the 37 best correspondences, that is the previous ones plus 7 correspondences (+7=37 because there are three correspondence at .23):

   \[
   \begin{align*}
   \text{Professor} & = .32 \text{ Boxology} & \text{Assistant} & = .32 \text{ Course} \\
   \text{Reference} & = .28 \text{ Boxology} & \text{Reference} & = .26 \text{ Philosophy} \\
   \text{Lecture} & = .26 \text{ Faculty} & \text{Reference} & = .25 \text{ Science} \\
   \text{Reference} & = .23 \text{ People} & \ldots & \text{or any of the other at .23}
   \end{align*}
   \]

**Solution D.9 (Alignment extraction).** Consider the two ontologies \( o \) and \( o' \) of Problem C.1 as illustrated in Fig. C.1. Assume that the similarity between their entities is expressed by the similarity table of Problem C.8.
1. The alignment extracted with the greedy algorithm, is:

\[
\begin{align*}
\text{PhDStudent} &= .92 \text{ Student} \\
\text{Professor} &= .62 \text{ People} \\
\text{Assistant} &= .58 \text{ Faculty} \\
\text{Reference} &= .28 \text{ Boxology}
\end{align*}
\]

\[
\begin{align*}
\text{Staff} &= .64 \text{ Science} \\
\text{Room} &= .62 \text{ Office} \\
\text{Lecture} &= .34 \text{ Philosophy}
\end{align*}
\]

2. The alignment provided by the greedy algorithm is stable;
3. The weights of the alignment extracted by the greedy algorithms sum to 4.0. The maximum weight alignment, given below, has weights summing to 4.03:

\[
\begin{align*}
\text{PhDStudent} &= .92 \text{ Student} \\
\text{Professor} &= .62 \text{ People} \\
\text{Assistant} &= .62 \text{ Science} \\
\text{Reference} &= .28 \text{ Boxology}
\end{align*}
\]

\[
\begin{align*}
\text{Staff} &= .63 \text{ Staff} \\
\text{Room} &= .62 \text{ Office} \\
\text{Lecture} &= .34 \text{ Philosophy}
\end{align*}
\]

Solution D.10 (Composing matchers). Consider the application described in Problem D.1.

1. The matching architecture developed so far would be:

![Diagram of matching architecture]

The advantage of this architecture is that it allows to test many approaches for terminological matching (D.5), instance-based matching (D.6), aggregation (D.7), applying threshold (D.8) or extracting (D.9) the alignment.

2. The figure above provides the name of the classes, available on the book web site, which implements the surrounding dotted composition. The implementation in a new Alignment class corresponds to composing the AggregateAlignment and ExtractAlignment processes. As implemented in the ComposedAlignment class, this can be achieved through:

```java
public void align(Alignment alignment, Properties params) throws AlignmentException {
    // Call AggregateAlignment
    matcher.align( alignment, params );

    // Call ExtractAlignment... that only do extraction from matcher
    extractor.align( matcher, params );

    // Extract alignment
    for ( Cell c : extractor ) { addCell( c ); }
}
```
D.6 Evaluation of matching systems

Solution D.11 (Precision and recall computation). The precision, recall, F-measure, overall and Hamming distance alignments $A_1$, $A_2$ and $A_3$ and $R$ is given in the following table:

<table>
<thead>
<tr>
<th>Alignment</th>
<th>precision</th>
<th>recall</th>
<th>F-measure</th>
<th>overall</th>
<th>Hamming distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>-.33</td>
<td>.2</td>
</tr>
<tr>
<td>$A_1$</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.33</td>
<td>.5</td>
</tr>
<tr>
<td>$A_2$</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
<td>-.33</td>
<td>.2</td>
</tr>
</tbody>
</table>

Hence $A_2$ seems to be the best alignment. The results for $A_3$ are different from those given by the Alignment API because, we have considered here that the correspondence $\text{Student} = \text{PhDStudent}$ is not correct because of the relation which should have been $\ge$, though the Alignment API only considers the entities and thus consider the correspondence correct (all values are .5).

Solution D.12 (Application specific evaluation). We have three systems and only one application. This application, given our previous answers speed (runtime=No) and automation (automatic=NO) are not important, precision is very important (correct=YES), recall moderately important (complete=Yes). Hence, we would assign weights 1 to speed and automation, 5 to precision, 4 to recall, which, once normalised yield .1, .5 and .4. Since we have no information about speed, it is not taken into account and all systems are equally automated.

<table>
<thead>
<tr>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$A_3$</th>
<th>Automatic</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>.33</td>
<td>.66</td>
<td>.33</td>
<td>1.</td>
<td>.33</td>
<td>.33</td>
</tr>
</tbody>
</table>

The results are very similar to those provided by F-measure.

D.7 Representing alignments

Solution D.13 (Representation generation). Consider $R$ as described in Fig. C.2.

1. $R$ in OWL as given by the Alignment API (see Solution D.16, Question 1, for the actual API output):

```xml
<?xml version='1.0' encoding='utf-8' standalone='no'?>
<!DOCTYPE rdf:RDF [ 
<!ENTITY admin "http://book.ontologymatching.org/exercise/admin.owl#" >
<!ENTITY lab "http://book.ontologymatching.org/exercise/lab.owl#" >
]>
<rdf:RDF
xmlns:owl="http://www.w3.org/2002/07/owl#" 
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
```
Appendix D: Solutions

```xml
<owl:Ontology rdf:about=""/>
<rdfs:comment>OWL Alignment rendering</rdfs:comment>
<rdfs:comment>(altered from OWLAxiomsRendererVisitor output)</rdfs:comment>
</owl:Ontology>

<owl:Class rdf:about="&admin;Office">
  <owl:equivalentClass rdf:resource="&lab;Room"/>
</owl:Class>

<owl:Class rdf:about="&lab;PhDStudent">
  <rdfs:subClassOf rdf:resource="&admin;Student"/>
</owl:Class>

<owl:ObjectProperty rdf:about="&admin;room">
  <owl:equivalentProperty rdf:resource="&lab;office"/>
</owl:ObjectProperty>

<owl:Class rdf:about="&admin;Faculty">
  <owl:equivalentClass rdf:resource="&lab;FullProfessor"/>
</owl:Class>

<owl:ObjectProperty rdf:about="&admin;teaches">
  <owl:equivalentProperty rdf:resource="&lab;teaches"/>
</owl:ObjectProperty>

<owl:Class rdf:about="&admin;Course">
  <owl:equivalentClass rdf:resource="&lab;Lecture"/>
</owl:Class>

2. R in C-OWL as given by the Alignment API:

```xml
<?xml version='1.0' encoding='utf-8' standalone='no'?>
<!DOCTYPE rdf:RDF [>
<!ENTITY admin "http://book.ontologymatching.org/exercise/admin.owl#" >
<!ENTITY lab "http://book.ontologymatching.org/exercise/lab.owl#" >
<rdf:RDF
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xml:base="http://www.itc.it/cowl#">
  <owl:Ontology
</owl:sourceOntology>
  <owl:targetOntology
    rdf:about="http://book.ontologymatching.org/exercise/lab.owl"/>
</owl:bridgeRule>
  <owl:Equivalent
    <owl:source
      <owl:Class rdf:about="&admin;Office"/>
</owl:source>
  <owl:target
      <owl:Class rdf:about="&lab;Room"/>
</owl:target>
  </owl:Equivalent>
</owl:bridgeRule>
<owl:bridgeRule>
  <owl:source
      <owl:Class rdf:about="&admin;Student"/>
</owl:source>
</owl:bridgeRule>
</owl:RDF>
```
3. \( R \) in SWRL as given by the Alignment API:

```xml
<?xml version='1.0' encoding='utf-8' standalone='no'?>
<!DOCTYPE swrlx:Ontology [ ]>
<!ENTITY admin "http://book.ontologymatching.org/exercise/admin.owl#" >
<!ENTITY lab "http://book.ontologymatching.org/exercise/lab.owl#" >

<swrlx:Ontology swrlx:name="generatedAl"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:swrlx="http://www.w3.org/2003/11/swrlx#"
xmlns:owlx="http://www.w3.org/2003/05/owl-xml"
xmlns:ruleml="http://www.w3.org/2003/11/ruleml#">
<owlx:Imports
<ruleml:imp>
<ruleml:_body>
<swrlx:ClassAtom>
<owlx:Class owlx:name="&admin;Office"/>
```
Appendix D: Solutions

<ruleml:var>x</ruleml:var>
</swrl:individualPropertyAtom>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;room"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;office"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;teaches"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;teaches"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;teaches"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;teaches"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;teaches"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>

<ruleml:imp>
<ruleml:_body>
<swrl:individualPropertyAtom swrl:name="&admin;lecture"/>
<ruleml:var>y</ruleml:var>
</ruleml:_body>
</ruleml:_head>
</ruleml:imp>
4. $R$ in the SEKT mapping language as given by the Alignment API:

MappingDocument(<""
  source("http://book.ontologymatching.org/exercise/admin.owl")
  target("http://book.ontologymatching.org/exercise/lab.owl")
  classMapping( <"#s075689">
    bidirectional
    <"http://book.ontologymatching.org/exercise/lab.owl#Room">
      <"http://book.ontologymatching.org/exercise/admin.owl#Office">
  }
  classMapping( <"#s035024">
    unidirectional
    <"http://book.ontologymatching.org/exercise/admin.owl#Student">
      <"http://book.ontologymatching.org/exercise/lab.owl#PhDStudent">
  }
  attributeMapping( <"#s074544">
    bidirectional
    <"http://book.ontologymatching.org/exercise/admin.owl#room">
      <"http://book.ontologymatching.org/exercise/lab.owl#office">
  }
  classMapping( <"#s026252">
    bidirectional
    <"http://book.ontologymatching.org/exercise/admin.owl#Faculty">
      <"http://book.ontologymatching.org/exercise/lab.owl#FullProfessor">
  }
  attributeMapping( <"#s006240">
    bidirectional
    <"http://book.ontologymatching.org/exercise/admin.owl#teaches">
      <"http://book.ontologymatching.org/exercise/lab.owl#teaches">
  }
  classMapping( <"#s016467">
    bidirectional
    <"http://book.ontologymatching.org/exercise/admin.owl#Course">
      <"http://book.ontologymatching.org/exercise/lab.owl#Lecture">
  )

5. $R$ in SKOS as given by the Alignment API:

<?xml version='1.0' encoding='utf-8' standalone='no'?>
<!DOCTYPE rdf:RDF >
<!ENTITY admin "http://book.ontologymatching.org/exercise/admin.owl#" >
<!ENTITY lab "http://book.ontologymatching.org/exercise/lab.owl#" >
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:skos="http://www.w3.org/2008/01/skos#">
<skos:Concept rdf:about="#admin;Office">
  <skos:exactMatch rdf:resource="#lab;Room"/>
</skos:Concept>
<skos:Concept rdf:about="#admin;room">
  <skos:exactMatch rdf:resource="#lab;office"/>
</skos:Concept>
The discussion on the advantages of alignment formats is rather generic. The OWL representation provides a merge between the two ontologies. Once loaded, this OWL does not distinguish an alignment and two ontologies, but only one ontology. The other representations maintain the two ontologies and represent the alignments as mappings (C-OWL, SEKT or SKOS) or rules (SWRL). SWRL and C-OWL use oriented mappings (from source to target) while the others are not oriented. OWL, C-OWL and SWRL are immediately operational for reasoning and expressed in a formally defined language, while SEKT and SKOS require a processor. With respect to the RDF format, all these representations lose information such as the metadata on the alignments or the confidence measures.

D.8 Explaining alignments

Solution D.14 (Alignment explanation). Consider the ontologies of Fig. C.2 and the architecture and a solution devised for Problem D.10.

1. A decision flow for the correspondence \( \text{Student} = \text{PhDStudent} \) is provided below. Following the discussion of Solution D.11, the relation of this correspondence is not specific enough, and hence, it is an incorrect one. The above presented decision flow would be that of a matching system able to compute only equivalence correspondences.

2. However, a system able to compute subsumption relations, e.g., through a common suffix rule, hence being able to generate the \( \text{Student} \geq \text{PhDStudent} \) as an initial candidate, would manage to deliver it as the final and correct correspondence with the same explanation flow.

3. Both the terminological substring matcher and the instance-based single linkage matcher agree that there is a relation between entities Student and PhDStudent by assigning them a high similarity, which was further reinforced through an aggregation via weighted product. This correspondence, with the similarity of .98, passed the hard threshold of .6 and was also retained by the greedy alignment extraction algorithm. Hence, it was delivered as the final one. The (in)correctness depends of the ability of a matching system to compute a specialisation relation.
D.9 Processing alignments

Solution D.15 (Merging ontologies).

1. The merge between $o$ and $o'$ according to $R$ can be expressed with OWL import by:

```xml
<owl:Ontology rdf:about="">
  <rdfs:comment>Matched ontologies</rdfs:comment>
</owl:Ontology>

<owl:Class rdf:about="http://book.ontologymatching.org/exercise/admin.owl#Office">
  <owl:equivalentClass rdf:resource="http://book.ontologymatching.org/exercise/lab.owl#Room"/>
</owl:Class>

<owl:Class rdf:about="http://book.ontologymatching.org/exercise/lab.owl#PhDStudent">
  <rdfs:subClassOf rdf:resource="http://book.ontologymatching.org/exercise/admin.owl#Student"/>
</owl:Class>

<owl:ObjectProperty rdf:about="http://book.ontologymatching.org/exercise/admin.owl#room">
</owl:ObjectProperty>

<owl:Class rdf:about="http://book.ontologymatching.org/exercise/admin.owl#Faculty">
  <owl:equivalentClass rdf:resource="http://book.ontologymatching.org/exercise/lab.owl#FullProfessor"/>
</owl:Class>
```
This is what is indeed rendered by the OWLAxiomsRendererVisitor delivered with the Alignment API.

2. It can also be expressed as an integrated OWL ontology like the one of Figure D.3, in which Faculty replaces FullProfessor, Staff in $o$ has been renamed Staff#2, Room replaces Office, Lecture replaces Course, office replaces room and both teaches have been unified.

![Diagram](image)

**Fig. D.3.** Merge of the ontology $o$ and $o'$ with regard to the reference alignment $R$.

**Solution D.16 (Data translation).** Assume that we want to transform the data instances from ontology $o$ into instances of $o'$. Consider the data instances in the first table of Problem C.6.
1. The transformation of data instances in the first table of Problem C.6 with regard to the reference alignment $R$ would lead to the creation of 6 Full professors with no name, no hiring date and no office.

2. There may be two ways to perform the translation using the functions of the Alignment API. If the data is or can be expressed in XML, then the Alignment API is able to generate an XSLT transformation thanks to its XSLTRendererVisitor serialiser. This transformation may then be applied to the XML data. Otherwise, it is necessary to use the alignment from Java for performing the adequate actions. Here is the core of a program that does this with the Alignment API and Jena:

```java
public void transform(OntModel onto1, OntModel onto2, Alignment al) {
    Hashtable<String,Individual> ht = new Hashtable<String,Individual>();
    URI base = al.getOntology2URI();
    // For each correspondence involving a class, transform individuals
    for (Cell c : al) {
        OntClass cl = onto1.getOntClass(c.getObject1AsURI(al).toString());
        if (cl != null && (c.getRelation().getRelation().equals(“=”)
            || c.getRelation().getRelation().equals(“<”))) {
            OntClass cl2 = onto2.getOntClass(c.getObject2AsURI(al).toString());
            for (Individual i : listInstances(cl)) {
                String id = base.getScheme() + base.getSchemeSpecificPart()
                    + URLEncoder.encode(i.getLabel(), "UTF-8");
                Individual o = onto2.createIndividual(id, cl2);
                ht.put(i.getURI(), o);
                System.out.println(o.getURI() + " rdf:type " + o.getOntClass() + ".");
            }
        }
        // For each correspondence involving a property, transform statements
        for (Cell c : al) {
            OntProperty pr = onto1.getOntProperty(c.getObject1AsURI(al).toString());
            if (pr != null) {
                OntProperty pr2 = onto2.getOntProperty(c.getObject2AsURI(al).toString());
                StmtIterator stmtit = onto1.listStatements(null, pr, null);
                for (; stmtit.hasNext();)
                    System.out.println(stmtit.nextStatement().toString());
            }
        }
    }
}
```

A working example is given in the exercise web site.

Solution D.17 (Mediation).

---

$^3$ It would have been better to ask for transferring from $o'$ to $o$, to have the rooms instead of the ids in $o'$'s table. Even this would only allow to transfer PhD and Full professors.
1. The translation of the query ‘SELECT ?x.room WHERE Faculty(?x)’ expressed with respect to ontology o involves finding the correspondences for attribute room and class Faculty. Hence the translated query is ‘SELECT ?x.office WHERE FullProfessor(?x)’. Rewritten in terms of SPARQL, the query is:

   SELECT ?y
   WHERE { ?x rdf:type Faculty AND. ?x office ?y }

and its translation is:

   SELECT ?y
   WHERE { ?x rdf:type FullProfessor . ?x room ?y }

2. When evaluated against the second table, the answer will be \{B12\} because the only instance of the class FullProfessor in the second ontology is Federico di Guava whose office is B12 (or rather the room identified by B12).

3. A mediator able to perform the transformations of any query q based on the reference alignment R will have to transform the query with regard to R \(q' = \text{TransformQuery}(q, R)\), it will then have to evaluate the result \(r' = \text{Eval}(q')\) and to return the results by translating them with regard to the converse of R \(r = \text{Translate}(r', \text{Invert}(R))\). Hence, the full mediator would be implemented by:

   \[\text{Translate}(\text{Eval}(\text{TransformQuery}(q, R)), \text{Invert}(R))\]