Evaluating the creation and understanding of uplift mappings

1. Motivation

The majority of data in the Web still resides in other formats than the Resource Description Framework[1] (RDF). RDF is a W3C recommendation for representing information in the Web, facilitating data exchange, data integration and others. One of the main tasks when upgrading legacy systems to the Semantic Web is the conversion of data. The process of converting data in any format into RDF is called *uplift*. The key stakeholders in this process are web developers, software programmers specialized in the development of systems for the web, and ontology engineers, experts in semantic web technologies such as ontologies, RDF and so on.

Several solutions have been proposed, however, these still focus on Semantic Web experts. To facilitate the uplift process and to make the technology available to a wider set of stakeholders, I have developed a method to represent uplift mappings visually. The method draws inspiration from visual programming languages such as Google's Blockly. Blockly has been used in many projects, such as code.org's introduction courses. In the visual representation, blocks represent a mapping that automatically generates an uplift mapping. In this experiment, I aim to investigate if such a visual representation: (i) facilitates the creation of accurate uplift mappings; (ii) eases the understandability of uplift mappings; and (iii) imposes an optimal mental workload on users.

2. Process

Half of the participants will use a text editor and the other half will use the aforementioned developed visual representation. The rationale for this being the comparison of the mappings created.

The study will consist of the following steps:

- a presentation about relevant technologies to the experiment will be presented;
- participants will be asked to fill a pre-questionnaire;
- participants will be asked to work on one task using either Juma Uplift or R2RML directly. This experiment design ensured that participants would be expose to only one representation;
- participants will be asked to fill out post task questionnaires.

3. Briefing arrangements

At any point throughout or after this study, all participant questions about the rationale for the study, methods of analysis etc. will be answered by the lead researcher.

4. Method Analysis

The following analysis will be done:

- for the mapping task (Section 7) - accuracy of the R2RML mapping created (either from a text editor or the developed visual representation) will be evaluated by analyzing the output of the mapping created by the participant and an expected output. For every time participants run the mapping, this will be stored as to evaluate the progress of the participant;
- for the understandability task (Section 8) – analysis of the answers in terms of precision and recall;
- usability questionnaire's responses (see Section 11.1). This questionnaire will only be used for participants executing the mapping task (Section 7);
- time needed for each task and overall for all participants;

[1] http://www.w3.org/TR/rdf11-concepts/

- mental workload questionnaire's responses (see Section 10.2 and 11.2) for all participants.

5. Ethical considerations

All data will be collected with participant consent. If, at any point during this study, the participant wishes to end their participation, their data will be immediately deleted.

All participant responses will be anonymised.

6. Relevant legislation

The Data Protection Act (1988 and 2003) applies to this research, specifically the section titled 'Protection of Privacy of Individuals with regard to Personal Data'. This section governs: (i) The conditions under which data may be obtained and analysed (ii) The obligations of the individual storing and analysing the data.

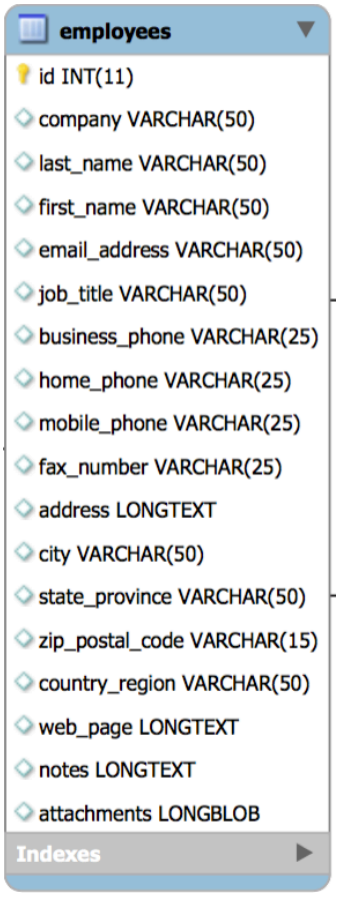
This study will only data collected with with participant consent. Participants will be provided with an information sheet beforehand explaining the purpose of the study and the information that will be collected. Any questions they have at any point during the study will be answered fully by the lead researcher. Their data will be stored on the lead researcher's password-protected computer. This computer will not be taken off the campus of Trinity College Dublin. No participant data will be transmitted in any way. The only exception to this rule would be if a participant requested a copy of their own data. If, at any time during the study, the participant wishes to end their participation, their data will be deleted immediately. Once the processing of an individual participant's data is complete, that participant's data will be deleted immediately. As such, this study is in accordance with the requirements of the Data Protection Act.

7. Mapping Task

Instructions:

- There is some material related to this experiment available;
- You will create 1 mapping;
- The task is divided in 3 parts;
- It is possible to run the mapping and check the output presented in each part;
- If you have any doubts you can ask the researcher.

The table to be mapped is shown below:



Part 1:

- Map the table EMPLOYEES shown above;
- The mapping should create subjects with the URI http://data.example.org/employee/{id}, where *id* refers to the *id* column;
- The subject should be of URI type class foaf:Person;
- The mapping should generate the predicate foaf:givenName for every value of the column first_name;
- The mapping should generate the predicate foaf:familyName for every value of the column last_name;
- The mapping should generate the predicate foaf:name concatenating the values of the columns last_name and first_name separated by a comma;

Example output for employee 1

<http://data.example.org/employee/1> a <http://xmlns.com/foaf/0.1/Person> ;
<http://xmlns.com/foaf/0.1/familyName> "Freehafer" ;
<http://xmlns.com/foaf/0.1/givenName> "Nancy" ;
<http://xmlns.com/foaf/0.1/name> "Freehafer, Nancy" .

Part 2

- Map the table EMPLOYEES again;
- The mapping should create subjects with the URI http://data.example.org/city/{city}, where *city* refers to the *city* column;
- The subject should be of URI type class foaf:Spatial_Thing;
- The mapping should generate the predicate rdfs:label for every value of the column city;

Example output for employee 1

<http://data.example.org/city/Seattle>
a <http://xmlns.com/foaf/0.1/Spatial_Thing> ;
<http://www.w3.org/2000/01/rdf-schema#label> "Seattle" .

Part 3:

- Relate the subject created in Part 1 with the subject created in Part 2 using the predicate foaf:based_near ;

Example output for employee 1

<http://data.example.org/employee/1>
<http://xmlns.com/foaf/0.1/based_near>
<http://data.example.org/city/Seattle> .

8. Understandability Task

The input database and mappings are shown below:

Person

ID	NAME	AGE	CITY_FK
1	Ana	29	100

City

CITY_ID	NAME
100	Dublin

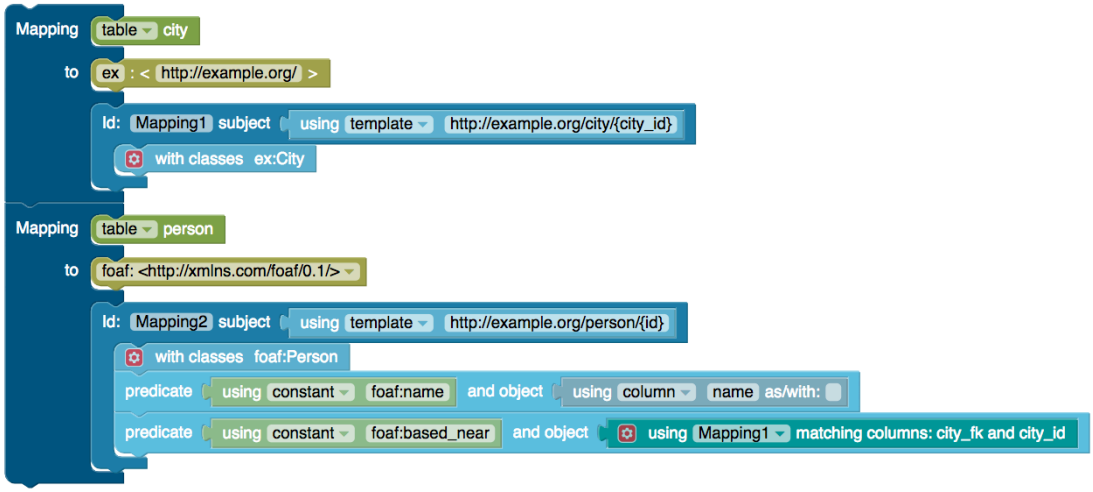
R2RML:

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMapMapping1>
rr:logicalTable [  rr:tableName "city"; ];
rr:subjectMap [
   rr:template "http://example.org/city/{city_id}";
   rr:class ex:City;
 ];.

<#TriplesMapMapping2>
rr:logicalTable [ rr:tableName "person";];
rr:subjectMap [
   rr:template "http://example.org/person/{id}";
   rr:class foaf:Person;
];
rr:predicateObjectMap [
  rr:predicateMap [     rr:constant foaf:name;    ];
  rr:objectMap [     rr:column "name";     ];
];
rr:predicateObjectMap [
   rr:predicateMap [ rr:constant foaf:based_near; ];
   rr:objectMap [
     rr:parentTriplesMap <#TriplesMapMapping2>;
     rr:joinCondition [
      rr:child "city_fk";
      rr:parent "city_id";
    ]; ]; ];.
```

Juma Uplift representation:



Select all correct triples that will be generated from the mapping presented above

1. `<http://example.org/person/1> a foaf:Person .`
2. `<http://example.org/person/1> foaf:name "Ana" .`
3. `<http://example.org/city/100> a ex:City .`
4. `<http://example.org/person/1> foaf:based_near <http://example.org/city/100> .`
5. `<http://example.org/person/id> class foaf:Person .`
6. `<http://example.org/person/id> foaf:name "name" .`
7. `<http://example.org/city/city_id> class ex:City .`
8. `<http://example.org/person/id> foaf:based_near <http://example.org/city/city_id>.`
9. `foaf:Person a "1" .`
10. `foaf:Person based_near <#TriplesMapMapping1> .`
11. `ex:City a "100" .`
12. `foaf:name a "name" .`
13. I cannot make sense of this mapping.

9. Questionnaires

The following header is on the top of every page: Each question is optional. Feel free to omit a response to any question; however, the researcher would be grateful if all questions are responded to.

10. Pre-task questionnaire

All participants of the experiment will answer this questionnaire.

10.1. Familiarity with semantic web technologies

- I am familiar with Semantic Web technologies (OWL, RDF, amongst others).

STRONGLY AGREE						STRONGLY DISAGREE
1	2	3	4	5	6	7

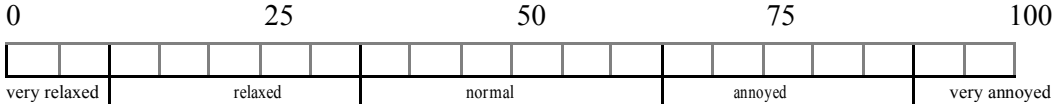
- I am familiar with R2RML.

STRONGLY AGREE						STRONGLY DISAGREE
1	2	3	4	5	6	7

10.2. Mental workload [3]

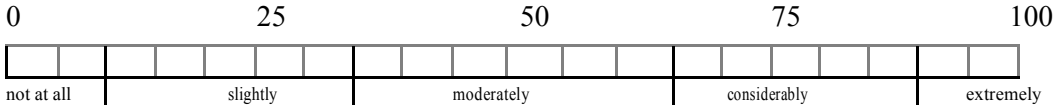
Frustration

Are you frustrated? In other words, are you relax or annoyed?



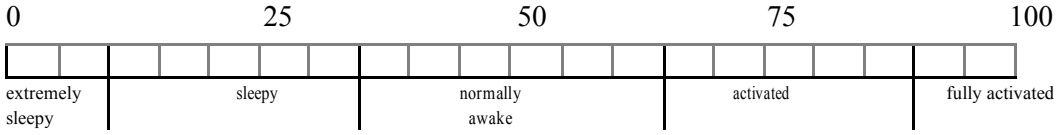
Motivation

How motivated are you to be doing this experiment?



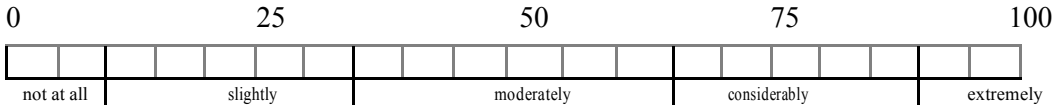
Alertness

Are you alerted? In other words, are you sleepy/tired or fully activated/awake?



Emotional State

Is some thought or something that occurred recently in your life mentally bothering you right now?



11. Post-questionnaire

11.1. PSSUQ - Post-Study System Usability Questionnaire

This questionnaire uses the following 7 likert scale. Every question has a text field for comments (not shown) [2]. Only participants executing the task presented in Section 7 will answer this questionnaire.

This message is shown before the comment text field: Please do not name any third parties in any open text field of the questionnaire. Any such reply will be anonymized.

STRONGLY AGREE						STRONGLY DISAGREE	
1	2	3	4	5	6	7	N/A

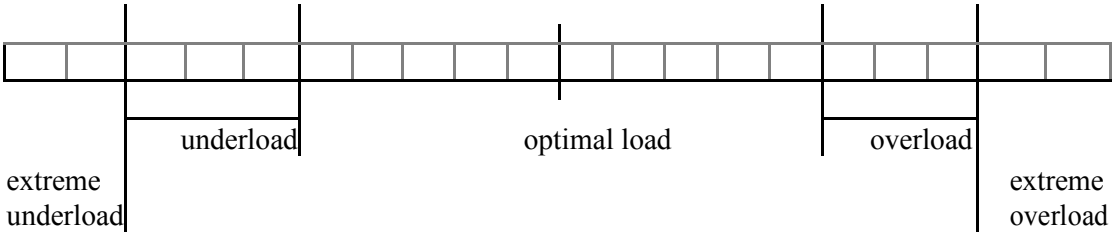
1. Overall, I am satisfied with how easy it is to use this system
2. It was simple to use this system
3. I could effectively complete the tasks and scenarios using this system
4. I was able to complete the tasks and scenarios quickly using this system
5. I was able to efficiently complete the tasks and scenarios using this system
6. I felt comfortable using this system
7. It was easy to learn to use this system
8. I believe I could become productive quickly using this system
9. The system gave error messages that clearly told me how to fix problems

10. Whenever I made a mistake using the system, I could recover easily and quickly
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear
12. It was easy to find the information I needed
13. The information provided for the system was easy to understand
14. The information was effective in helping me complete the tasks and scenarios
15. The organization of information on the system screens was clear
16. The interface of this system was pleasant
17. I liked using the interface of this system
18. This system has all the functions and capabilities I expect it to have
19. Overall, I am satisfied with this system

11.2. Mental workload [3]

All participants of the experiment will answer this questionnaire.

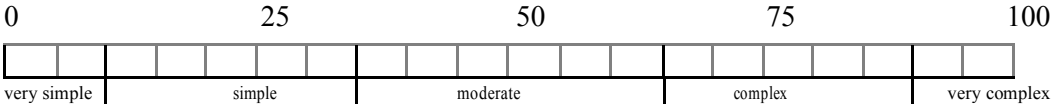
How much mental workload the task imposed on you?



In order to answer the following questions, please place an 'x' in one of the 20 boxes.

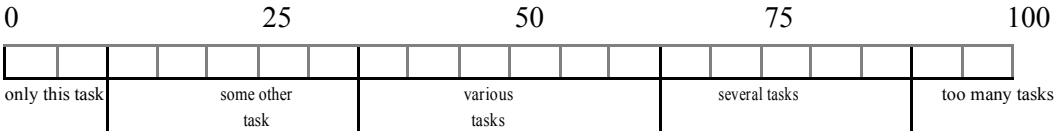
Mental demand

How much mental and perceptual activity was required by the task? In other words, was the task easy or demanding, simple or complex?



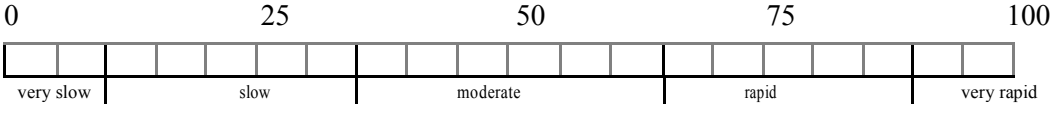
Parallelism

Did you perform just this task or were you engaged in other parallel tasks (mobile browsing/social networks, chatting, reading, conversations etc.)?



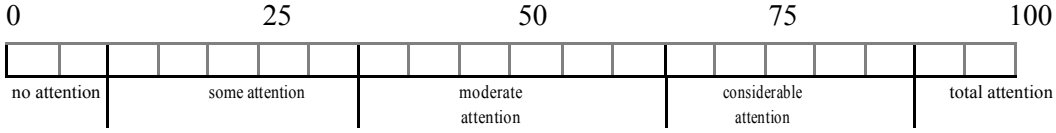
Temporal demand

How much time pressure did you feel due to the pace at which the task occurred? In other words, was the pace slow or rapid?



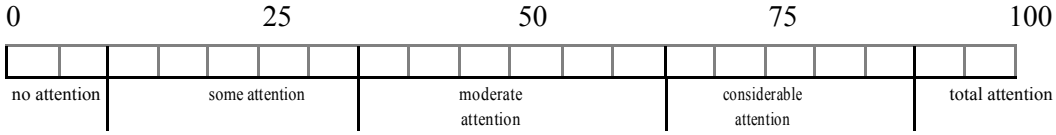
Manual activity

How much attention was required for manually responding to the elements/activities of the task (writing, drawing, etc)?



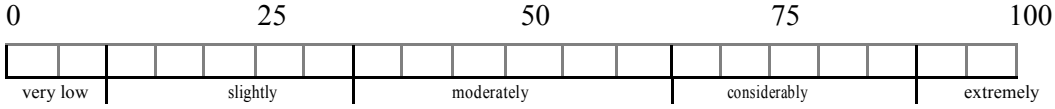
Visual attention

How much attention was required for the task based on the information visually received?



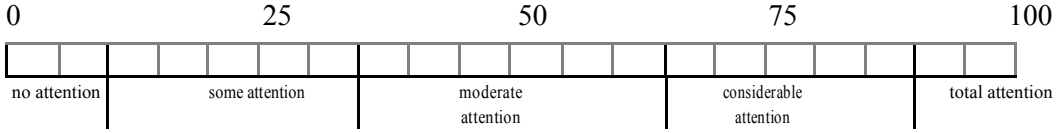
Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance during the task?



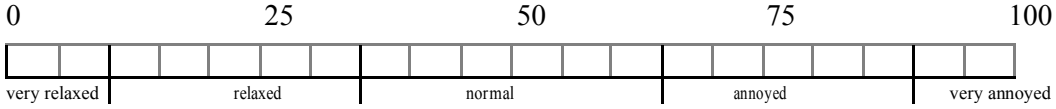
Solving & deciding

How much attention was required by the task for activities like remembering, problem solving, decision-making and perceiving (e.g. detecting, recognizing and identifying objects)?



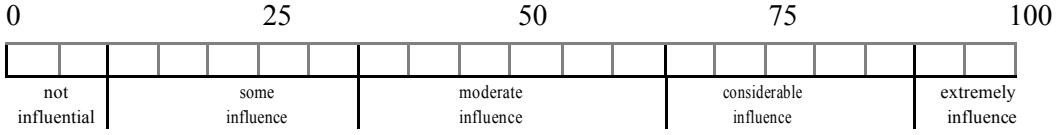
Frustration

Were you frustrated during the task? In other words, were you feeling relaxed or annoyed?



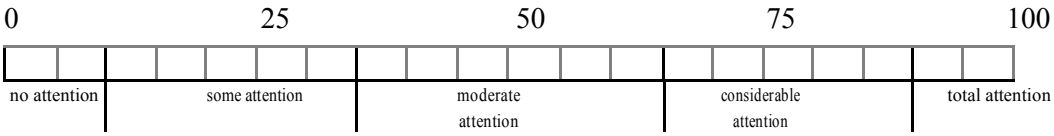
Context bias

How often interruptions during the task occurred? In other words, were distractions (mobile, interruptions, noise, questions, other participants, etc.) not influential or did they negatively influence your attention towards the task?



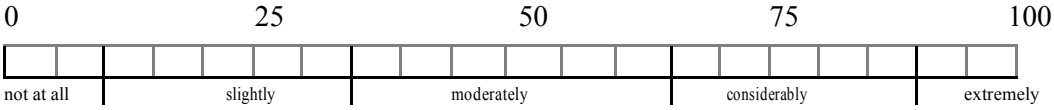
Task and space

How much attention was required by the task for spatial processing (spatially pay attention around you)?



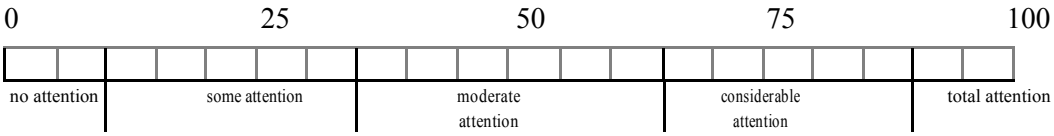
Motivation

Were you motivated by the task you have just finished?



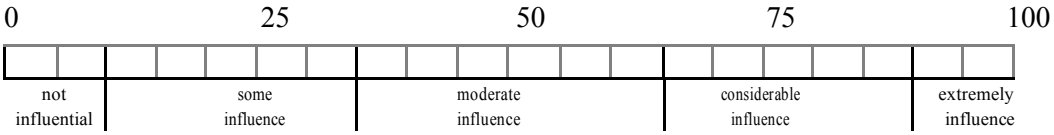
Verbal material

How much attention was required by the task for verbal material (e.g. reading or processing linguistic material or listening to verbal conversations)?



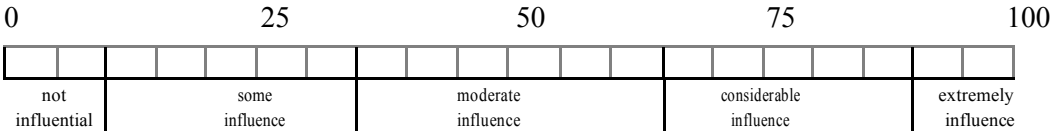
Skill

Did your skills have no influence or did they help you execute the task?



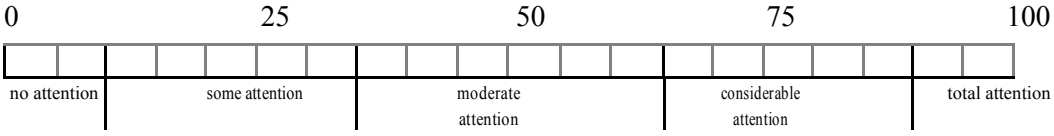
Auditory attention

How much attention was required for the task and executing its activities based on the information auditorily received?



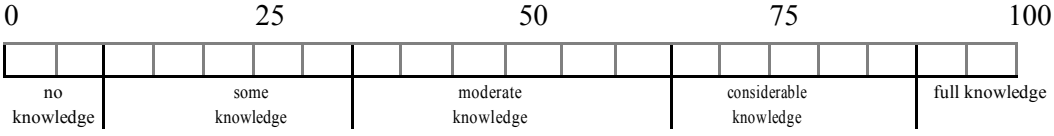
Speech response

How much attention was required by the task for producing speech responses (e.g. engaging in conversation or talk or answering questions)?



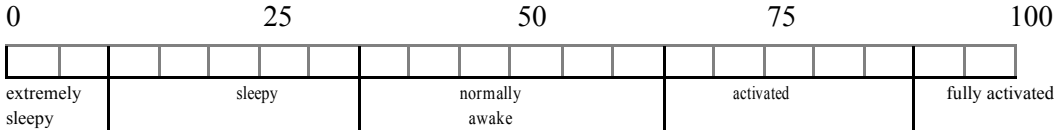
Past knowledge/expertise

How much experience/knowledge did you already had about the task?



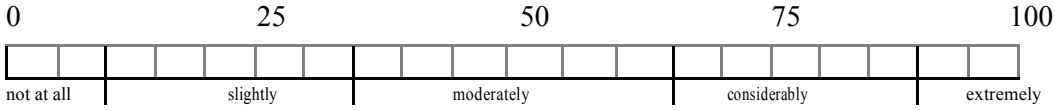
Alertness

Were you alerted during the task? In other words, were you sleepy/tired or fully activated/awake?



Performance

How successful were you in the task? In other words, how satisfied were you with your level of performance?



References

1. Hitzler, P., Krotzsch, M., Rudolph, S.: Foundations of semantic web technologies. CRC Press, (2009).
2. Lewis, J.R. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. International Journal of Human-Computer Interaction, 7(1), 57-78 (1995).
3. Longo, L. A defeasible reasoning framework for human mental workload representation and assessment. Behaviour & Information Technology, 34(8), 758-786, (2015).

TRINITY COLLEGE DUBLIN
INFORMED CONSENT FORM

LEAD RESEARCHERS: Ademar Crotti Junior

BACKGROUND OF RESEARCH:
The majority of data in the Web still resides in other formats than the Resource Description Framework (RDF). RDF is a W3C recommendation for representing information in the Web, facilitating data exchange, data integration and others. One of the main tasks when upgrading legacy systems to the Semantic Web is the conversion of data. The process of converting data in any format into RDF is called *uplift* . The key stakeholders in this process are web developers, software programmers specialized in the development of systems for the web, and ontology engineers, experts in semantic web technologies such as ontologies, RDF and so on.
Several solutions have been proposed, however, these still focus on Semantic Web experts. To facilitate the uplift process and to make the technology available to a wider set of stakeholders, I have developed a method to represent uplift mappings visually. The method draws inspiration from visual programming languages such as Google's Blockly. Blockly has been used in many projects, such as code.org's introduction courses. In the visual representation, blocks represent a mapping that automatically generates an uplift mapping. In this experiment, I aim to investigate if such a visual representation: (i) facilitates the creation of accurate uplift mappings; (ii) eases the readability and understandability of uplift mappings and (iii) imposes an optimal mental workload on users.

PROCEDURES OF THIS STUDY: You have been chosen for this study because: (i) you responded to the lead researcher's email requesting your participation. This email was sent to you because you are a part of the same research group as the lead researcher in this study (note that this may create a conflict of interest). (ii) You have knowledge on Semantic Web technologies or more specifically on uplifting data into RDF. (iii) You are a student of technical disciplines in Computer Science.

The experiment should take about 30 minutes where:

- The researcher will present basic information about relevant technologies related to this study.
- You will be asked to fill a pre-questionnaire.
- You will be asked to work on 1 task.
- Finally, you will be asked to fill questionnaires about the use of the tool.

Participation in this study is entirely voluntary. If, for any reason, you wish to terminate your participation, you are free to do so.

In the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities.

PUBLICATION: It is my intention to publish the results of this evaluation in conferences and/or scientific journals. It is also my intention to use these results in my PhD thesis.

Individual results may be aggregated anonymously and research reported on aggregate results.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.

- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- *<If the research involves viewing materials via a computer monitor>* I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
- I have received a copy of this agreement.

PARTICIPANT'S NAME:

PARTICIPANT'S SIGNATURE:

Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:
Email: crottija@scss.tcd.ie
Phone: 089 XXXXX02

INVESTIGATOR'S SIGNATURE:

Date: