Universidade do Minho
Escola de Engenharia
Departamento de Informática

Sérgio Maia Dias

Automatic creation of CVs from Online Social Networks

Universidade do Minho

Escola de Engenharia

Departamento de Informática

Sérgio Maia Dias

Automatic creation of CVs
from Online Social Networks

Master dissertation

Master Degree in Software Engineering

Dissertation supervised by

Pedro Rangel Henriques

Alda Gançarski

October 2015

ABSTRACT

This document consists in a thesis report for a master work in Software Engineering, under the topic *automatic creation of CVs from online social networks*.

The thesis is the main component of the second year of the masters degree in Software Engineering at Universidade do Minho.

This master work entails the creation of a platform that will retrieve information about the user from popular professional-oriented social network platforms and based on that information it will generate the user's *curriculum vitae* in several popular formats.

CONTENTS

Contents

LIST OF FIGURES

List of Figures

LIST OF TABLES

g

LIST OF LISTINGS

INTRODUCTION

In this first chapter of the report the master work is introduced. The purpose of this master work is exposed and is contextualized, and the necessary goals to complete it are also detailed.

1.1 CONTEXT

In a professional environment, individuals need a way to expose their employment history, qualifications and education, to prove their worth in a competitive marketplace.

In an academic environment, individuals also have this requirement, although it is tailored to the nature of the individuals' work, i.e., more focused in their research and teaching.

Previously, this requirement was satisfied by a *curriculum vitae* (CV), which is a document that contains an overview of the aforementioned aspects of an individual's career. With the emergence of the web, those aspects started to be exposed in personal or institutional web pages.

Both CV and personal web pages have limited visibility: a CV can only be shared when requested, either in a physical form or otherwise, and personal or institutional web pages can have limited visibility on the search engines that are used to find them.

As such, with the popularity of social media, professional-oriented social networks started to emerge, and became essential for personal marketing in the labor market. In these social networks, users share their professional experience and the projects (thematics and collaborators) they are involved with or they have been involved with in the past.

Examples of these are *LinkedIn* [11], where users enumerate the companies or institutions that they have worked in the past, they list the projects that they have been involved in and they describe their education and create an explicit network of past or present collaborators, and more academic focused social networks, like *Academia* [1] and *Research Gate* [14], where the users' publications and projects are displayed.

Nonetheless, this method of self-marketing is insufficient for some processes; to apply for a position at a company it is usually required for the candidate to present his professional experience in a succinct manner, ordinarily in a document for that effect, i.e., a CV.

1.2 MOTIVATION AND OBJECTIVES

As mentioned above, currently most individuals prefer to use professional oriented social networking platforms to display their professional information, and the creation and maintenance their CV is usually just a formality required by the employment process in most companies.

As such, it is clear that there is a need to utilize the information that is available in these new forms of professional exposure for the automatic generation of the documents that are accepted and required by the existing industry of the labor market.

This master work entails the creation of a system that will accomplish precisely that: using the information available online on specific social networking platforms, like LinkedIn [11] or Research Gate [14], these documents will be generated for the end user, which can then use as he sees fit.

1.3 DOCUMENT STRUCTURE

This document consists in several chapters.

Firstly there is an introductory chapter, where the context, the motivation and the objectives of this master work are detailed.

This is followed by the state of the art, where several *curriculum vitae* formats, online social networks and existing solutions are all analyzed and compared.

The third chapter aims to define in detail the proposed solution.

The fourth chapter elaborates on the implementation of said solution.

The fifth chapter ilustrates how the final implementation of the proposed solution functions, by displaying its main features and exploring selected case studies.

The final chapter will reflect on conclusions and future work.

STATE OF THE ART

Before proposing a solution to the issue at hand, a thorough analysis must be conducted to perceive the current state of the art of the area in which this master work is included.

With that purpose, in this chapter existing solutions are analyzed, as well the platforms that will be the source of the data and the existing formats for the target documents.

2.1 CURRICULUM VITAE STANDARD FORMATS

CVs are documents that are widely used. As such, a multitude of formats exist today, each defined and supported by different entities, institutions or countries. Unfortunately there is not a standard that is globally used and accepted; the Europass format [6], however, comes close: it is a product of an initiative by the Directorate-General for Education and Culture of the European Union to standardize this type of documents, and is widely accepted in several countries of the European Union, like Portugal, Spain and Italy.

Apart from the Europass format, there are other formats that are more focused on specific areas of expertise, such as the recommended format of The Career Center of the University of Washington [2], which contains specific information dedicated for academical oriented individuals, or others that serve as a standard in a more limited geographical area.

Since these formats are tailor made for a specific set of individuals or a specific situation, it becomes clear that they could not satisfy everyone, and thus are not universally accepted.

These formats are also not very strict, since they do not define a clear and strict set of rules that a CV must comply with, delegating that responsibility to the individual writing the document.

It can be concluded that CV formats vary according to the context of the individuals (geographical location and area of expertise) and their need to describe themselves in a simpler or more complex manner.

The complexity of a CV varies from format to format, but a core set of information is common among most formats.

To define the basic structure of a CV and to identify the common information associated with it, the analysis of different CV formats presented above should start with a format that is popular and

sufficiently generic to be used by individuals in a wide range of areas of expertise. The chosen format for this purpose is the Europass format.

2.1.1 *Europass*

The Europass format is simple but versatile.

It typically starts with the individual's basic information, such as name, gender, date of birth, home address and contact information.

This is followed by the individual's work experience, in which the previous positions are enumerated chronologically, describing the company or institution, its location, and the responsibilities the individual had in that position.

The individual's education and training is presented next, similarly describing the institution where it occurred and its location, the subject matters that were studied and the resulting evaluation. If the education followed the european standards, it is often identified as such.

The individual's skills are then revealed, starting with his language capabilities, where each language that the individual can use is described, and his proficiency in understanding, speaking and writing in those languages is rated according to a standard metric.

This is followed by the rest of the individual's skill set, separated by categories such as "Communication skills", "Organizational/managerial skills", "Computer Skills", among others. If the individual possesses a driver's license, it is also mentioned.

The document ends with a complementary section where additional information is displayed: the individual's affiliations, the seminars and conferences he has attended, the publications that he has contributed to, and references that potential employers can contact to inquire about the candidate.

Apart from the common information identified above, each specific format has a set of information dedicated to the specific business area that it caters to.

2.1.2 *The Career Center of the University of Washington's Academic Format*

The Career Center of the University of Washington's [2] academic CV format is intended to display all of the relevant information about an academic individual, and as such contains extensive information for that purpose.

Several extra categories are listed below.

The honors, grants, awards and fellowships should be enumerated, describing the name of the award, the institution granting it and the date awarded. The entries should be described to indicate their significance, and should be sorted by latest to oldest.

The work experience has a different focus; as an academic, both the teaching experience and the research experience of the individual should be detailed.

The teaching experience should list the positions the individual was responsible for, indicating the job title, the course title, the name of the university or institution, the time in which the position was held.

The research experience should present several aspects, including the publications the individual participated in, articles and other works that are not yet completed, the research interests, grants received in the context of research work, the research positions that were held, and presentations in their various forms and types of events.

Other type of work experience may also be presented.

If the individual participated in any activities that contributed to the department, university or the surrounding community, this should also be indicated. This includes participating in committees, boards, or in events like scientific contests, for example.

2.1.3 *College Art Association's Visual Artist Format*

The College Art Association's visual artist CV format [18] is similar to the previous format in that it also caters to a specific type of individual, and as such it also contains specific information appropriate for their line of work. This specific format is dedicated to artists with academic careers.

As with the previous format, the work experience is more detailed, and is divided between teaching experience and artistic endeavors.

In comparison with the previous format, the exhibition record takes the place of the publication record; it lists the exhibitions where the individual participated, the location of the exhibition, its date and other relevant details. The individual's prominence in the exhibition (if it was a solo exhibition, a two-person exhibition, or an exhibition of several artists) should be indicated, perhaps by separating the entries in different subcategories. It is one of the most important sections, and should be one of the first in the document for that reason.

Other type of projects that the individual was involved in should also be displayed.

The awards, grants and fellowships must also be detailed.

Any noteworthy commissions should be indicated, with moderate detail about the commissioner and the work itself.

A bibliography of any media records, produced by other people, that concerns the individual and his work should be kept; this includes interviews, articles and reviews.

Similarly to the previously format, the individual's own publications can also be listed, as do the lectures, presentations and other similar events.

Lastly if the individual's works are part of one or more collections, these should be listed and described.

2.1.4 *Comparison*

The information that is supported by the Europass format, the Career Center of the University of Washington's [2] academic CV format and the north-american College Art Association's visual artist CV format [18] is displayed in table 1.

	Europass	UW's Academic CV	CAA's Artist CV
Name	Yes	Yes	Yes
Gender	Yes	No	No
Date of Birth	Yes	No	Yes
Addresses	Yes	Yes	Yes
Contacts	Yes	Yes	Yes
Education	Yes	Yes	Yes
Work Experience	Yes	Yes	Yes
Skill Set	Yes	Yes	Yes
Affiliations	Yes	Yes	Yes
Honors and Awards	Yes	Yes	Yes
Publications	Yes	Yes	Yes
Projects	Yes	No	No
Citations	Yes	Yes	Yes
Memberships	Yes	Yes	No
Conferences	Yes	Yes	Yes
Seminars	Yes	Yes	Yes
Workshops	Yes	Yes	Yes
References	Yes	Yes	Yes
Courses	Yes	Yes	No
Certifications	Yes	Yes	No
Exhibitions	Yes[1]	No	Yes
Commissions	Yes[1]	No	Yes

Table 1.: Common information for existing formats

These last two formats are simply two examples of the many formats that are available. The purpose of this master work is not to study a large number of these formats, but to understand the need to generate these types of documents, and the level of detail that they should contain.

They will serve as case studies for the purpose of designing the proposed system.

This is the information that is relevant for the system to be developed, which needs to be collected from the available online platforms.

---

1 Not by default, but it can be supported.

2.1.5   *The Europass XML Format*

Europass is an initiative that developed several document formats that can be used to describe an individual's skills and qualifications, and present that information in a clear way to be easily understood and widely accepted through out Europe.

Since this CV format is one of the most widely used, it is generic enough to be able to support a wide range of areas of expertise and is in fact an open stardard with extensive support, it will be the base format used in the proposed solution.

These document formats created by Europass are open standards, meaning that their specification is open and can be explored by any party that wishes to do so; in fact, the "European Centre for the Development of Vocational Training" (CEDEFOP) [10] even has a portal available to describe this standard in technical terms and also to present the web services that they have available.

Europass defines a specific data model and vocabulary to hold its information. This is done via a data schema defined in two languages: XML and JSON. Of the two, the XML version of this schema is more widely used by the tools that support the Europass standard (like the "Europass Online Editor" [12], for example). It also benefits from additional resources, such as XSL Transformations, that update Europass CV documents defined in an older version. For these reasons, the XML format will be the one used in the proposed system and detailed in this section.

Please note that this document will not explore the complete schema, as it would be unnecessary and redundant, since that information is readily available as mentioned above. It will however give an overview of the schema, and present the details of the sections that are pertinent to this master work. The version of the schema detailed here is the latest (v3.3.0).

2.1.5.1   *Overview*

The Europass XML schema was designed to be generic; this allows several types of documents to be defined using the same schema. Due to this, it includes sections to hold the document's meta-data, such as the type of the document that it is, and the way the data should be presented.

The "SkillsPassport" element is the root element. It contains in turn the following elements:

- **DocumentInfo** Contains meta-data that describes the document, such as the date and time when the document was created and last updated, the version of the schema that the document uses, among other fields. Most importantly, it contains the "DocumentType" element, which determines the type of document that it is, among the different document types supported by Europass (for example, "Curriculum Vitae" or "European Skills Passport").

- **LearnerInfo** Contains the actual data of the document.

- **AttachmentList (Optional)** Contains the attachments that are referenced somewhere else in the document.

- **PrintingPreferences** Contains the meta-data that describes how the information available in the document should be displayed when printed to a readable format, such as PDF. Each field can be individually defined to be displayed or not, in a per-document basis. The order and the formatting of the elements can also be specified.

- **CoverLetter (Optional)** Contains information used to create a cover letter.

2.1.5.2 *User Data*

The Europass XML schema supports a wealth of data pertaining an individual's skills and qualifications. In this section, most of the fields supported by the Europass CV XML format will be presented and exemplified, with reference to where they are located in this format via their paths.

The most relevant element to a CV's information in the Europass XML format is the "LearnerInfo" element, since it contains most of that information. It contains several elements, which are presented, together with an example, in what follows.

- **Identification**

  Contains basic information that identifies an individual.

```
1 <Identification>
2     <PersonName> (...) </PersonName>
3     <ContactInfo> (...) </ContactInfo>
4     <Demographics> (...) </Demographics>
5 </Identification>
```

Listing 1: Europass "Identification" element example

- Identification/**PersonName**

  The person's first and last names.

```
1 <PersonName>
2     <FirstName>Betty</FirstName>
3     <Surname>Smith</Surname>
4 </PersonName>
```

Listing 2: Europass "PersonName" element example

- Identification/**ContactInfo**

  The person's various contacts, including his addresses, emails and telephone numbers, among others.

```
1 <ContactInfo>
2     <Address>
3         <Contact>
4             <AddressLine>32 Reading rd</AddressLine>
5             <PostalCode>B26 3QJ</PostalCode>
6             <Municipality>Birmingham</Municipality>
7             <Country>
```

```
8              <Code>UK</Code>
9              <Label>United Kingdom</Label>
10          </Country>
11       </Contact>
12    </Address>
13    <Email>
14       <Contact>smith@kotmail.com</Contact>
15    </Email>
16    <TelephoneList>
17       <Telephone>
18          <Contact>+44 7123456789</Contact>
19          <Use>
20              <Code>mobile</Code>
21              <Label>Mobile</Label>
22          </Use>
23       </Telephone>
24    </TelephoneList>
25    <InstantMessagingList>
26       <InstantMessaging>
27          <Contact>betty.smith</Contact>
28          <Use>
29              <Code>aim</Code>
30              <Label>AOL Instant Messenger (AIM)</Label>
31          </Use>
32       </InstantMessaging>
33    </InstantMessagingList>
34 </ContactInfo>
```

Listing 3: Europass "ContactInfo" element example

- Identification/**Demographics**

  The person's birth date, gender and nationalities.

```
1 <Demographics>
2     <Birthdate year="1975" month="--03" day="---01"/>
3     <Gender>
4         <Code>F</Code>
5         <Label>Female</Label>
6     </Gender>
7     <NationalityList>
8         <Nationality>
9             <Code>UK</Code>
10            <Label>British</Label>
11        </Nationality>
12    </NationalityList>
13 </Demographics>
```

Listing 4: Europass "Demographics" element example

- **Headline**

  A headline for the current document, used to justify the document's purpose.

```
1 <Headline>
```

```
2     <Type>
3         <Code>job_applied_for</Code>
4         <Label>JOB APPLIED FOR</Label>
5     </Type>
6     <Description>
7         <Label>European project manager</Label>
8     </Description>
9 </Headline>
```

Listing 5: Europass "Headline" element example

- **WorkExperienceList**

  Contains a list of items which describe experiences in a work environment.

```
1 <WorkExperienceList>
2     <WorkExperience>
3         <Period> (...) </Period>
4         <Position> (...) </Position>
5         <Activities> (...) </Activities>
6         <Employer> (...) </Employer>
7     </WorkExperience>
8     <WorkExperience> (...) </WorkExperience>
9 </WorkExperienceList>
```

Listing 6: Europass "WorkExperienceList" element example

- WorkExperienceList/**WorkExperience**

  A single work experience, where several parameters can be detailed: the employer, the position that was held, a description of the activities that the position encompassed, the period of time that the position was held by the individual, and complementing documentation.

```
1 <WorkExperience>
2     <Period>
3         <From year="2002" month="--08"/>
4         <Current>true</Current>
5     </Period>
6     <Position>
7         <Label>Independent consultant</Label>
8     </Position>
9     <Activities>
10         Evaluation of European Commission youth training support measures for youth
                national agencies and young people
11     </Activities>
12     <Employer>
13         <Name>British Council</Name>
14         <ContactInfo>
15             <Address>
16                 <Contact>
17                     <AddressLine>123, Bd Ney</AddressLine>
18                     <PostalCode>75023</PostalCode>
19                     <Municipality>Paris</Municipality>
20                     <Country>
```

```
21                    <Code>FR</Code>
22                    <Label>France</Label>
23                </Country>
24            </Contact>
25        </Address>
26      </ContactInfo>
27    </Employer>
28 </WorkExperience>
```

Listing 7: Europass "WorkExperience" element example

- **EducationList**

  Contains a list of items which describe periods or achievements of formal or informal training that amount to the individual's education.

```
1 <EducationList>
2    <Education>
3        <Period> (...) </Period>
4        <Title> (...) </Title>
5        <Activities> (...) </Activities>
6        <Organisation> (...) </Organisation>
7    </Education>
8    <Education> (...) </Education>
9 </EducationList>
```

Listing 8: Europass "EducationList" element example

- EducationList/**Education**

  Describes one period or achievement of formal or informal training that amounts to the individual's education. Several parameters can be detailed: the title of the course (or similar), the organization, the field, a level (the classification of the individual's performance in the course), a description of the activities that the course encompassed, the period of time in which the course was completed, and complementing documentation.

```
1 <Education>
2    <Period>
3        <From year="1993"/>
4        <To year="1997"/>
5        <Current>false</Current>
6    </Period>
7    <Title>Bachelor of Science in Sociology and Psychology</Title>
8    <Activities>
9        - Sociology of risk
10       - Sociology of scientific knowledge / information society
11       - Anthropology
12       - E-learning and Psychology
13       - Research methods
14    </Activities>
15   <Organisation>
16       <Name>Brunel University</Name>
17       <ContactInfo>
```

```
18          <Address>
19              <Contact>
20                  <Municipality>London</Municipality>
21                  <Country>
22                      <Code>UK</Code>
23                      <Label>United Kingdom</Label>
24                  </Country>
25              </Contact>
26          </Address>
27      </ContactInfo>
28  </Organisation>
29  <Documentation>
30      <ReferenceTo idref="ATT_123456789"/>
31  </Documentation>
32 </Education>
```

Listing 9: Europass "Education" element example

- **Skills**

  Contains skills the individual affirms to have. They are sorted into several types, each having a specific level of detail and parametrization. These include: linguistic skills, communication skills, organizational skills, job-related skills, computer skills, driving skills, and other skills.

```
1 <Skills>
2      <Linguistic> (...) </Linguistic>
3      <Communication> (...) </Communication>
4      <Organisational> (...) </Organisational>
5      <Computer> (...) </Computer>
6      <Driving> (...) </Driving>
7      <Other> (...) </Other>
8      (...)
9 </Skills>
```

Listing 10: Europass "Skills" element example

- Skills/**Linguistic**

  Specifies linguistic skills, including the mother tongue as well as foreign languages; for the latter, the proficiency levels (using european norms) can be specified, as well as the time it was acquired, the certificates used to attest the claim and documentation.

```
1 <Linguistic>
2      <MotherTongueList>
3          <MotherTongue>
4              <Description>
5                  <Code>en</Code>
6                  <Label>English</Label>
7              </Description>
8          </MotherTongue>
9      </MotherTongueList>
10     <ForeignLanguageList>
11         <ForeignLanguage>
```

12

```
12          <Description>
13              <Code>fr</Code>
14              <Label>French</Label>
15          </Description>
16          <ProficiencyLevel>
17              <Listening>C1</Listening>
18              <Reading>C2</Reading>
19              <SpokenInteraction>B2</SpokenInteraction>
20              <SpokenProduction>C1</SpokenProduction>
21              <Writing>C2</Writing>
22          </ProficiencyLevel>
23        </ForeignLanguage>
24     </ForeignLanguageList>
25 </Linguistic>
```

Listing 11: Europass "Linguistic" element example

- Skills/**Other**

  Specifies skills that are not featured by default on the Europass CV format. It allows for defining a description that consists in formatted text, and referencing annexes. Most other skills elements, like "Communication" and "Job-Related" are identical in structure.

```
1 <Other>
2     <Description>Creating pieces of Art and visiting Modern Art galleries. Enjoy all
          sports particularly hockey, football and running. Love to travel and
          experience different cultures.
3     </Description>
4 </Other>
```

Listing 12: Europass "Other" element example

- **AchievementList**

  Contains a list of achievements the individual reports having.

```
1 <AchievementList>
2     <Achievement> (...) </Achievement>
3     (...)
4 </AchievementList>
```

Listing 13: Europass "AchievementList" element example

- AchievementList/**Achievement**

  An achievement that the individual reports having. Its type has to be specified either by a label or a code. If the code is set it must have a predefined value ("honors_awards", "publications", "projects" and so on) but the label can have any value. It also supports referencing annexes.

```
1 <Achievement>
2     <Title>
3         <Code>publications</Code>
4         <Label>Publications</Label>
```

13

```
5        </Title>
6        <Description>'How to do Observations: Borrowing techniques from the Social
             Sciences to help Participants do Observations in Simulation Exercises' Coyote
             EU/CoE Partnership Publication, (2002).
7        </Description>
8    </Achievement>
9
10   <Achievement>
11       <Title>
12           <Label>Scholarship</Label>
13       </Title>
14       <Description>
15           - Standard High School Scholarship
16           - Standard University Scholarship
17       </Description>
18   </Achievement>
```

Listing 14: Europass "Achievement" element example

The "AttachmentList" element, that is under the "SkillsPassport" root element, can contain the data of a set of attachments.

It can contain "Attachment" elements, in which each attachment's meta-data, like the type of attachment, its name, its description, etc, is specified, along with the attachment's data, encoded in a Base64 string of characters.

Each attachment is also identified by a unique id, by which it can be referenced throughout several sections of a CV, such as education and achievements.

```
1  <AttachmentList>
2      <Attachment id="ATT_370337481435">
3          <Name>Bachelor Certificate.pdf</Name>
4          <MimeType>application/pdf</MimeType>
5          <Data>JVBERi0xLjUKJeLjz9MKMiAwIG9iago (...)</Data>
6          <MetadataList>
7              <Metadata key="number-of-pages" value="4"/>
8          </MetadataList>
9          <Description>Bachelor Certificate.pdf</Description>
10     </Attachment>
11     (...)
12 </AttachmentList>
```

Listing 15: Europass "AttachmentList" element example

2.2 ONLINE SOCIAL NETWORKS

As mentioned previously, there are several social networks and other online repositories of information that should be taken into account when selecting the sources of data for the system to be implemented. The data sources considered are *LinkedIn*, *Academia*, *Research Gate*, *Google Scholar* and *Behance*, and the consequent analysis is presented below.

2.2.1 *LinkedIn*

LinkedIn [11] is an obvious choice, since it is one of the most widely used professional oriented social network online today.

A user's profile contains most of the information needed to generate a CV following the Europass standard, and LinkedIn provides an API to easily access it. To access the information an authenticated user is required. If the user is the owner of the profile being accessed, all of the profile's information is available; if the user is not the owner of that profile, the information that is available is determined by the profile owner's privacy settings.

2.2.2 *Academia and ResearchGate*

Academia [1] and *Research Gate* [14] are academic oriented social networks, which contain academic related information, like the publications of an individual. They are relevant in the scope of this master work since for academic oriented individuals, their professional information is contained in these platforms.

They also have a particularity that is interesting and will be useful: they often contain information about the individuals, even if they have not entered it by themselves, since often their peers will do so for them (mostly due to cooperation on publications).

These social networks are very similar; they both allow a user to enumerate their publications and other works, and even publish those works' documents online. However, they have one key difference: while Academia allows users to simply upload an existing CV as a file and provide it for download, Research Gate allows users to add the information that is typically in a CV to their profile page directly. This makes Research Gate a better choice as a data source.

2.2.3 *Google Scholar*

Google Scholar [9] is a Google product that indexes publications from a large variety of sources, from journals to publishers and even online media like Research Gate users' publications.

Besides working as a simple web search engine, it also allows users to manage their own profiles, where they can colaborate with Google to identify their own works, moderating what Google automatically indexes.

For each work it identifies, it fetches and makes available a large quantity of metadata that makes it invaluable as a source for an individual's publications.

2.2.4   *Behance*

Behance [13] is an online portfolio, more focused in the display of works of art and similar endeavors. It is therefore of limited use, since a CV usually only describes the individual and at most enumerates his works; nonetheless, for artistic oriented individuals, it is a useful source of information.

2.3 EXISTING SOLUTIONS FOR CV GENERATION FROM ONLINE SOCIAL NETWORKS

The work explored in this master thesis concerns the automatic generation of CV using users' online social networks' information. It is only natural that for a necessity as evident as this one several applications (software packages) already exist.

In this section these solutions will be analyzed, focusing on their main objectives and features.

The solutions that will be explored are the following: Yevgeniy Brikman's *Resume Builder* [19], *DoYouBuzz.com* [5], *VisualCV* [20], *Create-CV.com* [3] and EGrabber's *ResumeGrabber Suite* [17].

All of the solutions mentioned above are capable of extracting information from social networks, being *LinkedIn* the most common source of data. Some, like *Create-CV.com*, even support other social networking and messaging services, like *Facebook*, *Twitter*, *Skype*, etc, although most of them are either poor sources of professional information or require extensive mechanisms for extracting data, which makes them not feasible.

The main objective of some of the systems above is to collect the information and make it available online in structure commonly associated with CV, with the option of further editing it with the integrated tools, and possibly export it as a file ready for printing. This is the case of *Resume Builder*, *DoYouBuzz.com* and *VisualCV*. *Create-CV.com* simply allows to export the end result as a file, and does not allow displaying it online.

These solutions focus on being complete and self-sufficient, once the information as been retrieved from the data sources. In doing so they become less useful, because they cannot be reused in other systems as easily, or use the functionality of other systems.

For example, none of these tools generate files in a file format that is easily editable externally, or that follow a standard (like Europass, which uses an XML based format) and therefore can be integrated with other systems.

Table 2 depicts the features of each tool, establishing a comparison and identifying the main issues that were detected.

	Re-sume Builder	DoYouBuzz	VisualCV	Create CV
Import info from online media	Yes, from LinkedIn	Yes, from LinkedIn and Viadeo	Yes, from LinkedIn	Yes, from LinkedIn, Facebook, Twitter, Skype, etc (although it only retrieves basic information)
Import info from files	No	No	Yes, from PDF and Word	No
Export CV in read-only file format	Yes, in PDF	Yes, in PDF, although it only exports the basic information	Yes, in PDF	Yes, in PDF
Export CV in editable file format	No	Yes, in Word, although it only exports basic information	No	No
Has an integrated editor	Yes	Yes, with many features	Yes, with many features	No
Can share CV online	Yes, through a link	Yes, through a link, social networks or publicly on search engines	Yes, through a link, social networks or publicly on search engines	Yes, through a link
Includes info for specific areas of expertise	No	No	No	No

Table 2.: Comparison of features of the identified CV generation systems

As it can be easily concluded, the main objective of these systems is to collect the individual's data, allow for it to be easily edited inside the system and easily share it online, or export it as a read-only format.

The biggest flaws in these systems are:

- In general, they do not allow exporting the CV in an editable format;

- In general, they only import information from the most popular source of professional data, i.e., LinkedIn;

- The information that they include in the CV is only the core set of information identified previously; more specific information, like detailed academic information, is not considered.

These aspects will be considered in the proposed system for this master work.

There are also other systems that make use of CV, although for other purposes or with other objectives in mind.

EGrabber's *ResumeGrabber Suite* is a substantially different solution, since, unlike the other solutions that have been presented, it is not focused on the individual described in the CV as the user, but as a tool to be used in the context of an organization for integration of CV information of the organization's collaborators or job applicants.

Although it is considerably different from the purpose of the system that will be developed for this master work, it identifies a new possible end user for that system, this being not the individuals them-

selves, but a third party that wants to generate CVs for a group of individuals. This, however, is not a common situation, and such a tool would be of limited use.

<div align="right">

3

</div>

PROPOSED SOLUTION

As stated in chapter 2, existing systems that generate CVs from OSN information have some draw-backs: they do not use standard CV formats; they are not capable of export the CVs in an editable for-mat, which limits their interoperability; usually the only available source of information is LinkedIn; and they do not manage information from specific areas of expertise.

To overcome all of these issues, we propose a solution by building a system which enables users to create their CVs in an automated fashion by utilizing their existing information available in selected online social networks. We call our system CVGenie.

This chapter will elaborate the system's requirements, its planned system architecture and data model, and finally a thorough examination on the most important features.

3.1 REQUIREMENTS

The proposed system will be designed towards users who want to generate their own CVs, from the information about themselves available online. The system shall provide the following features:

- **Use a standard and versatile CV format:** the format that dictates the structure of the CV documents managed and generated by the system should be widely used, versatile and, idealy, an open standard, to enhance the system's interoperability. With this purpose, the Europass CV format was selected.

- **Import information from files:** the system should be able to import information from files in the standard Europass XML format.

- **Import information from online media:** The information required for the CV should be obtained from the selected online platforms (*LinkedIn*, *Google Scholar*, *Research Gate* and *Behance*). The latter should be used mainly for information for the academic and artistic sections of the CV.

- **Include information for specific areas of expertise:** the system should manage information for the aforementioned case studies (academic CV and academic artist CV).

- **Export CV in an editable format:** the system should be able to export CV documents in the Europass XML format, making it possible to later import these documents in Europass compatible systems.

- **Export CV in a read-only format:** the system should be able to export CV documents as PDF files or simple HTML websites.

3.2 SOLUTION ARCHITECTURE

To secure the functionality that was planned, an appropriate and detailed system architecture must be designed and implemented. Figure 1 presents the architecture of the proposed CVGenie system.

As it can be noted, the system should have the following modules:

- **XML Validator** Validates if an XML file is valid according to the Europass XML format schema;

- **File Manager** Saves and loads files from the storage system;

- **XML Converter** Converts data between an intermediate format appropriate for manipulation by the system and its XML-based representation;

- **HTML Generator** Creates an HTML web page with the information from a CV;

- **XML Manager** Manages XML files the insertion and extraction of information from a CV, by using "XML Validator", "XML Converter" and "File Manager";

- **File Importer** Obtains the CV XML File from a user and uses "XML Manager" to integrate that file into the system;

- **HTML Exporter** Retrieves information from the CV using "XML Manager", generates a HTML web page using "HTML Generator" and supplies it to the user;

- **PDF and Others Exporter** Retrieves information from the CV using "XML Manager", generates a file in the selected format and supplies it to the user;

- **Data Fetcher** Obtains the data from an online media platform;

- **Data Normalizer** Normalizes data before it is imported;

- **Online Media Importer** Retrieves information from each online media platform by using a "Data Fetcher" component, normalizes the retrieved data with a "Data Normalizer" component and inserts it into the CV via "XML Manager".

Figure 1.: UML component diagram of CVGenie's architecture

3.3 DATA MODEL

The proposed system's data model is displayed in fig. 2.
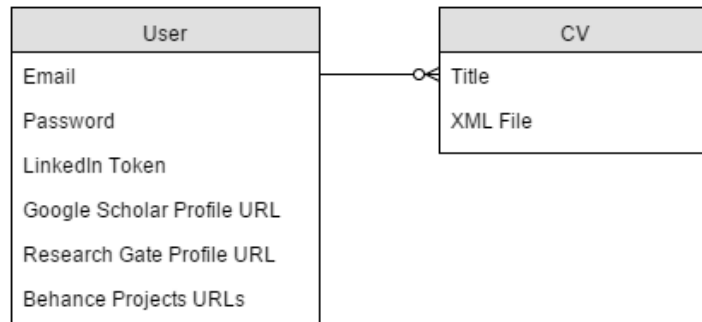
Figure 2.: CVGenie's data model

Each user has an account and can create an unlimited amount of CVs.

Some extra parameters will be saved for the user's convinience; these will allow a user to repeat the importing process without having to login or specify his profile in the online social networks that will be supported.

Each CV is an entry that contains an Europass XML file.

The data relative to the users' CV documents is stored in the system in the Europass XML format directly, since this is the main format that the system works with. This format is described in detail in section 2.1.5.2.

By choosing this format for storage, the system will not require having redundant databases or complex mechanisms to input and output the data by transforming it between the different formats. This would be the case if the data was kept, for example, in a relational database, which is commonly used. The downside is that this system is therefore very dependent on this format, and more specifically to the current version of this format. However, since the Europass team provides support to this format and has available XSL transformers to convert the documents across different versions of the Europass document standard, this problem poses only a small concern.

3.4 INCLUSION OF INFORMATION

As it was mentioned previously, the CV format (as well as its implementation on a technical standpoint) that was chosen for the storage of information in the proposed system will be the Europass CV format, using its XML implementation.

In this section, the elected sources of information are analyzed, as well as how that information can be imported into this particular format. It will stand to prove that the Europass CV format is sufficiently

versatile to manage, in the very least, the case studies that have been selected (generic, academic and artistic information about an individual).

Since inevitably the OSNs will only be able to provide with a subset of the information that is desired (not to mention required, in some cases) by the previously considered CV formats in section 2.1, first an analysis of those CV formats will be made and how the Europass CV format can support the same information that the more specific CV formats do; afterwards an analysis of the OSNs that were considered in section 2.1 will be made, and how the information contained within can also be integrated in that format.

3.4.1 *CV Formats Analysis*

In section 2.1, several formats were considered: firstly, the Europass format, to serve as a generic CV document format capable of including standard information about an individual, as well as two CV formats that are custom-tailored to two different profiles, respectively: the Career Center of the University of Washington's CV format, for academic individuals, and the College Art Association's visual artist CV format, for artistic individuals.

3.4.1.1 *Academic CVs*

The Career Center of the University of Washington's CV format distinguishes itself from other CV formats in that it considers the academic experience of an individual and displays it appropriately. As such, disregarding common information, it features the following sections or fields:

- Teaching experience;
- Research experience;
- Other professional experience;
- Professional training;
- Honors, awards, fellowships and grants;
- Publications, possible divided by type, and with status of completion on display;
- Articles, presented similarly to publications;
- Research interests;
- Participation in related events, such as conferences, seminars, and similar;
- Professional service, such as contribution to committees, boards, task forces and similar activities.

All of these sections should be integrable in the Europass CV format.

In an academic area of expertise, there are several types of experience that can be considered work experience; this CV format features teaching and research experience, as well as other indiscriminate professional experience. They are however separated by type, to better distinguish and highlight

each type. Nonetheless they can be grouped into a single work experience section, and be further differentiated by prefixing pertinent labels, for example. The result could then be integrated into the "Work Experience" section of the Europass CV format.

Professional training can be included in the "Education and Training" section.

The other sections, such as honors, awards, publications and participation and/or contribution to events, can all be sorted into the "Additional Information" section, which is technically named "AchievementList" in the Europass XML format. As previously mentioned in section 2.1.5.2, this section is the crux of the versatility of the Europass CV format; each section from the academic CV format can be included here under an individual item, with a description as elaborate as necessary, and it even can be supported by annexing documents which are referenced in that item.

3.4.1.2 *Artistic CVs*

The College Art Association's visual artist CV format is targeted to individuals that also have academic careers and additionally are artists. Apart from the usual CV sections, it also features the following sections:

- Teaching experience;
- Awards, grants, fellowships;
- Exhibition record, which can be divided by the prominence of the individual (solo/two-person/group exhibition);
- Commissions;
- Bibliography, where entries published about the individual's work are referenced;
- Publications;
- Participation in related events, such as conferences, seminars, and similar.

The approach of integration of information that is featured in this CV format into the Europass CV format is very similar to the previously analyzed format: teaching experience can be considered work experience, and as such is included in the "Work Experience" section, properly labeled; all of the other fields can be included as items of the "Additional Information" section, each with its proper level of detail and formatting.

3.4.1.3 *Other CVs*

From the previous two examples, a strategy can be established to adapt the Europass CV format to satisfy the need of any individual, regardless of his area of expertise and consequently the information that he wants to include in his CV.

When the individual holds a professional position at a company or an institution for a relatively long amount of time (in the range of months), irrespective of the type of work (for example, teaching and

researching) those positions can be considered work experience, and as such they can be included in the "Work Experience" section of the Europass CV format.

Any other information that does not have a standard dedicated section in this format can be included in the "Additional Information" section; by default, this section contains items such as "Certifications", "Conferences", "Honors and Awards", "Memberships", "Projects", "Publications", "References" and "Seminars", but new types of items can be defined freely. The content of these items is formatted text, unbound by any structure. Any format can be successfully implemented; this makes it completely versatile. Additionally, it can also contain references to annexes.

This analysis shows that due to the low strictness of the Europass CV format in these specific sections, this format proves to be generic and simultaneously as specific and detailed as required, and so it can support a wide range of situations, besides the ones presented in this master work.

3.4.2 *Online Social Networks Analysis*

In this section the selected OSNs will be analyzed to determine what information can be obtained from them and how is that information is integrated into the Europass CV format.

3.4.2.1 *Importing from LinkedIn*

A LinkedIn profile can include large amount of information; let us consider the example in figs. 3 to 5.

Figure 3.: LinkedIn Profile Example (1/3)

Figure 4.: LinkedIn Profile Example (2/3)

Honors & Awards

Employee of the Month
Software4U
October 2008

Best Student in Class
Universidade do Minho
July 2008

Volunteer Experience & Causes

Kitchen Aid
Soup Kitchen
November 2002 – February 2006 (3 years 4 months) | Poverty Alleviation

Projects

Raftr
July 2007 – November 2008

Boat Sharing Manager

3 team members

João Miguel Rodrigues
Software Developer at Crucible Softwar...

Roberto Maia

Manuel Silva

Publications

Portugal's Best Sailing Locations
Small Publisher
2013

Figure 5.: LinkedIn Profile Example (3/3)

From the page header we can determine the first and last names of the individual, the city and country he lives in, his area of expertise, and other elements that are expanded upon below.

After that, his experience his listed, with two items in two different companies. The names of the companies are specified, as is their location and the period in which the individual has worked there.

The user's education follows, specifying the institution where he obtained a degree, that degree's title and level, and the period of time he took to finish the degree.

His birthday and marital status are listed, as well as his skills and languages.

A few awards are indicated, mentioning the title, the awarding entity and the time it was awarded to the individual.

One volunteer experience is also mentioned, as well as one project in which several properties are displayed, including the other team members. Finally the individual's publications are shown.

There are other fields that are not featured in this example, but, as mentioned previously, they could be handled by integrating them in the "Additional Information" section of the Europass CV format, like it happens with the "Volunteer Experience" field.

All of this information can be easily included in an Europass CV. Table 3 shows where each field can be included in the Europass XML format.

Field	XPath
First Name	`/SkillsPassport/LearnerInfo/Identification/PersonName/FirstName`
Last Name	`/SkillsPassport/LearnerInfo/Identification/PersonName/Surname`
Birth Date	`/SkillsPassport/LearnerInfo/Identification/Demographics/Birthdate/`
Address	`/SkillsPassport/LearnerInfo/Identification/ContactInfo/Address`
Work Experiences	`/SkillsPassport/LearnerInfo/WorkExperienceList`
Education	`/SkillsPassport/LearnerInfo/EducationList`
Honors And Awards	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Code[.="honors_awards"]]`
Skills (Job Related)	`/SkillsPassport/LearnerInfo/Skills/JobRelated`
Skills (Languages)	`/SkillsPassport/LearnerInfo/Skills/Linguistic/ForeignLanguageList`
Projects	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Code[.="projects"]]`
Publications	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Code[.="publications"]]`
Volunteer Experience	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Volunteer Experience"]]`
<Other Field>	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="<Other field>"]]`

Table 3.: Paths of fields from LinkedIn integrated into the Europass XML format

3.4.2.2 *Importing from Google Scholar*

Google Scholar handles publications of several types, matches these with their authors and also manages citations between publications. For each author, it provides a dedicated profile page where those author's publications are listed. Each publication also has a dedicated page, where the following properties are detailed:

- Title
- Authors
- Date of Publication
- Publisher
- Conference
- Journal
- Pages
- Issue
- Volume
- Book

Four types of publications were identified: books, conference papers, journal articles and chapters in books.

Figures 6 and 7 show examples of a Google Scholar profile and publication, respectively.

3.4. Inclusion of Information

Figure 6.: Google Scholar Profile Example

Figure 7.: Google Scholar Publication Example

Publications have a standard location in the Europass CV format; they are located in the "Additional Information" section, in an item designated for that purpose. Technically, they are included inside an "Achievement" element, characterized by a specific predefined type. The XPath that identifies this element is

```
/SkillsPassport/LearnerInfo/AchievementList/
Achievement[descendant::Code[.="publications"]].
```

However, since Google Scholar's information is sufficiently detailed to thoroughly identify the type of publication, these publications can be divided into 4 different sections, one for each of the identified types, and each can have its own formatting.

The paths of the resultant elements in an Europass XML document are described in table 4.

Publi-cation Type	XPath
Books	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Publications (Books)"]]`
Chapters in Books	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Publications (Chapters in` `Books)"]]`
Journal Articles	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Publications (Journal` `Articles)"]]`
Con-ference Papers	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Publications (Conference` `Papers)"]]`

Table 4.: Paths of publications from Google Scholar integrated into the Europass XML format
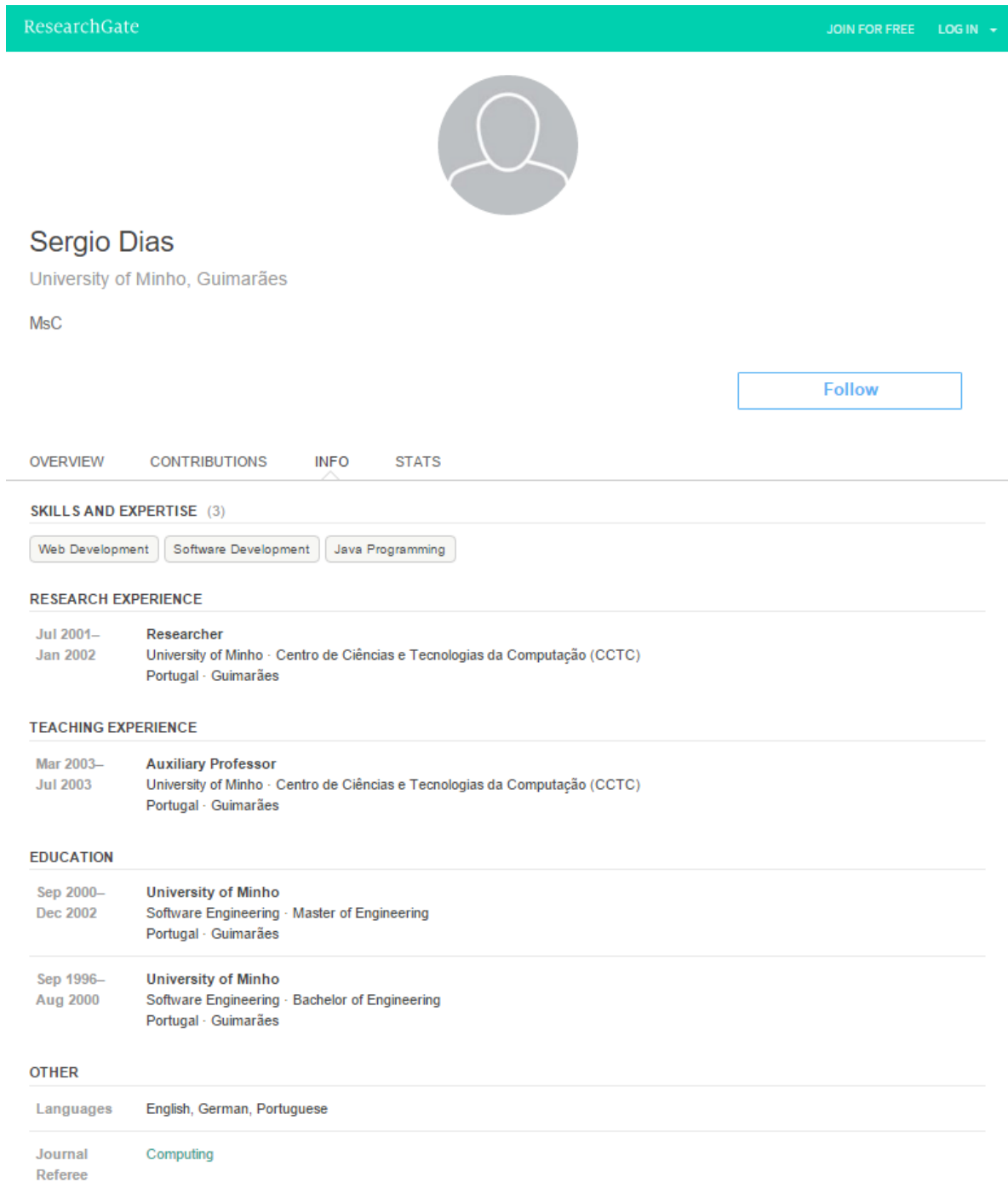
3.4.2.3 *Importing from Research Gate*

Research Gate is an academic social network, where users can maintain a profile about themselves. That profile has a page of personal information, and another to display the user's publications. Since that Google Scholar already supplies publications information and that Research Gate is poor in the bibliographic sense, this aspect was disregarded. In each user's personal page, the following fields may be available:

- Full name;
- Education, in which each item contains the period of time, the institution, the location and the degree;
- Research experience, in which each item contains the period of time, the institution, the location, the title and a description;
- Teaching experience, in which each item contains the period of time, the institution, the location, the title and a description;
- Skills;
- Grants;
- Scholarships;
- Honors and Awards;
- Languages;
- Journal Referee;
- Scientific Societies.

Figure 8 shows an example of a Research Gate profile.

Figure 8.: Research Gate Profile Example

Most of the fields that can be available in a Research Gate profile are directly integratable into the Europass CV format, since they have a designated location. To include the "Research Experience" and "Teaching Experience" fields, they can be merged into a single set, and integrated into the "Work Experience" section, while still be distinguished by prefixing "Teaching" and "Researching" to the positions' titles. "Grants", "Scholarships", "Journal Referee" and "Scientific Societies" do not have a designated location; as such, they can be included under the "Additional Information" section each inside appropriately named items.

The location of these fields in the Europass XML format are defined in table 5.

Field	XPath
First Name	`/SkillsPassport/LearnerInfo/Identification/PersonName/FirstName`
Last Name	`/SkillsPassport/LearnerInfo/Identification/PersonName/Surname`
Education	`/SkillsPassport/LearnerInfo/EducationList`
Work Experiences	`/SkillsPassport/LearnerInfo/WorkExperienceList`
Skills (Job Related)	`/SkillsPassport/LearnerInfo/Skills/JobRelated`
Skills (Languages)	`/SkillsPassport/LearnerInfo/Skills/Linguistic/ForeignLanguageList`
Honors and Awards	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Code[.="honors_awards"]]`
Grants	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Grants"]]`
Scholarships	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Scholarships"]]`
Journal Referee	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Journal referee"]]`
Scientific Societies	`/SkillsPassport/LearnerInfo/AchievementList/` `Achievement[descendant::Label[.="Scientific societies"]]`

Table 5.: Paths of fields from Research Gate integrated into the Europass XML format

3.4.2.4 *Importing from Behance*

Behance is a repository of projects of various types, where most are represented by pictures. Each user has a personal page, where his projects are listed. Each project also has a dedicated page, where its properties are displayed, and are as follows:

- Title
- Related Topics
- Description
- Date

- Pictures

Figure 9 shows an example of a Behance project.

Figure 9.: Behance Project Example

The description of the projects themselves can all be included under a single item in the "Additional Information" section: the "Projects" field, that is already predefined in the Europass CV format, for which its XPath is

`/SkillsPassport/LearnerInfo/AchievementList/`
`Achievement[descendant::Code[.="projects"]]`.

The projects' pictures must be saved as annexes of the CV document, and then referenced in the "Projects" section. For this purpose, they are saved under the

`/SkillsPassport/AttachmentList/Attachment`

path, and further referenced under the

`/SkillsPassport/LearnerInfo/Documentation/ReferenceTo`

path and the "Projects" path as well.

<div style="text-align: right; font-size: 4em; color: gray;">4</div>

IMPLEMENTATION

Once the proposed system is fully conceptualized, its implementation can begin.

This chapter aims to clarify on the implementation of the proposed solution, by describing the framework and the methodology behind the implementation of the system's processes and consequently exposing the application of that methodology by presenting how the system's key features were implemented.

4.1 APPLICATION FRAMEWORK

The proposed system, named "CVGenie", consists in a web-based application, where users will be able to create an account and create or upload their CVs, manage them, and import and export information regarding those CVs.

The application was built using the "Ruby on Rails" web application framework.

"Ruby on Rails" [15] is a web application framework written in the "Ruby" [16] programming language. It has several key features [8]:

- **MVC architecture** It is based on the model-view-controller paradigm.

- **Database Access Library** It includes "ActiveRecord", which maps database tables to object-oriented classes and table rows to objects; this object-relational mapping (ORM) makes it easier to manipulate information that is stored in a relational database, as well as painlessly using and integrating any relational database software.

- **Basic Libraries** It includes by default a large number of libraries to perform basic web related tasks (form validation, session management).

- **Guiding Principles** It is guided by two principles that improve productivity: "Do not Repeat Yourself" (DRY) aims the decrease repetition of information of all kinds, like for example, the schema of a database, or recurring code; "Convention over Configuration" (CoC), which entails that a developer only has to specify and configure unconventional aspects of his application, otherwise convention always is applied, like if a class's name is "Car", its database table name is "Cars".

- **Extensibility** The framework nurtures an environment of extensibility, and a very large amount of additional libraries are available to support any type of application.

Several additional libraries were used, besides the default set included by Ruby on Rails. These enabled to implement the required features in CVGenie. They are as follows:

- "Devise", for authentication management;
- "Paperclip", for file management;
- "Nokogiri", for XML parsing and manipulation;
- "OAuth2", for authentication and authorization with web services, using the OAuth standard;
- "Mechanize", for basic web scraping, as detailed in section 4.2.2.1;
- "Watir" and "Watir-webdriver", for advanced web scraping, as detailed in section 4.2.2.2, used in conjunction with the "PhantomJS" browser.

4.2 DATA RETRIEVAL METHODOLOGY

To retrieve data relevant to a CV from online media, several different methods can be employed. The methods that were used are described in this section.

4.2.1 *Application Program Interface (API)*

To obtain data from a website, the most typical approach is to use an API, if it is available. An API will provide data in a reliable and organized manner, following industry standards, making it much simpler to process that data afterwards.

Frequently, the access to APIs is restrained. Developers must register their applications for the platform to grant them access to the API, and to set up how the platform should react to their queries.

Additionally, if an application acts on the behalf of a user of the platform that is being accessed, elaborate authentication methods, such as OAuth, can be employed. OAuth is an open standard for authorization. It allows a user to mediate the authorization process between two systems, allowing access to the user's resources that are present in a system to a third party system, without compromising his credentials.

Website APIs typically follow the REST architecture style, in that they communicate via Hypertext Transfer Protocol, using ordinary HTTP verbs (such as GET and POST) and referring to resources by specific URLs. Each request is self-sufficient, using a token for authentication, and other parameters to specify the request. Responses also adhere to standards, and often use JSON, XML or HTML.

The advantages of this method are clear:

- The API is supported by the website that acts as a source of information, having the guarantee that the website will supply data via their API reliably;

- The website gives explicit authorization to retrieve information using the API;

- The data is provided in an organized manner, decreasing the need to parse it and normalize it posteriorly.

However, there are also some disadvantages:

- Developers are limited by what the API provides, regardless of what information the website actually contains, which can be a big limitation.

4.2.2 *Web Scraping*

When a website does not provide an API, the only viable alternative to retrieve its information is by web scraping.

Web scraping, also known as web harvesting or web data extraction, is a process by which data from websites can be retrieved.

Typically, most websites consist of HTML. HTML is a structured language, and as such the information that is elected to be retrieved can be pinpointed in the HTML documents that websites serve, and can then be retrieved.

Simple websites function essentially that way; when the page is loaded you can simply gather from the page whatever information you desire, since the location of that information in the page is consistent.

However, nowadays most websites are dynamic in nature, using technologies such as JavaScript to load extra information at specific times or when specific events happen, such as when the user interacts with a button, or scrolls down the page. At those times, extra information can be loaded via AJAX, for example, and the page layout can consequently change.

There are various tools and methods that can be used to perform web scraping:

- Retrieving web pages directly by sending HTTP requests directly to web servers, and posteriorly using a parser to analyze and retrieve information from the HTML page.

- Using embedded web-browsers that load web pages and process any client-side scripts present in the page; the information can then be retrieved from the DOM tree the browser creates.

- Using automated web analyzers that try to analyze a web page as a human being would.

The first two methods will be described in greater detail in the following sections.

Web scraping is very good method to retrieve information from websites. It has the advantages of allowing access to everything that is usually available on a web page and of being relatively simple to implement.

Unfortunately, there are some serious disadvantages with this method:

- The web pages of a website can be altered at any time by the webmaster. If that happens, the web pages might not be available at the same address, or may have their layout changed; the pages might not be found or, if they are, the HTML elements where the information used to be

retrieved from might be displaced, or no longer exist; either way the scraping mechanism fails. This makes this method very fragile.

- It might not be legal in some websites; depending on the Terms of Service of the website in question, retrieving information using automatic processes might not be allowed, and doing so might have legal consequences.

- There might be increased polling limits; if web pages are requested at a much faster rate than an ordinary user would, the website in question might reject requests temporarily or permanently after that rate has been exceeded. This limits the software to request within those limits, which are established with the regular user in mind, not automated processes, which can be detrimental to performance.

4.2.2.1 *Basic Web Scraping*

Most websites have a thing in common: they usually serve HTML documents for their pages. In some cases, these pages are mostly (or completely) static, meaning that the layout of the page and the structure and contents of the HTML document does not change after the page is loaded.

In these cases, web scraping can be performed easily.

First of all, the web pages have to be retrieved. This can be done by using socket programming to send a pre-built HTTP request to the website, which will then analyze and serve a web page that complies with the request. The HTTP response body usually consists of an HTML document. This HTML can then be supplied to an HTML parser. Several parsers can then supply a tree based interface to the document, and allow XPath and CSS Path based traversal and selection of elements on the HTML document. If the programmer defines a map of carefully pre-selected XPath or CSS Path expressions, a tool can be created that automatically retrieves the information from a website.

Using the Ruby on Rails "Mechanize" (for the web page retrieval and interaction) and "Nokogiri" (for the HTML parsing) gems, this process was successfully implemented in CVGenie.

4.2.2.2 *Advanced Web Scraping*

Often nowadays websites are very interactive and dynamic. As an example we can consider Facebook: we can submit new posts or comment on other users' posts without leaving the current page, we receive notifications when other users interact with us without reloading the page and if we scroll down, new posts continue to load indefinitely.

For all of these things to be possible, simple HTML documents are insufficient. Other technologies are used to do client-side scripting that enable these features, from simple layout modification of the web page, to dynamic loading of new content. One example of this is JavaScript, which enables DOM modification, and loading of new content via AJAX, respectively.

To perform web scraping on websites that make use of these technologies is challenging, since the software cannot simply retrieve an HTML document and parse it. The client-side code that is supplied with the web page must be executed to fully load the page properly, and only then can the information in the page be scraped.

There are two alternatives to obtain this information:

- Study how the client-side code interact with the server, to understand how it makes the dynamic requests for extra information, and replicate those requests, bypassing the execution of that client-side code that is embedded in the web page. This approach has many issues, since it requires a bigger effort due to the reverse engineering of that code, and often websites use authentication systems that make it impossible to generate valid requests.

- Run a fully-fledged web browser that is controlled programmatically. This way the whole page is loaded and processed as an ordinary web browser would, and the information that is required will be available. The disadvantage of this method is that it is much slower and has a larger impact on performance than other methods, since a real browser is being run, which entails a big overhead in processing time and power. It is however reliable, since a browser is used much the same way an ordinary user would use it, and the information that is required is equally available.

The second method was used in CVGenie, since it is much more viable.

In terms of usage, the process is very similar to the basic web scraping described previously; the only thing that differs is the libraries used to do so.

The Ruby on Rails "Watir" and "Watir-webdriver" gems were used, along with the "PhantomJS" browser.

PhantomJS is a headless full-stack web browser. It behaves like an ordinary modern web browser, with the exception that it does not provide a graphical user interface; it can only be controlled programmatically. Because of this, it has a smaller performance footprint than regular browsers.

Watir and "Watir-webdriver" are a family of libraries that allow web browser automation. They supply an easy-to-use API to control web browsers in the Ruby programming language.

In fact, the page traversal and selection of elements is very similar to the API provided by the "Mechanize" and "Nokogiri" gems, which allow using XPath and CSS paths to select elements. It however also provides an API to control the browser itself, like for example clicking links, scrolling through a web page, etc.

4.3 DATA INTEGRATION AND XML MANIPULATION

By using the Europass XML format as the data storage format for the CVs in the system, any process that imports or exports data from the system will forcefully require manipulating XML files in this format.

4.3.1   *Conversion Mechanism*

To facilitate the manipulation of data, it is useful to do so by using an intermediate format native to the programming language that is used whenever possible, and only use its XML representation when the data is ready to be imported, or when it needs to be exported. With that intent, a conversion module was implemented.

The intermediate format that was used is equivalent to the Europass XML format in terms of structure and contents. The only difference is that it is based on the Ruby programming language, which enables the developer to better manipulate data in this format, since a wider range of libraries and methods are available.

To convert data from the intermediate format into an XML representation, the algorithm is as follows:

- Data is supplied in an hash structure; its contents represent the node that is returned after conversion;
- It is detected if the hash contains a key associated with attributes; if it does and its value is a hash, each item in that hash is added as an attribute to the resulting XML node;
- It is detected if the hash contains a key associated with the resulting node's contents;
- The type of the contents value is assessed;
- If it is a simple string, the resulting node will have a single text node as its contents;
- If it is a hash, it is evaluated as well, and each item in that hash is analyzed;
- The type of each item is assessed;
- If it is a hash, the algorithm is called recursively, and the resulting node is added as a child of the current node;
- If it is an array of hashes, the algorithm is called recursively for each array element, and each resulting node is then added as a child of the current node.

Consider the example in listing 16 about the "WorkExperienceList" node in the intermediate format; hash structures are represented by curly brackets ({ }), arrays are represented by angle brackets ([ ]), key-value pairs are separated by colons (:), and "Contents" and "Attributes" are simplified to "Cts" and "Ats" respectively.

```
1    {
2       Education: {
3         Cts: {
4           [
5             {
6               Title: Job 1,
7               Period: {
8                 Cts: {
9                   From: {
10                    Ats: { year: 2001, month: 2 }
11                  },
12                  To: {
```

```
13                    Ats: { year: 2001, month: 3 }
14                  }
15                Current: false
16              }
17            }
18          },
19          {
20            Title: Job 2,
21            Period: {
22              Cts: {
23                From: {
24                  Ats: { year: 2001, month: 3 }
25                },
26                Current: true
27              }
28            }
29          }
30        ]
31      }
32    }
33  }
```

Listing 16: Intermediate representation of a "WorkExperienceList" element

This would be converted into the fragment in the Europass XML format represented by listing 17.

```
1   <WorkExperienceList>
2     <Education>
3       <Title> Job 1 </Title>
4       <Period>
5         <From year="2001" month="2"/>
6         <To year="2001" month="3"/>
7         <Current> false </Current>
8       </Period>
9     </Education>
10    <Education>
11      <Title> Job 2 </Title>
12      <Period>
13        <From year="2001" month="3"/>
14        <Current> true </Current>
15      </Period>
16    </Education>
17  </WorkExperienceList>
```

Listing 17: XML representation of a "WorkExperienceList" element

The opposite conversion works very similarly:

- A node is supplied to the algorithm;
- The contents of the node are analyzed;
- If the node's content is text, a hash is returned with its contents being that text;
- Otherwise, the node's children are analyzed;

- The children nodes' names are fetched; for each name, it is assessed how many children exist with that name;

- If only one child exists with that name, it is converted by calling the algorithm recursively and then it is added to the current hash's contents, by using the child's name as the key, and the resulting conversion of the child node as the value;

- If several children exist with that same name, they are all converted using the same algorithm recursively and then they are added to the current hash's contents, with the key being the children's name, and the value being an array of the resulting converted children nodes.

Do note that there is one use case that is not supported by the algorithms above: when nodes with both text content and other children, like the example represented by listing 18:

```
1   <Node1>
2     Text 1
3     <Node2> Text2 </Node2>
4   </Node1>
```

Listing 18: Use case that is supported by the conversion mechanism

This particular use case is not supported since it is not present in the Europass XML format's schema, and so it is not necessary for the purposes of CVGenie.

4.3.2 *Integration and Manipulation*

To handle the data integration process in a modular manner, the sections of the Europass XML format that are supported were enumerated, and the XPath of each was specified. The importing process, regardless of the source, works like this: for each section, a replacement node is fashioned from the data extracted from that source, and any existing node with the same path (hence in the same section) is replaced by the new node. If there is not a node for that path, and in fact the nodes referred in that path do not exist, they are created to accommodate the new node.

This algorithm is used in most cases; however there are some special cases where it is not sufficient.

One of these cases is the management of attachments. In the Europass XML format, all of the CV's attachments exist under a specific section. Consequently, when adding new attachments we cannot remove the previous attachments using the normal method, since it uses a static path for that section, and using so would remove all existing attachments; instead, a custom algorithm is used to remove only the attachments that should no longer be present in the CV, leaving the rest unaffected, and the new ones are added to the attachment section.

4.4 EUROPASS WEB SERVICES

As mentioned previously, CEDEFOP has a web portal in which they describe the public available set of web services that can be used to integrate the Europass format into existing software applications. These web services will prove invaluable to the implementation of CVGenie.

These web services include a REST API and a remote upload service.

The REST API exists mainly to satisfy use cases related to conversion. It can handle as input Europass XML files from version 3.0.0 and above, PDF files with the same XML files embedded, and Europass JSON files from version 3.0.0 and above as well. As an output, the formats that are supported are the same as the ones supported for input, as well as Microsoft Word 97-2003 files and OpenDocument version 2 Text files.

Additionally, the REST API also provides utility services to extract XML attachments that are embedded in Europass PDF documents, and another to upgrade Europass documents from previous versions (version 2 and below).

The remote upload service can receive an Europass XML file and open the Europass XML Online Editor with the data included in that file being pre-populated in the form.

4.5 IMPORTING FROM EUROPASS XML DOCUMENTS

To allow users to maintain the CV documents that they already have and complement them with up-to-date additional information, it must be possible to import existing documents into the system.

Importing information from CVs in a file format will be mainly used as the initial source of information. CVGenie's main purpose is to import information that is available online into existing CV documents, and as such the system will not be able to merge existing CVs with CVs that are available as offline files; files imported into the system will always completely replace existing information relative to a CV.

Europass documents are able to be imported into the system easily. Since the system is based on this format, these files can be immediately integrated.

In the importing process the integrity and validity of the documents will be evaluated thoroughly, by way of the XML Schema made available by the Europass team.

After the supplied document is deemed valid a CV's content will be completely replaced by the information in the imported document. Afterwards the user can import information from additional sources.

4.6 IMPORTING FROM LINKEDIN

4.6.1 *Data Retrieval*

LinkedIn is a professional social network, where the users define professionally relevant information on their profiles. There is a multitude of fields that LinkedIn has available per profile that are of interest, as described in section 3.4.2.1.

LinkedIn has several web services available for developers to use in their applications to integrate LinkedIn's features and data [7].

In order to access that API, CVGenie was registered as an application in the LinkedIn Developers platform.

In the importing process, users are required to be authenticated and to authorize CVGenie to access their data contained in LinkedIn. In this case, this process is established using the OAuth2 standard. When it is successful, LinkedIn provides CVGenie with a token that can be used to fetch data about that user from LinkedIn for a limited time.

Unfortunately, LinkedIn recently changed the terms of service of their API to restrict the information that the developers have, based on the purpose of their applications [4]. Basically, only developers that intend to use the information in the context of a company's job listing page can obtain explicit approval to use the complete profile information. Other developers can only make use of the following fields:

- First and last name;
- Location;
- Industry;
- Networking details, such as number of connections in the service;
- Specialties;
- Last active job position.

Of these, CVGenie imports the user's name, his location and his last active job position. For each of these fields, a request is made, specifying the field name and specifying the intended format as JSON. When the response is obtained, if deemed successful, it is parsed from JSON into an internal representation.

4.6.2 *Data Normalization*

Very little parsing and normalization are required after the retrieval of data from LinkedIn, since the data is already supplied in the standard JSON format. However, some location and date fields are parsed to separate them into logical parts, for posterior integration.

4.6.3 *Data Integration*

After the normalization process the data can be integrated in a Europass XML CV document. The location of each field was pre-determined in table 3 (section 3.4.2.1), and it was successfully implemented, for the fields that are available (first name, last name, address and work experience).

4.7 IMPORTING FROM GOOGLE SCHOLAR

4.7.1 *Data Retrieval*

In Google Scholar, each author has a personal profile page where his publications are enumerated; each publication in turn also has a page detailing the properties indicated in section 3.4.2.2.

Through the availability of these properties, four different types of publication can be identified: books, conference papers, journal articles and chapters in books.

Since Google Scholar does not have an API available, web scraping methods were used to retrieve the information from this source.

Analyzing the web pages that Google Scholar serves, it can be seen that the user profile page is mostly static; it presents the first 20 publications that a user has, in no apparent order. To view the next set of publications, the right arrow can be pressed; this will load a new web page using additional query parameters - `cstart` and `pagesize`, which determine the starting index and number of publications per page, respectively - which means that all of the user's publications can be loaded statically. This makes Google Scholar a prime target for basic web scraping using socket programming and HTML parsing.

Using the Ruby on Rails "Mechanize" (for the web page retrieval and interaction) and "Nokogiri" (for the HTML parsing) gems, this process was successfully implemented in CVGenie for Google Scholar. The algorithm is as follows:

- The supplied profile URL is used to load the web page for the user's profile using Mechanize.
- The publication pages are iterated over until no more publications can be found.
- For each iteration, each publication link is retrieved.
- For each link a new web page is loaded that details the information of a publication.
- All of the fields concerning that publication are retrieved. Each individual field has a respective XPath expression that selects the appropriate HTML element from the HTML document included in the web page, and the contents of that element are saved.
- The full list of publications is compiled and returned.

4.7.2 *Data Normalization*

After fetching the data related to publications, the type of each item is unknown, and the fields' values might be invalid. The final result should also be formatted appropriately. For these reasons, this data is submitted to a normalization process.

First of all, the list of publications is traversed and sorted by type:

- Publications that include the "book" field are identified as chapters in books;
- Publications that include the "journal" field are identified as journal articles;
- Publications that include the "conference" field are identified as conference papers;
- Publications that do not meet the criteria above are identified as books.

Then each publication's set of fields is further normalized:

- The "date" field is separated into individual parts (day, month and year);
- The "publisher" field is set as "Unknown Publisher" if none was provided, since it is deemed mandatory;
- For any type of publication other than books, the "pages" field is separated into individual parts (first and last pages).

Publications in a CV are described each as a citation in a single line of text. There are many citation styles; depending on the area of expertise, one is favored over the others. However, since this algorithm does not detect the area of expertise of the publications, only one citation style was chosen, and that is the Harvard Reference format.

The last step of normalization is to convert the list of publications into citations. The Harvard Reference format was followed as best as possible, considering that sometimes certain fields required by this format might not be available.

4.7.3 *Data Integration*

Once the data has been thoroughly normalized, it can be integrated in a Europass XML CV document. The section of the Europass XML schema where the publications were included was the "AchievementList" section. More precisely, for each type of publication, a new node is included as a child of the "AchievementList" node, each with an appropriate title; the contents of those nodes are each a list of the citations of the publications of that type. The location of each field was pre-determined in table 4 (section 3.4.2.2).

4.8 IMPORTING FROM RESEARCH GATE

4.8.1 *Data Retrieval*

Each Research Gate user can have a page where he presents his personal information. In that page, he can specify the fields specified in section 3.4.2.3.

Since Research Gate does not have an API available, web scraping methods were used to retrieve the information from this source.

Research Gate serves its web pages mostly in a static manner. The page that is of interest for data retrieval, the "Info" page, is, however, loaded dynamically; for this reason it was deemed necessary to use more advanced web scraping methods, which are described in detail in section 4.2.2.2.

The full algorithm is as follows:

- The supplied user profile URL is used to load the web page;
- The fields indicated previously are selected in the web page, and their values are retrieved;
- The fields are compiled into a single data set for posterior processing.

4.8.2 *Data Normalization*

To maintain coherence, the data that was retrieved must be parsed and normalized before being integrated into a CV. The following steps are executed to guarantee this:

- For the education, research and teaching experiences, the period of time is parsed, and the institution and the location are normalized (as well as the degree, for education items);
- For the achievements, their description and date of occurrence are parsed as well;
- The research and teaching experiences are merged into work experiences, while noting their type in the description of the position;
- The achievements are sorted by type (grant, scholarship and award);
- The remaining fields are parsed and normalized as well.

4.8.3 *Data Integration*

Once the data has been thoroughly normalized, it can be integrated in a Europass XML CV document. The location of each field was pre-determined in table 5 (section 3.4.2.3), and it was successfully implemented.

4.9   IMPORTING FROM BEHANCE

4.9.1   *Data Retrieval*

A Behance user can make available in his profile any number of projects, which have their dedicated pages as well. Each project has the properties mentioned in section 3.4.2.4.

Since Behance does not have an API available, web scraping methods were used to retrieve the information from this source.

By observing the web pages that Behance serves, it can be detected that Behance loads each web page in a completely dynamic manner, making it impossible to retrieve its information using basic web scraping methods. To do so, we then use instead the methods referred in section 4.2.2.2.

The data fetching begins with the user supplying a list of URLs to Behance projects. It has been decided to use this workflow for several reasons: a user might not want to display all of his projects in his CV, and this way he can simply select the ones he wants; when including projects with pictures, the scraping process may take a long period of time to complete, and the CV can easily become over-encumbered with attachments.

The full algorithm for data retrieval from Behance is as follows:

- The supplied projects' URLs are used to load the respective web pages;
- For each web page, the project's property values are retrieved;
- Depending on the settings, either the first, all or none of the pictures of that project are included;
- The full list of projects is compiled and returned.

4.9.2   *Data Normalization*

After fetching the data related to the projects, it is normalized to guarantee that the end result is compliant with the minimum standards necessary to include them in a CV. The steps in the normalization process for each project are as follows:

- For each picture, an identifier is assigned to be able to manage the CV picture attachments posteriorly;
- For each picture, the image referred by the image URL that was obtained is retrieved, and converted into a Base64 string;
- For each picture, the content type of the image file is also noted;
- The description is striped of leading and trailing white-space;
- The date of publication is parsed and formatted appropriately.

4.9.3 *Data Integration*

Once the data has been thoroughly normalized, it can be integrated in a Europass XML CV document. Behance data is very similar to Google Scholar data regarding its integration in an Europass XML CV document; Behance projects are included as an "Achievement" node under the "AchievementList" section. However, there is a key difference: Behance projects can contain pictures, which can be included in the CV if the user chooses so. Consequently, these attachments must also be included in the Europass XML file. The integration process occurs as follows:

- A single node is created containing all of the selected projects. Its title is "Projects", its description holds a list of the selected projects, and the references to the picture attachments are included as well via their previously generated image identifiers. The XPath expression for this node is
  ```
  /SkillsPassport/LearnerInfo/AchievementList/
  Achievement[descendant::Label[.="Projects"]].
  ```

- For all of the pictures in every project, a new node is created. It contains all of the picture's properties specified previously. The XPath expression for all of these nodes is
  ```
  /SkillsPassport/AttachmentList/Attachment.
  ```

- An additional list of references is set at the XPath path
  ```
  /SkillsPassport/LearnerInfo/Documentation/ReferenceTo.
  ```

Do note that importing projects from Behance does overwrite any existing projects in a CV. To implement this and still guarantee that any existing attachments related with something other than projects are not removed from the CV, a more complex algorithm was implemented:

- Before importing, the projects' "Achievement" node that is referred above is retrieved;

- If that node is retrieved successfully, the attachment references contained within are analyzed and the matching "Attachment" and "ReferenceTo" nodes contained in the document are removed.

- The projects' "Achievement" node is then removed.

4.10 CONFLICT MANAGEMENT

Depending on the source, the complexity of the conflict management in CVGenie varies.

In both Google Scholar and Behance, it was deemed unnecessary to create complex conflict handling mechanisms. With Google Scholar, all of a user's publications are imported, replacing the existing publications in the CV, and the user can posteriorly select which publications are worth keeping and which are not. With Behance, the same can be made when configuring the importing process, refering only the projects worth importing, replacing the existing ones in the CV as well.

LinkedIn and Research Gate, in the other hand, offer more varied information; opposite to the OSNs mentioned above, the information that LinkedIn and Research Gate contain can be used to fill several fields of a CV. As such, it was deemed necessary to provide users with a more intelligent mechanism to handle conflicts when importing information from these OSNs.

The mechanism that was implemented to handle conflicts between existing CV information and information from LinkedIn and Research Gate works as follows:

- The information from these OSNs is retrieved, normalized and converted into the intermediate format introduced in section 4.3.1 for comparison;
- The information currently included in the user's CV is retrieved as well, and converted to the same intermediate format;
- The presense of each field of a CV in both the current CV and the data retrieved online is assessed;
- Fields that are present in both are compared;
- If they contain exactly the same data, they are ignored; otherwise they are presented to the user, in a human readable format (see fig. 10 for an example), for him to make an informed decision about if each field should be overwritten or not, based on the current value of that field in the CV and the new value obtained from OSNs;
- When the user submits which fields should be overwritten, the information that was retrieved from OSNs for both those fields and the fields for which conflicts were not detected is imported into the user's CV.

1. ○ Position: Researching (Group Leader)
   ○ Employer: University of Minho, Departamento de Informática (DI), Language Processing Group (gEPL) (Contact information available)
   ○ From: 1994-01 (Currently employed)
2. ○ Position: Teaching (Associate Professor)
   ○ Employer: University of Minho, Departamento de Informática (DI) (Contact information available)
   ○ From: 1996-01 (Currently employed)
3. ○ Position: Teaching (Auxiliary Professor)
   ○ Employer: University of Minho, Departamento de Informática (DI) (Contact information available)
   ○ From: 1993-01
   ○ To: 1996-01
4. ○ Position: Teaching (Assistant Lecturer)
   ○ Employer: University of Minho, Departamento de Informática (DI) (Contact information available)
   ○ From: 1981-01
   ○ To: 1993-01

Figure 10.: Work experiences in the human readeable format used for displaying conflicts

4.11 EXPORT CV IN THE EUROPASS XML FORMAT

Users can at any time export their CV as a file in the Europass XML format, which can then be used in another software that is compatible with this format.

Since users' CVs are saved as Europass XML files at all times in CVGenie, to export them they are simply served directly to users without requiring any transformation.

It should be noted that, if the user has not imported information either from a file or online media, the CV file that will be supplied to the user will be devoid of information, but it will still be valid according to the Europass XML format.

4.12 EXPORT CV AS PDF DOCUMENTS (AMONG OTHER FORMATS)

As mentioned in section 4.4, CEDEFOP provides web services to convert Europass documents between a set of file formats. These formats include Portable Document Files (PDF), Microsoft Word 97-2003 files and OpenDocument version 2 files.

These web services are used in CVGenie to export CV documents in these formats.

The algorithm to run this process is as follows:

- Retrieve the Europass XML document for the user's CV;
- Generate an HTTP POST request, declaring the content type of the request as "application/xml", and including in the request body the XML document's contents;
- Send the HTTP request to the appropriate URL via HTTPS, according to the REST API; note that each format has a dedicated URL (for PDF it is
  `https://europass.cedefop.europa.eu/rest/v1/document/to/pdf`,
  for example);
- Parse the response; if deemed valid, the response's body is a file in the requested format;
- Serve the retrieved file to the user.

4.13 EXPORT CV AS HTML WEBSITES

In CVGenie, CV documents can also be provided as simple HTML static web pages. These include the most relevant information, like an individual's name, address, work experiences and education, as well as his publications and other achievements.

These static web pages are generated when the user requests them, using information included in his CV.

To perform conversion of information into a human readable format, the same module that was used for conflict handling was also used for exporting HTML web pages. An example of this format is

displayed in fig. 10 (section 4.10). The reason behind this is that both functionalities required CV information to be converted from the XML format into a human readable HTML fragment.

The web pages also offer navigation and styling that is simple but functional, implemented by embedding JavaScript and CSS in the document itself; this way a user can download it and immediately start using it as his personal web page, without concern for these aspects.

4.14 EDIT IN AN EXTERNAL EDITOR

As an unplanned feature, since there is an Europass web service that allows so, CVs in CVGenie can be exported directly into an external editor provided by the Europass team. Users can then use that editor to review the results of importing data from CVGenie and manipulate the CV how they see fit.

This feature was implemented by creating a form on the main CV web page that provides a copy of the CVs Europass XML file's contents when submitted into the address
`https://europass.cedefop.europa.eu/editors/remote-upload`,
which then redirects the user to the Europass editor, and prompts them if he wants to include the supplied data into the editor.

<div style="text-align: right; font-size: 3em;">5</div>

FEATURES DISPLAY AND CASE STUDIES

After implementing the system, we tested it with different case studies, in order to prove that it was correctly implemented and guarantee that it satisfies the requirements stated in section 3.1.

With that goal, we first present the final product's features (section 5.1); then (section 5.2), we describe how the system was subjected to tests by exploring selected case studies.

Of all the system's features, the following were deemed most important and will be displayed:

- Importing from LinkedIn;
- Importing from Google Scholar;
- Importing from Research Gate;
- Importing from Behance;
- Managing conflicts in the importing process;
- Exporting CV documents in several formats;
- Editing a CV in the external Europass editor.

5.1 FEATURE DISPLAY

5.1.1 *Importing from LinkedIn*

After a user creates a new CV in CVGenie, the main CV page is presented to him. This page is shown in fig. 11. At this point, the user's CV is empty.

If he navigates to the "Importing from LinkedIn" page and has not signed in on LinkedIn, he will be suggested to do so (appendix A.1).

The LinkedIn sign-in prompt will appear (appendix A.2), and when the user signs in, he will be allowed to import information from LinkedIn (appendix A.3).

If he does so, the importing process starts, and if no conflicts are found, the user's CV will be modified to include information from his LinkedIn profile.

Figure 11.: CVGenie's main CV page

Figure 12 shows an example LinkedIn profile, and fig. 13 shows the result of importing the information from that profile into an empty CV.

Figure 12.: LinkedIn profile example

Figure 13.: CV with imported LinkedIn information example

5.1.2 *Importing from Google Scholar*

From the main CV page, if a user navigates to the "Importing from Google Scholar" page, he will be prompted to submit his Google Scholar profile URL, and a form will be available to help him search for it (appendix A.4).

If he has already submitted a Google Scholar profile URL previously, it will already be available to be selected as a source (fig. 14).

Regardless of he specifies a new profile URL or uses one specified previously, when pressing the "Import" button, the importing process will start, and once executed successfully he will be redirected to the main CV page.

Figures 15 and 16 show examples of a user's Google Scholar profile and one of his publications, and fig. 17 shows the result of the importing process.

Figure 14.: CVGenie's "Import from Google Scholar" page

Figure 15.: Google Scholar profile example

Figure 16.: Google Scholar publication example

5.1. Feature Display

![europass logo] Curriculum vitae

PERSONAL INFORMATION	Not Assigned
ADDITIONAL INFORMATION	

Publications (Books)

- Gropper, S & Smith, J 2012, *Advanced nutrition and human metabolism*, Cengage Learning
- 1995, *Cyclohexanone or cyclohexanol derivatives*, Unknown Publisher
- 1994, *Unguinol and analogs are animal growth permittants*, Unknown Publisher
- 1993, *Antibiotic eicosenoic acids*, Unknown Publisher
- Smith, JL 1993, *Sexual maturation in sea-and freshwater phases of Pacific salmonids and investigations of reproductive hormone involvement in osmoregulatory adaptation*, Unknown Publisher
- 1992, *Antibiotic agent*, Unknown Publisher
- Smith, JL 1983, *Effect of isolated soy protein on the bioavailability of copper in ground beef*, Unknown Publisher
- Smith, J, Berkheimer, SD, Werner, CJ & Bundy, BC , *Supplemental Information*, Unknown Publisher

Publications (Journal Articles)

- Zwart, R, Sher, E, Ping, X, Jin, X, Sims, J, Chappell, A, Gleason, S, Hahn, P, Gardinier, K, Gernert, D, Hobbs, J, Smith, J, Valli, S & Witkin, J 2014, 'Perampanel, an Antagonist of α-Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid Receptors, for the Treatment of Epilepsy: Studies in Human Epileptic Brain and Nonepileptic Brain and in Rodent Models', *Journal of Pharmacology and Experimental Therapeutics*, vol. 351, no. 1, pp. 124-133
- Devlin, RH, Sakhrani, D, Biagi, CA, Smith, JL, Fujimoto, T & Beckman, B 2014, 'Growth and endocrine effect of growth hormone transgene dosage in diploid and triploid coho salmon', *General and comparative endocrinology*, vol. 196, pp. 112-122
- Leggatt, RA, Biagi, CA, Smith, JL & Devlin, RH 2012, 'Growth of growth hormone transgenic coho salmon Oncorhynchus kisutch is influenced by construct promoter type and family line', *Aquaculture*, vol. 356, pp. 193-199
- Higgs, D, Sutton, J, Kim, H, Oakes, J, Smith, J, Biagi, C, Rowshandeli, M & Devlin, R 2009, 'Influence of dietary concentrations of protein, lipid and carbohydrate on growth, protein and energy utilization, body composition, and plasma titres of growth hormone and insulin-like growth factor-1 in non-transgenic and growth hormone transgenic coho salmon, Oncorhynchus kisutch (Walbaum)', *Aquaculture*, vol. 286, no. 1, pp. 127-137
- Campbell, PM, Fernandez, MP, Royston, S, Smith, JL, Poppelen, PV, Ikonomou, MG & Devlin, RH 2006, 'Male Coho Salmon(Oncorhynchus kisutch) Exposed to a Time-Course of Urban Sewage Effluent Exhibit a Sporadic Low Incidence of Sex Reversal and Intersex', *Water quality research journal of Canada*, vol. 41, no. 3, pp. 235-243
- Afonso, LO, Smith, JL, Ikonomou, MG & Devlin, RH 2002, 'Y-chromosomal DNA markers for discrimination of chemical substance and effluent effects on sexual differentiation in salmon.', *Environmental health perspectives*, vol. 110, no. 9, p. 881
- Afonso, LO, Iwama, GK, Smith, J & Donaldson, EM 2000, 'Effects of the aromatase inhibitor Fadrozole on reproductive steroids and spermiation in male coho salmon (Oncorhynchuskisutch) during sexual maturation', *Aquaculture*, vol. 188, no. 1, pp. 175-187
- Afonso, L, Iwama, G, Smith, J & Donaldson, E 1999, 'Effects of the aromatase inhibitor fadrozole on plasma sex steroid secretion and oocyte maturation in female coho salmon (Oncorhynchus kisutch) during vitellogenesis', *Fish Physiology and Biochemistry*, vol. 20, no. 3, pp. 231-241
- Afonso, LO, Iwama, GK, Smith, J & Donaldson, EM 1999, 'Effects of the aromatase inhibitor fadrozole on plasma sex steroid secretion and ovulation rate in female coho salmon, Oncorhynchus kisutch, close to final maturation', *General and comparative endocrinology*, vol. 113, no. 2, pp. 221-229
- Freind, KO, Smith, JL & Waninger, KN 1998, 'Nutrition Services in NCAA Athletic Programs', *Journal of the American Dietetic Association*, vol. 98, no. 9, p. 81

Figure 17.: CV with imported Google Scholar information example

5.1.3 *Importing from Research Gate*

Importing from Research Gate is very similar to importing from Google Scholar: a profile URL must also be provided, and if it was provided previously, it will already be available as well (fig. 18). When importing, however, if conflicts occur they will be displayed to the user.

Figure 19 shows an example of a user's Research Gate profile and fig. 20 shows the result of the importing process.

Figure 18.: CVGenie's "Import from Research Gate" page

Figure 19.: Research Gate profile example

europass Curriculum vitae

PERSONAL INFORMATION	Graham Colditz
WORK EXPERIENCE	

January 2008–December 2010	**Researching** Barnes Jewish Hospital, Saint Louis (USA)
Jan 2007–Dec 2011	**Researching** Washington University in St. Louis, Division of Public Health Sciences, Saint Louis (USA)
Oct 2006–Present	**Researching (Adjunct Professor)** Harvard University, Department of Epidemiology, Boston (United States)
Jan 2004	**Researching** Royal Brisbane Hospital, Brisbane (Australia)
Jan 2003	**Researching** American Cancer Society, Atlanta (USA)
Jan 2002–Dec 2003	**Researching** National Cancer Institute, Division of Cancer Epidemiology and Genetics, Bethesda (USA)
Jan 2002–Dec 2004	**Researching** University of Queensland , School of Population Health, Brisbane (Australia)
Jan 2000–Dec 2007	**Researching** Massachusetts Department of Public Health, Boston (USA)
Jan 1997–Dec 2007	**Researching** Dana-Farber Cancer Institute, Department of Medical Oncology, Boston (USA)
Jan 1996–Dec 2011	**Researching** Brigham and Women's Hospital, Channing Division of Network Medicine, Boston (USA)
Jan 1983–Dec 2012	**Researching (Harvard Medical School)** Brigham and Women's Hospital, Department of Medicine, Boston (USA)
Jan 1997–Dec 2005	**Teaching** Harvard University, Department of Epidemiology, Cambridge (USA)
Jun 1988–Jul 2002	**Teaching** University of Michigan, Ann Arbor (USA)

Figure 20.: CV with imported Research Gate information example

5.1.4 *Importing from Behance*

Importing from Research Gate is somewhat similar to importing from Google Scholar and Research Gate: instead of a profile URL, one or more project URLs must be provided (fig. 21), as well as an option on how many pictures from each project may be included in the CV. The project URLs will be saved, so that they are immediately selected the next time the user wants to import Behance projects, respectively.

Figures 22 to 25 respectively show examples of a user's Behance projects and the corresponding CV resulting from importing those projects.

Figure 21.: CVGenie's "Import from Behance" page

Figure 22.: Behance profile example

Figure 23.: Behance project example

Figure 24.: CV page with imported Behance projects example

Figure 25.: CV page with imported Behance project example

5.1.5 *Conflict Management*

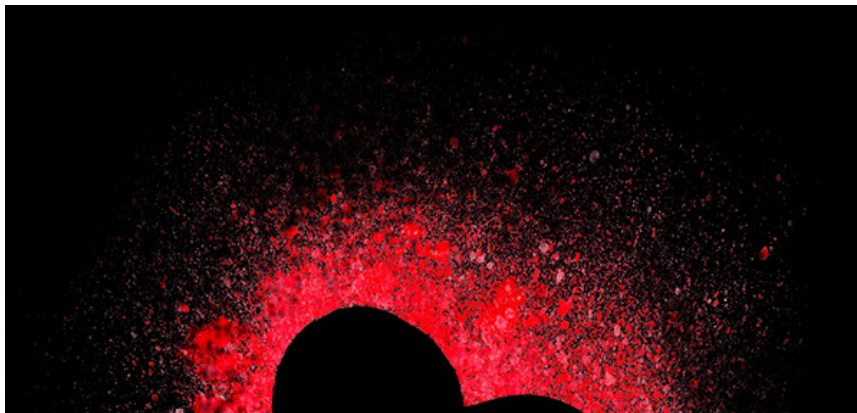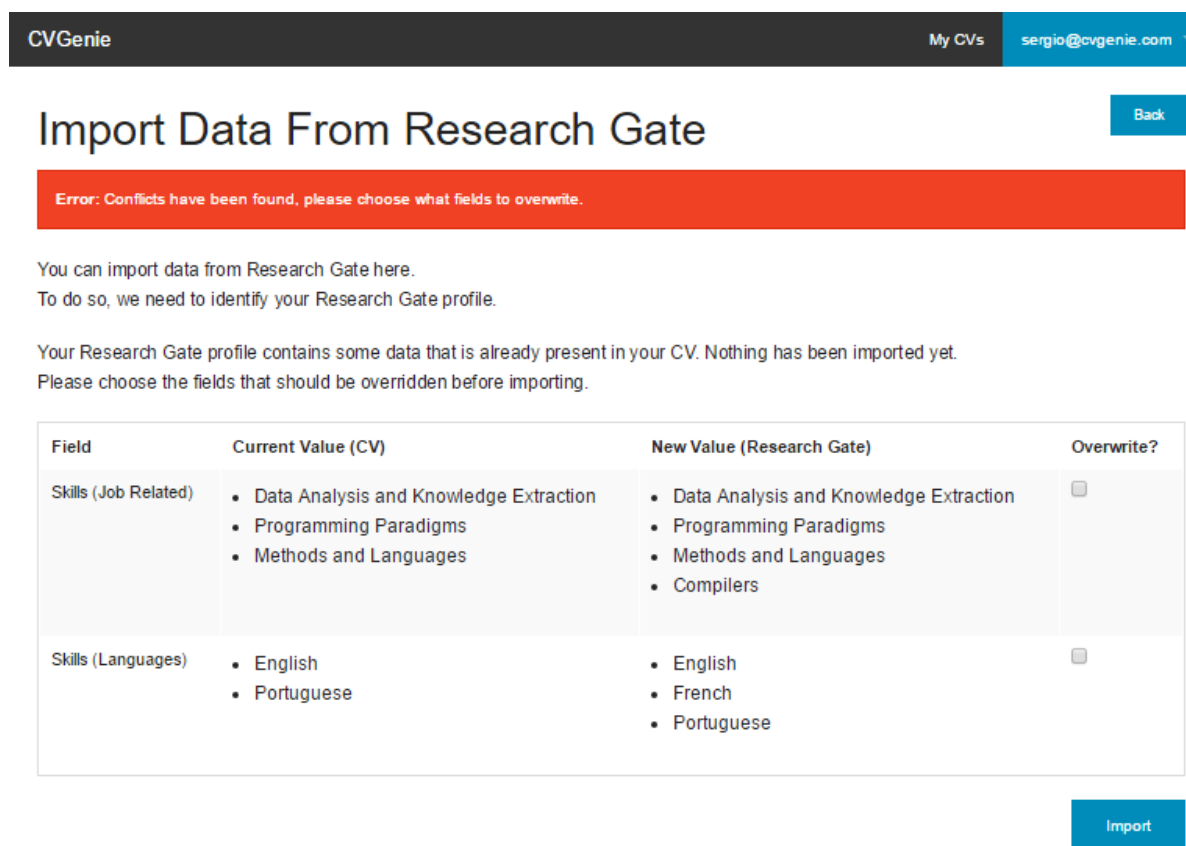When importing from LinkedIn and Research Gate, the existing information on a user's CV and the information that is being imported from those online platforms may cause conflicts, by having different values for the same sections of the CV (as detailed in section 4.10).

When that happens, a page is provided to the user that lists, for each section of the CV, the old value currently on the CV and the new value from the online platform, allowing the user to choose for each section if he wants to overwrite the value or keep the previous one.

Figure 26 displays an example of the conflict management page.

Figure 26.: CVGenie's conflict management page when importing from Research Gate

5.1.6 *Export CV in Various Formats*

At any time, the user can choose to export his CV in a variety of formats.

For that purpose, he simply needs to navigate to main CV page and press one of the buttons under the "Export" section. The following items specify the options that are available under that section, each representing a file format in which the user's CV can be exported.

- **XML Document** Allows the user to download the CV as an Europass XML document;
- **PDF Document** Allows the user to download the CV as an Europass PDF document (appendix A.5);
- **Word Document** Allows the user to download the CV as a Word document (appendix A.6);
- **Open Document** Allows the user to download the CV as Open Document (appendix A.7);
- **HTML Web Page** Allows the user to download the CV as a simple HTML page, with separate and interconnected sections (appendix A.8).

5.1.7 *Edit in an External Editor*

A user can also elect to, at any time, export his CV from CVGenie into the official Europass Online Editor, to enhance his CV. This is a lot more convenient than exporting the CV from CVGenie as an XML file and then uploading it to the Europass Editor manually.

Unfortunately the CV cannot then be saved to CVGenie automatically, although it can still be re-uploaded manually.

Appendices A.5 and A.9 present a CV in PDF format as it is on CVGenie and the result of selecting the "Edit on Europass Editor" option, respectively.

5.2 CASE STUDIES

5.2.1 *Case Study 1*

As the first case study, let us consider Professor Pedro Rangel Henriques's CV.

Professor Henriques has his professional information available in several locations: his institutional web page, as a CV document in the LaTeX format, and across several profiles in academic social networks, like Research Gate and Academia. In addition to these, he also has a profile in Google Scholar that is automatically updated by Google to include all of his publications.

To automatize the process of updating his publications on both his institutional web page and his CV, Professor Henriques decides to start using CVGenie, since it can automatically fetch his publications from his Google Scholar account, and export new versions of his CV both in web page and PDF formats.

To begin, he starts creating his own Europass CV, since he did not have one previously. After filling a few fields, he starts growing tired of doing it manually. The fields he has filled were his name, his

contacts and address, his date of birth and nationality, one entry of work experience and his job-related skills. At that point his CV is as displayed in fig. 27 with the corresponding XML Europass instance in listing 19.

Figure 27.: Professor Henriques's initial CV state

5.2. Case Studies

```xml
1  <?xml version='1.0' encoding='UTF-8'?>
2  <SkillsPassport xmlns="http://europass.cedefop.europa.eu/Europass" xmlns:xsi="http://www.w3.
       org/2001/XMLSchema-instance" xsi:schemaLocation="http://europass.cedefop.europa.eu/
       Europass http://europass.cedefop.europa.eu/xml/v3.3.0/EuropassSchema.xsd" locale="en">
3    <DocumentInfo>
4      <DocumentType>ECV</DocumentType>
5      <CreationDate>2015-10-19T10:15:21.235Z</CreationDate>
6      <LastUpdateDate>2015-10-23T09:34:29.614Z</LastUpdateDate>
7      <XSDVersion>V3.3</XSDVersion>
8      <Generator>EWA</Generator>
9      <Comment>Europass CV</Comment>
10     <EuropassLogo>true</EuropassLogo>
11   </DocumentInfo>
12   <PrintingPreferences>
13     <Document type="ECV">
14       <Field name="LearnerInfo.Identification.PersonName" show="true" order="FirstName
             Surname"/>
15       <Field name="LearnerInfo.Identification.Demographics.Birthdate" show="false" format="
             numeric/long"/>
16       <Field name="LearnerInfo.WorkExperience[0].Period" show="false" format="text/short"/>
17       <Field name="LearnerInfo.Education[0].Period" show="false" format="text/short"/>
18       <Field name="LearnerInfo" show="false" order="Identification Headline WorkExperience
             Education Skills Achievement ReferenceTo"/>
19     </Document>
20   </PrintingPreferences>
21   <LearnerInfo>
22     <Identification>
23       <PersonName>
24         <FirstName>Pedro Manuel Rangel Santos</FirstName>
25         <Surname>Henriques</Surname>
26       </PersonName>
27       <ContactInfo>
28         <Address>
29           <Contact>
30             <AddressLine>Rua Candido de Oliveira, num. 195 Hab-01</AddressLine>
31             <PostalCode>4715-012</PostalCode>
32             <Municipality>Braga</Municipality>
33             <Country>
34               <Code>PT</Code>
35               <Label>Portugal</Label>
36             </Country>
37           </Contact>
38         </Address>
39         <Email>
40           <Contact>pedrorangelhenriques@gmail.com</Contact>
41         </Email>
42         <TelephoneList>
43           <Telephone>
44             <Contact>+351 91 432 40 21</Contact>
45             <Use>
46               <Code>home</Code>
47             </Use>
48           </Telephone>
49         </TelephoneList>
```

```
50              <WebsiteList>
51                <Website>
52                  <Contact>http://www.di.uminho.pt/~prh</Contact>
53                </Website>
54              </WebsiteList>
55            </ContactInfo>
56          <Demographics>
57            <Birthdate year="1955" month="--06" day="---30"/>
58            <NationalityList>
59              <Nationality>
60                <Code>PT</Code>
61                <Label>Portuguese</Label>
62              </Nationality>
63            </NationalityList>
64          </Demographics>
65        </Identification>
66        <WorkExperienceList>
67          <WorkExperience>
68            <Period>
69              <From year="1996"/>
70              <Current>true</Current>
71            </Period>
72            <Position>
73              <Label>Associate Professor</Label>
74            </Position>
75            <Employer>
76              <Name>University of Minho</Name>
77              <ContactInfo>
78                <Address>
79                  <Contact>
80                    <Country>
81                      <Code>PT</Code>
82                      <Label>Portugal</Label>
83                    </Country>
84                  </Contact>
85                </Address>
86              </ContactInfo>
87            </Employer>
88          </WorkExperience>
89        </WorkExperienceList>
90        <Skills>
91          <JobRelated>
92            <Description>&lt;p>&lt;strong>Domain of specialization:&lt;/strong>&lt;/p>&lt;p>&lt;
                  strong>&lt;/strong>Computer science (Compilers)&lt;/p>&lt;p>&lt;strong>  Areas
                  of Research:&lt;/strong>&lt;/p>&lt;ul>&lt;li>Formal Languages Processing;&lt;/li
                  >&lt;li>Programming Paradigms;&lt;/li>&lt;li>Methods and Languages;&lt;/li>&lt;
                  li>Data Analysis and Knowledge Extraction.&lt;/li>&lt;/ul>&lt;p>&lt;strong>
                   Current  Research:&lt;/strong>&lt;/p>&lt;p>&lt;/p>&lt;ul>&lt;li>Compiling
                  methods;&lt;/li>&lt;li>Techniques and tools;&lt;/li>&lt;li>Program Comprehension
                  ;&lt;/li>&lt;li>Visualization and Animation;&lt;/li>&lt;li>Data Cleaning (Data
                  Mining preparation).&lt;/li>&lt;/ul></Description>
93          </JobRelated>
94        </Skills>
95      </LearnerInfo>
96 </SkillsPassport>
```

Listing 19: Professor Henriques's initial CV state in XML

Professor Henriques then wonders if CVGenie, in addition from importing his publications from Google Scholar, also imports from other sources, which would accelerate the initial process of filling the CV. Professor Henriques signs in to his CVGenie account, creates a new CV entry and verifies that, indeed, CVGenie supports other sources, and among them is Research Gate, for which he has an account.

He then imports his new CV into CVGenie as is. Once imported, he navigates to the "Import from Research Gate" page. It is requested that he enters the URL for his Research Gate profile. Using the search box in that page, he searches for his profile on Research Gate, and quickly finds it. Professor Henriques's Research Gate profile is displayed in figs. 28 and 29.

Figure 28.: Professor Henriques's Research Gate profile (1/2)

EDUCATION

Jan 2002– Jan 2012	**University of Minho** Computer Science · Habilitation Portugal · Guimarães
Jan 1988– Jan 1993	**University of Minho** Computer Science · Doctor of Philosophy Portugal · Guimarães
Jan 1981– Jan 1984	**University of Minho** Computer Science · Master of Science Portugal · Guimarães
Jan 1972– Jan 1977	**University of Porto** Electro-technical Engineering · Bachelor of Engineering Portugal · Porto

OTHER

| Languages | English, French, Portuguese |
| Scientific
Societies | Portuguese Association for Artificial Intelligence (APPIA): President of the Coordination Board for 2006-2007, 2008-2009, and 2010-2011. |

Figure 29.: Professor Henriques's Research Gate profile (2/2)

After confirming that his profile information is correct and making a few last tweaks, he submits the URL to his Research Gate profile into CVGenie. CVGenie retrieves his Research Gate profile information, and compares it with is current CV.

The fields that could be retrieved from the professor's profile were his name, his job-related skills, his work experience (though his researching and teaching experience), his education, his language skills and the scientific societies he belongs to.

Professor Henriques is pleasingly surprised to find that CVGenie has found conflicts but, instead of mindlessly replacing sections of his current CV, it offered him the opportunity to select what conflicting fields should be overwritten through the formulaires shown in figs. 30 and 31.

Figure 30.: Conflicting Information (1/2)

Figure 31.: Conflicting Information (2/2)

Professor Henriques finds that some information that he filled in the version of the CV he uploaded to CVGenie is incomplete when compared to his Research Gate profile; he selects the fields corresponding to that information, in order for them to be overwritten. One instance of this is the "Work Experience" section, for which the original CV value and the new value retrieved from Research Gate are respectively displayed in listings 20 and 21.

```xml
 1  <WorkExperienceList>
 2    <WorkExperience>
 3      <Period>
 4        <From year="1996"/>
 5        <Current>true</Current>
 6      </Period>
 7      <Position>
 8        <Label>Associate Professor</Label>
 9      </Position>
10      <Employer>
11        <Name>University of Minho</Name>
12        <ContactInfo>
13          <Address>
14            <Contact>
15              <Country>
16                <Code>PT</Code>
17                <Label>Portugal</Label>
18              </Country>
19            </Contact>
20          </Address>
21        </ContactInfo>
22      </Employer>
23    </WorkExperience>
24  </WorkExperienceList>
```

Listing 20: Current CV work experience in XML

```xml
 1  <WorkExperienceList>
 2    <WorkExperience>
 3      <Period>
 4        <From year="1994" month="--01"/>
 5        <Current>true</Current>
 6      </Period>
 7      <Position>
 8        <Label>Researching (Group Leader)</Label>
 9      </Position>
10      <Employer>
11        <Name>University of Minho, Departamento de Inform tica (DI), Language Processing
            Group (gEPL)</Name>
12        <ContactInfo>
13          <Address>
14            <Contact>
15              <Municipality>Guimar es</Municipality>
16              <Country>
17                <Label>Portugal</Label>
18              </Country>
19            </Contact>
```

```
20          </Address>
21        </ContactInfo>
22      </Employer>
23    </WorkExperience>
24    <WorkExperience>
25      <Period>
26        <From year="1996" month="--01"/>
27        <Current>true</Current>
28      </Period>
29      <Position>
30        <Label>Teaching (Associate Professor)</Label>
31      </Position>
32      <Employer>
33        <Name>University of Minho, Departamento de Inform tica (DI)</Name>
34        <ContactInfo>
35          <Address>
36            <Contact>
37              <Municipality>Guimar es</Municipality>
38              <Country>
39                <Label>Portugal</Label>
40              </Country>
41            </Contact>
42          </Address>
43        </ContactInfo>
44      </Employer>
45    </WorkExperience>
46    <WorkExperience>
47      <Period>
48        <From year="1993" month="--01"/>
49        <To year="1996" month="--01"/>
50        <Current>false</Current>
51      </Period>
52      <Position>
53        <Label>Teaching (Auxiliary Professor)</Label>
54      </Position>
55      <Employer>
56        <Name>University of Minho, Departamento de Inform tica (DI)</Name>
57        <ContactInfo>
58          <Address>
59            <Contact>
60              <Municipality>Guimar es</Municipality>
61              <Country>
62                <Label>Portugal</Label>
63              </Country>
64            </Contact>
65          </Address>
66        </ContactInfo>
67      </Employer>
68    </WorkExperience>
69    <WorkExperience>
70      <Period>
71        <From year="1981" month="--01"/>
72        <To year="1993" month="--01"/>
73        <Current>false</Current>
74      </Period>
```

```
75      <Position>
76        <Label>Teaching (Assistant Lecturer)</Label>
77      </Position>
78      <Employer>
79        <Name>University of Minho, Departamento de Inform tica (DI)</Name>
80        <ContactInfo>
81          <Address>
82            <Contact>
83              <Municipality>Guimar es</Municipality>
84              <Country>
85                <Label>Portugal</Label>
86              </Country>
87            </Contact>
88          </Address>
89        </ContactInfo>
90      </Employer>
91    </WorkExperience>
92 </WorkExperienceList>
```

Listing 21: New CV work experience retrieved from Research Gate in XML

Professor Henriques also notes that the "Skills" section of his Research Gate profile is a bit incomplete: it should rather be sectioned and detailed like the one in the Professor's current CV. The professor already filled the "Skills" section in this detailed fashion, and therefore selects to not overwrite the "Skills" field and keep the current information. Figures 32 and 33 represent the state of his CV after importing the information from his Research Gate profile. Listing 22 offers the corresponding Europass XML representation.

PERSONAL INFORMATION	Pedro Manuel Rangel Santos Henriques

Rua Cândido de Oliveira, num. 195 Hab-01, 4715-012 Braga (Portugal)

☎ +351 91 432 40 21

✉ pedrorangelhenriques@gmail.com

🌐 http://www.di.uminho.pt/~prh

Date of birth 30/06/1955 | Nationality Portuguese

WORK EXPERIENCE

Jan 1994–Present	Researching (Group Leader)
	University of Minho, Departamento de Informática (DI), Language Processing Group (gEPL), Guimarães (Portugal)

Jan 1996–Present	Teaching (Associate Professor)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

Jan 1993–Jan 1996	Teaching (Auxiliary Professor)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

Jan 1981–Jan 1993	Teaching (Assistant Lecturer)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

EDUCATION AND TRAINING

Jan 2002–Jan 2012	Computer Science, Habilitation
	University of Minho, Guimarães (Portugal)

Jan 1988–Jan 1993	Computer Science, Doctor of Philosophy
	University of Minho, Guimarães (Portugal)

Jan 1981–Jan 1984	Computer Science, Master of Science
	University of Minho, Guimarães (Portugal)

Jan 1972–Jan 1977	Electro-technical Engineering, Bachelor of Engineering
	University of Porto, Porto (Portugal)

PERSONAL SKILLS

Other language(s)

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	
English				
French				

Figure 32.: Professor Henriques's CV after importing from Research Gate (1/2)

europass   Curriculum vitae   Pedro Manuel Rangel Santos Henriques

Portuguese

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Job-related skills   **Domain of specialization:**

Computer science (Compilers)

Areas of Research:

- Formal Languages Processing;
- Programming Paradigms;
- Methods and Languages;
- Data Analysis and Knowledge Extraction.

Current Research:

- Compiling methods;
- Techniques and tools;
- Program Comprehension;
- Visualization and Animation;
- Data Cleaning (Data Mining preparation).

ADDITIONAL INFORMATION _____

Scientific societies   Portuguese Association for Artificial Intelligence (APPIA): President of the Coordination Board for
                           2006-2007, 2008-2009, and 2010-2011.

Figure 33.: Professor Henriques's CV after importing from Research Gate (2/2)

5.2. Case Studies

```xml
1  <?xml version='1.0' encoding='UTF-8'?>
2  <SkillsPassport xmlns="http://europass.cedefop.europa.eu/Europass" xmlns:xsi="http://www.w3.
       org/2001/XMLSchema-instance" xsi:schemaLocation="http://europass.cedefop.europa.eu/
       Europass http://europass.cedefop.europa.eu/xml/v3.3.0/EuropassSchema.xsd" locale="en">
3    <DocumentInfo>
4      <DocumentType>ECV</DocumentType>
5      <Bundle/>
6      <CreationDate>2015-10-19T10:15:21.235Z</CreationDate>
7      <LastUpdateDate>2015-10-23T11:24:35.678Z</LastUpdateDate>
8      <XSDVersion>V3.3</XSDVersion>
9      <Generator>EWA</Generator>
10     <Comment>Europass CV</Comment>
11     <EuropassLogo>true</EuropassLogo>
12   </DocumentInfo>
13   <PrintingPreferences>
14     <Document type="ECV">
15       <Field name="LearnerInfo.Identification.PersonName" show="true" order="FirstName
             Surname"/>
16       <Field name="LearnerInfo.Identification.Demographics.Birthdate" show="false" format="
             numeric/long"/>
17       <Field name="LearnerInfo.WorkExperience[0].Period" show="false" format="text/short"/>
18       <Field name="LearnerInfo.WorkExperience[1].Period" show="true" format="text/short"/>
19       <Field name="LearnerInfo.WorkExperience[2].Period" show="true" format="text/short"/>
20       <Field name="LearnerInfo.WorkExperience[3].Period" show="true" format="text/short"/>
21       <Field name="LearnerInfo.Education[0].Period" show="false" format="text/short"/>
22       <Field name="LearnerInfo.Education[1].Period" show="true" format="text/short"/>
23       <Field name="LearnerInfo.Education[2].Period" show="true" format="text/short"/>
24       <Field name="LearnerInfo.Education[3].Period" show="true" format="text/short"/>
25       <Field name="LearnerInfo" show="false" order="Identification Headline WorkExperience
             Education Skills Achievement ReferenceTo"/>
26     </Document>
27   </PrintingPreferences>
28   <LearnerInfo>
29     <Identification>
30       <PersonName>
31         <FirstName>Pedro Manuel Rangel Santos</FirstName>
32         <Surname>Henriques</Surname>
33       </PersonName>
34       <ContactInfo>
35         <Address>
36           <Contact>
37             <AddressLine>Rua Candido de Oliveira, num. 195 Hab-01</AddressLine>
38             <PostalCode>4715-012</PostalCode>
39             <Municipality>Braga</Municipality>
40             <Country>
41               <Code>PT</Code>
42               <Label>Portugal</Label>
43             </Country>
44           </Contact>
45         </Address>
46         <Email>
47           <Contact>pedrorangelhenriques@gmail.com</Contact>
48         </Email>
49         <TelephoneList>
```

```
50            <Telephone>
51              <Contact>+351 91 432 40 21</Contact>
52              <Use>
53                <Code>home</Code>
54              </Use>
55            </Telephone>
56          </TelephoneList>
57          <WebsiteList>
58            <Website>
59              <Contact>http://www.di.uminho.pt/~prh</Contact>
60            </Website>
61          </WebsiteList>
62        </ContactInfo>
63        <Demographics>
64          <Birthdate year="1955" month="--06" day="---30"/>
65          <NationalityList>
66            <Nationality>
67              <Code>PT</Code>
68              <Label>Portuguese</Label>
69            </Nationality>
70          </NationalityList>
71        </Demographics>
72      </Identification>
73      <WorkExperienceList>
74        <WorkExperience>
75          <Period>
76            <From year="1994" month="--01"/>
77            <Current>true</Current>
78          </Period>
79          <Position>
80            <Label>Researching (Group Leader)</Label>
81          </Position>
82          <Employer>
83            <Name>University of Minho, Departamento de Inform tica (DI), Language Processing
                  Group (gEPL)</Name>
84            <ContactInfo>
85              <Address>
86                <Contact>
87                  <Municipality>Guimar es</Municipality>
88                  <Country>
89                    <Label>Portugal</Label>
90                  </Country>
91                </Contact>
92              </Address>
93            </ContactInfo>
94          </Employer>
95        </WorkExperience>
96        <WorkExperience>
97          <Period>
98            <From year="1996" month="--01"/>
99            <Current>true</Current>
100         </Period>
101         <Position>
102           <Label>Teaching (Associate Professor)</Label>
103         </Position>
```

```
104            <Employer>
105              <Name>University of Minho, Departamento de Inform tica (DI)</Name>
106              <ContactInfo>
107                <Address>
108                  <Contact>
109                    <Municipality>Guimar es</Municipality>
110                    <Country>
111                      <Label>Portugal</Label>
112                    </Country>
113                  </Contact>
114                </Address>
115              </ContactInfo>
116            </Employer>
117          </WorkExperience>
118          <WorkExperience>
119            <Period>
120              <From year="1993" month="--01"/>
121              <To year="1996" month="--01"/>
122              <Current>false</Current>
123            </Period>
124            <Position>
125              <Label>Teaching (Auxiliary Professor)</Label>
126            </Position>
127            <Employer>
128              <Name>University of Minho, Departamento de Inform tica (DI)</Name>
129              <ContactInfo>
130                <Address>
131                  <Contact>
132                    <Municipality>Guimar es</Municipality>
133                    <Country>
134                      <Label>Portugal</Label>
135                    </Country>
136                  </Contact>
137                </Address>
138              </ContactInfo>
139            </Employer>
140          </WorkExperience>
141          <WorkExperience>
142            <Period>
143              <From year="1981" month="--01"/>
144              <To year="1993" month="--01"/>
145              <Current>false</Current>
146            </Period>
147            <Position>
148              <Label>Teaching (Assistant Lecturer)</Label>
149            </Position>
150            <Employer>
151              <Name>University of Minho, Departamento de Inform tica (DI)</Name>
152              <ContactInfo>
153                <Address>
154                  <Contact>
155                    <Municipality>Guimar es</Municipality>
156                    <Country>
157                      <Label>Portugal</Label>
158                    </Country>
```

```
159                  </Contact>
160                </Address>
161              </ContactInfo>
162            </Employer>
163          </WorkExperience>
164        </WorkExperienceList>
165        <EducationList>
166          <Education>
167            <Period>
168              <From year="2002" month="--01"/>
169              <To year="2012" month="--01"/>
170              <Current>false</Current>
171            </Period>
172            <Title>Computer Science, Habilitation</Title>
173            <Organisation>
174              <Name>University of Minho</Name>
175              <ContactInfo>
176                <Address>
177                  <Contact>
178                    <Municipality>Guimar es</Municipality>
179                    <Country>
180                      <Label>Portugal</Label>
181                    </Country>
182                  </Contact>
183                </Address>
184              </ContactInfo>
185            </Organisation>
186          </Education>
187          <Education>
188            <Period>
189              <From year="1988" month="--01"/>
190              <To year="1993" month="--01"/>
191              <Current>false</Current>
192            </Period>
193            <Title>Computer Science, Doctor of Philosophy</Title>
194            <Organisation>
195              <Name>University of Minho</Name>
196              <ContactInfo>
197                <Address>
198                  <Contact>
199                    <Municipality>Guimar es</Municipality>
200                    <Country>
201                      <Label>Portugal</Label>
202                    </Country>
203                  </Contact>
204                </Address>
205              </ContactInfo>
206            </Organisation>
207          </Education>
208          <Education>
209            <Period>
210              <From year="1981" month="--01"/>
211              <To year="1984" month="--01"/>
212              <Current>false</Current>
213            </Period>
```

```
214            <Title>Computer Science, Master of Science</Title>
215            <Organisation>
216              <Name>University of Minho</Name>
217              <ContactInfo>
218                <Address>
219                  <Contact>
220                    <Municipality>Guimar es</Municipality>
221                    <Country>
222                      <Label>Portugal</Label>
223                    </Country>
224                  </Contact>
225                </Address>
226              </ContactInfo>
227            </Organisation>
228          </Education>
229          <Education>
230            <Period>
231              <From year="1972" month="--01"/>
232              <To year="1977" month="--01"/>
233              <Current>false</Current>
234            </Period>
235            <Title>Electro-technical Engineering, Bachelor of Engineering</Title>
236            <Organisation>
237              <Name>University of Porto</Name>
238              <ContactInfo>
239                <Address>
240                  <Contact>
241                    <Municipality>Porto</Municipality>
242                    <Country>
243                      <Label>Portugal</Label>
244                    </Country>
245                  </Contact>
246                </Address>
247              </ContactInfo>
248            </Organisation>
249          </Education>
250        </EducationList>
251        <Skills>
252          <Linguistic>
253            <ForeignLanguageList>
254              <ForeignLanguage>
255                <Description>
256                  <Label>English</Label>
257                </Description>
258              </ForeignLanguage>
259              <ForeignLanguage>
260                <Description>
261                  <Label>French</Label>
262                </Description>
263              </ForeignLanguage>
264              <ForeignLanguage>
265                <Description>
266                  <Label>Portuguese</Label>
267                </Description>
268              </ForeignLanguage>
```
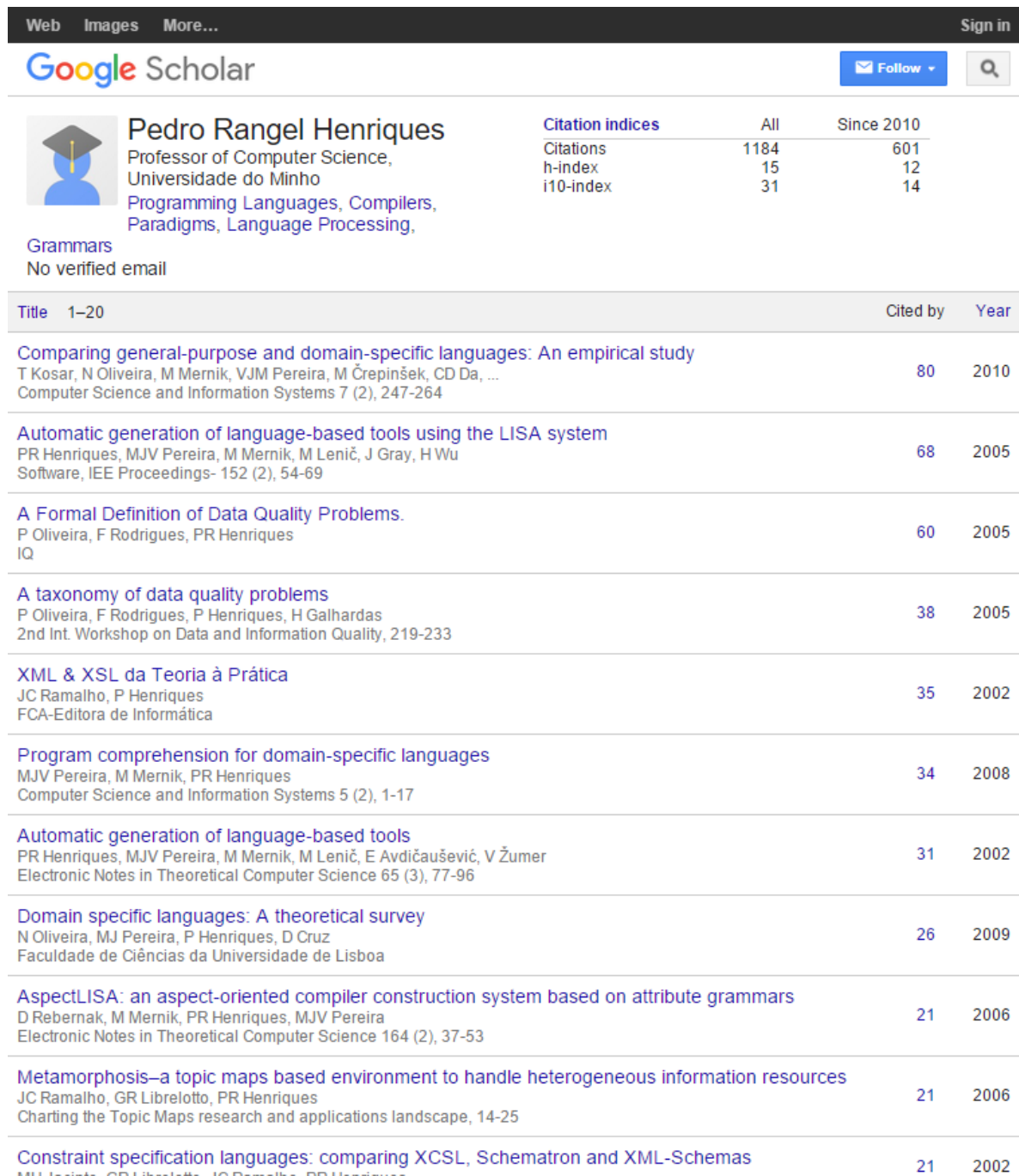
```
269          </ForeignLanguageList>
270        </Linguistic>
271        <JobRelated>
272          <Description>&lt;p>&lt;strong>Domain of specialization:&lt;/strong>&lt;/p>&lt;p>&lt;
                 strong>&lt;/strong>Computer science (Compilers)&lt;/p>&lt;p>&lt;strong> Areas
                 of Research:&lt;/strong>&lt;/p>&lt;ul>&lt;li>Formal Languages Processing;&lt;/li
                 >&lt;li>Programming Paradigms;&lt;/li>&lt;li>Methods and Languages;&lt;/li>&lt;
                 li>Data Analysis and Knowledge Extraction.&lt;/li>&lt;/ul>&lt;p>&lt;strong>
                  Current  Research:&lt;/strong>&lt;/p>&lt;p>&lt;/p>&lt;ul>&lt;li>Compiling
                 methods;&lt;/li>&lt;li>Techniques and tools;&lt;/li>&lt;li>Program Comprehension
                 ;&lt;/li>&lt;li>Visualization and Animation;&lt;/li>&lt;li>Data Cleaning (Data
                 Mining preparation).&lt;/li>&lt;/ul></Description>
273        </JobRelated>
274      </Skills>
275      <AchievementList>
276        <Achievement>
277          <Title>
278            <Label>Scientific societies</Label>
279          </Title>
280          <Description>Portuguese Association for Artificial Intelligence (APPIA): President
                 of the Coordination Board for 2006-2007, 2008-2009, and 2010-2011.</Description>
281        </Achievement>
282      </AchievementList>
283    </LearnerInfo>
284 </SkillsPassport>
```

Listing 22: Professor Henriques's CV after importing from Research Gate in XML

Professor Henriques then finally navigates to the "Import from Google Scholar" page to import his publications. He is similarly prompted to enter his Google Scholar profile URL; he opens it and supplies the URL. A sample of his profile can be seen in fig. 34.

Figure 34.: Professor Henriques's Google Scholar profile

Professor Henriques instructs CVGenie to import from his Google Scholar profile. After a short while, several hundreds of publications that Google Scholar has indexed that pertain to Professor Henriques are cited in his CV.

These publications are sorted by type, and are held in 4 different categories: books, chapters in books, conference papers and journal articles.

Figures 35 and 36 illustrate the resulting Europass CV in the PDF format. Listing 23 presents its respective Europass XML representation.

europass   Curriculum vitae   Pedro Manuel Rangel Santos Henriques

Portuguese

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Job-related skills

Domain of specialization:

Computer science (Compilers)

Areas of Research:

- Formal Languages Processing;
- Programming Paradigms;
- Methods and Languages;
- Data Analysis and Knowledge Extraction.

Current Research:

- Compiling methods;
- Techniques and tools;
- Program Comprehension;
- Visualization and Animation;
- Data Cleaning (Data Mining preparation).

ADDITIONAL INFORMATION ─────────────────────────────────

Scientific societies

Portuguese Association for Artificial Intelligence (APPIA): President of the Coordination Board for 2006-2007, 2008-2009, and 2010-2011.

Publications (Books)

- Pereira, MJ, Berón, M, Perez, N & Henriques, P 2015, *Comprensión de sistemas paralelos*, RedUNCI-Universidade Nacional de Salta
- Berón, M, Bernardis, H, Miranda, E, Riesco, D, Pereira, MJ & Henriques, P 2015, *WSDLUD: a metric to measure the understanding degree of WSDL descriptions*, Unknown Publisher
- Pereira, N, Pereira, MJ & Henriques, P 2014, *Comment-based concept location over system dependency graphs*, Unknown Publisher
- Azcurra, J, Berón, M, Montejano, G, Farnese, A, Henriques, P & Pereira, MJ 2014, *AId: uma Ferramenta para análise de identificadores de programas Java*, Unknown Publisher
- Carvalho, N, Almeida, JJ, Pereira, MJ & Henriques, P 2014, *Conclave: writing programs to understand programs*, Unknown Publisher
- Fonseca, J, Pereira, MJ & Henriques, P 2014, *Converting ontologies into DSLs*, Maria João Varanda Pereira, José Paulo Leal e Alberto Simões
- Leal, JP, Rocha, R, Simões, A, Favre, J, Zhou, N, Santos, J, Miranda, E, Berón, M, Montejano, G, Pereira, MJV, Henriques, PR, Fourtounis, G, Papaspyrou, NS, Chodarev, S, Pietriková, E, Kollár, J, Costa, T, Queirós, R, Campos, J, Nosál, M, Porubän, J, Guinovart, XG, Fonte, D, Cruz, D, Gançarski, AL, Oliveira, N, Halupka, I, Baciková, M, Lakatos, D, Azevedo, BM, Almeida, JJ, Santos, AC, Cardoso, JM, Diniz, PC, Ferreira, DR, Orosz, G, Laki, LJ, Novák, A, Siklósi, B, Medeiros, H, Batista, F, Moniz, H, Trancoso, I, Nunes, L, Freitas, T, Baptista, J & Mamede, N 2013, *2nd Symposium on Languages, Applications and Technologies*, Unknown Publisher
- Pereira, MJ, Oliveira, N, Cruz, D & Henriques, P 2013, *Choosing grammars to support language processing courses*, Unknown Publisher
- Miranda, E, Berón, M, Montejano, G, Pereira, MJ & Henriques, P 2013, *NESSy: a new evaluator for software development tools*, Unknown Publisher
- Azcurra, J, Berón, M, Henriques, P & Pereira, MJ 2012, *Análisis de información informal para facilitar la comprensión de programas*, RedUNCI
- Oliveira, N, Pereira, MJ, Gancarski, A & Henriques, P 2012, *Learning spaces for knowledge generation*, Alberto Simões and Ricardo Queirós and Daniela da Cruz

Figure 35.: Professor Henriques's CV after importing from Google Scholar (1/2)

5.2. Case Studies

e Processamento, Unknown Publisher

- Berón13, MM, Henriques, PR, Varanda, MJ & Uzal, R , *Program inspection to inter-connect the operational and behavioral views for program comprehension*, Unknown Publisher
- Areias, S, Cruz, DD, Henriques, PR & Pinto, JS , *GammaPolarSlicer A Contract-based Tool to help on Reuse*, Unknown Publisher

Publications (Chapters in Books)

- Martini, R, Guimarães, M, Librelotto, G & Henriques, P 2015, Storing Archival Emigration Documents to Create Virtual Exhibition Rooms in *New Contributions in Information Systems and Technologies*, Springer International Publishing, pp. 403-409
- Carvalho, NR, Almeida, JJ, Henriques, PR & Pereira, MJV 2014, Conclave: Ontology-driven measurement of semantic relatedness between source code elements and problem domain concepts in *Computational Science and Its Applications–ICCSA 2014*, Springer International Publishing, pp. 116-131
- Cruz, DD, Henriques, PR & Pinto, JS 2010, Contract-based slicing in *Leveraging Applications of Formal Methods, Verification, and Validation*, Springer Berlin Heidelberg, pp. 106-120
- Lopes, S, Henriques, P & Dias, P 2008, Making concept maps available on the web to the students in *Computers and Education*, Springer London, pp. 179-188
- Librelotto, GR, Azevedo, RP, Ramalho, JC & Henriques, PR 2008, Comparing topic maps constraint specification languages in *Scaling Topic Maps*, Springer Berlin Heidelberg, pp. 86-97
- Ramalho, JC, Librelotto, GR & Henriques, PR 2006, Metamorphosis–a topic maps based environment to handle heterogeneous information resources in *Charting the Topic Maps research and applications landscape*, Springer Berlin Heidelberg, pp. 14-25
- Rodrigues, MF, Ramos, C & Henriques, PR 1998, Extracting knowledge patterns from ticket data in *Discovey Science*, Springer Berlin Heidelberg, pp. 435-436
- Mendes, JA & Henriques, PR 1991, Heuristic parsing and search space pruning in *EPIA 91*, Springer Berlin Heidelberg, pp. 76-91
- Henriques, PR 1989, A semantic evaluator generating system in Prolog in *Programming Languages Implementation and Logic Programming*, Springer Berlin Heidelberg, pp. 201-218

Publications (Journal Articles)

- Schwab, A, Werbel, JD, Hofmann, H & Henriques, PL 2015, 'Managerial Gender Diversity and Firm Performance An Integration of Different Theoretical Perspectives', *Group & Organization Management*, p. 1059601115588641
- Curado, C, Henriques, PL & Ribeiro, S 2015, 'Voluntary or mandatory enrollment in training and the motivation to transfer training', *International Journal of Training and Development*, vol. 19, no. 2, pp. 98-109
- Carvalho, NR, Almeida, JJ, Henriques, PR & Varanda, MJ 2015, 'From source code identifiers to natural language terms', *Journal of Systems and Software*, vol. 100, pp. 117-128
- Sanchez, A, Oliveira, N, Barbosa, LS & Henriques, P 2015, 'A perspective on architectural re-engineering', *Science of Computer Programming*, vol. 98, pp. 764-784
- Carvalho, P, Oliveira, N & Henriques, PR 2014, 'Unfuzzying Fuzzy Parsing', *3rd Symposium on Languages, Applications and Technologies}*, vol. 38, pp. 101-108
- Henriques, PL, Werbel, J & Curado, C 2014, 'Leader's Human Capital as an Alternative for Relational Leadership', *Journal of Modern Accounting and Auditing*, vol. 10, no. 5, pp. 560-574
- Pereira, MJ, Oliveira, N, Cruz, D & Henriques, P 2014, 'An effective way to teach language processing courses', *Innovative Teaching Strategies and New Learning Paradigms in Computer Programming*, p. 131
- Soares, M, Garanito, L, Henriques, P, Rodrigues, F & Freire, F 2014, 'Upper airways sarcoidosis', *Annales françaises d'Oto-rhino-laryngologie et de Pathologie Cervico-faciale*, vol. 131, no. 4, p. 130
- Fonte, D, Boas, IV, Oliveira, N, Cruz, D, Gançarsky, AL & Henriques, PR 2014, 'Partial correctness and continuous integration in computer supported education', *CSEDU 2014-Proceedings of the 6th International Conference on Computer Supported Education*, vol. 2, pp. 205-212
- Martins, VT, Fonte, D, Henriques, PR & Cruz, D 2014, 'Plagiarism detection: A tool survey and comparison', *InProceedings {martins_et_al: OASIcs*, p. 4566
- Martins, R, Rocha, J & Henriques, P 2013, 'Segurança dos Web Services no Comércio Eletrônico Móvel', *Universidade do Minho Departamento de Informática–Campus de Gualtar-Braga>. Acesso*

Figure 36.: Professor Henriques's CV after importing from Google Scholar (2/2)

```xml
1  <AchievementList>
2    <Achievement>
3      <Title>
4        <Label>Scientific societies</Label>
5      </Title>
6      <Description>Portuguese Association for Artificial Intelligence (APPIA): President of
            the Coordination Board for 2006-2007, 2008-2009, and 2010-2011.</Description>
7    </Achievement>
8    <Achievement>
9      <Title>
10       <Code/>
11       <Label>Publications (Books)</Label>
12     </Title>
13     <Description>
14       &lt;ul&gt;&lt;li&gt; Pereira, MJ, Ber n , M, Perez, N &amp; Henriques, P 2015, &lt;i&
            gt;Comprensi n de sistemas paralelos&lt;/i&gt;, RedUNCI-Universidade Nacional de
            Salta &lt;/li&gt;&lt;li&gt; Ber n , M, Bernardis, H, Miranda, E, Riesco, D,
            Pereira, MJ &amp; Henriques, P 2015, &lt;i&gt;WSDLUD: a metric to measure the
            understanding degree of WSDL descriptions&lt;/i&gt;, Unknown Publisher &lt;/li&gt
            ;&lt;li&gt;
15       <!-- (...) -->
16     <Description>
17   </Achievement>
18   <Achievement>
19     <Title>
20       <Code/>
21       <Label>Publications (Chapters in Books)</Label>
22     </Title>
23     <Description>
24       &lt;ul&gt;&lt;li&gt; Martini, R, Guimar es , M, Librelotto, G &amp; Henriques, P 2015,
             Storing Archival Emigration Documents to Create Virtual Exhibition Rooms in &lt;i
            &gt;New Contributions in Information Systems and Technologies&lt;/i&gt;, Springer
            International Publishing, pp. 403-409 &lt;/li&gt;&lt;li&gt;
25       <!-- (...) -->
26     </Description>
27   </Achievement>
28   <Achievement>
29     <Title>
30       <Code/>
31       <Label>Publications (Journal Articles)</Label>
32     </Title>
33     <Description>
34       &lt;ul&gt;&lt;li&gt; Schwab, A, Werbel, JD, Hofmann, H &amp; Henriques, PL 2015, '
            Managerial Gender Diversity and Firm Performance An Integration of Different
            Theoretical Perspectives', &lt;i&gt;
35       <!-- (...) -->
36     <Description>
37   </Achievement>
38   <Achievement>
39     <Title>
40       <Code/>
41       <Label>Publications (Conference Papers)</Label>
42     </Title>
43     <Description>
```

```
44          &lt;ul&gt;&lt;li&gt; Tavares, PC, Gomes, EF &amp; Henriques, PR 2015, 'Animation and
                automatic evaluation in supporting the teaching of programming', &lt;i&gt;
                Information Systems and Technologies (CISTI), 2015 10th Iberian Conference on&lt;/
                i&gt;, IEEE, pp. 1-4 &lt;/li&gt;&lt;li&gt;
45          <!-- (...) -->
46        </Description>
47      </Achievement>
48  </AchievementList>
```

Listing 23: Professor Henriques's CV after importing from Google Scholar in XML

After analyzing the result, Professor Henriques realizes that the number of publications that were imported is really quite large, and that not every publication is perfectly cited.

With that in mind, he would like to cull and correct those publications; to do so, he selects the "Edit in External Editor" option in CVGenie, which opens the Europass Online Editor supplied with a copy of his current CV.

A warning is displayed indicating that any changes made in the external editor will not be propagated to CVGenie's version of the CV. Having read this warning, the professor recognizes that, after editing his CV in the external editor, he must download the result in the Europass XML format and re-upload it into CVGenie. He does so, and then considers CVGenie's "Export" section, in order to understand which formats are available for download, and how doing so can be accomplished.

He downloads his CV as PDF to replace his previous CV. He also downloads it as a web page, and studies the result. The resulting web page has a pleasant appearance and a handy sectioning system; however, it does not feature all of the information his current web page features, since not all of it is included in his CV. He decides to not replace his current web page with the new one, but makes a mental note about doing so in the future, after transcribing the information. Figures 37 to 42 present the resulting Europass CV in PDF and HTML formats.

europass Curriculum vitae

PERSONAL INFORMATION	Pedro Manuel Rangel Santos Henriques

Rua Cândido de Oliveira, num. 195 Hab-01, 4715-012 Braga (Portugal)

+351 91 432 40 21

pedrorangelhenriques@gmail.com

http://www.di.uminho.pt/~prh

Date of birth 30/06/1955 | Nationality Portuguese

WORK EXPERIENCE

Jan 1994–Present	Researching (Group Leader)
	University of Minho, Departamento de Informática (DI), Language Processing Group (gEPL), Guimarães (Portugal)

Jan 1996–Present	Teaching (Associate Professor)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

Jan 1993–Jan 1996	Teaching (Auxiliary Professor)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

Jan 1981–Jan 1993	Teaching (Assistant Lecturer)
	University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

EDUCATION AND TRAINING

Jan 2002–Jan 2012	Computer Science, Habilitation
	University of Minho, Guimarães (Portugal)

Jan 1988–Jan 1993	Computer Science, Doctor of Philosophy
	University of Minho, Guimarães (Portugal)

Jan 1981–Jan 1984	Computer Science, Master of Science
	University of Minho, Guimarães (Portugal)

Jan 1972–Jan 1977	Electro-technical Engineering, Bachelor of Engineering
	University of Porto, Porto (Portugal)

PERSONAL SKILLS

Other language(s)

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	

English

French

Figure 37.: Professor Henriques's final CV in the PDF format (1/2)

europass   Curriculum vitae   Pedro Manuel Rangel Santos Henriques

Portuguese

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Job-related skills   **Domain of specialization:**

Computer science (Compilers)

Areas of Research:

- Formal Languages Processing;
- Programming Paradigms;
- Methods and Languages;
- Data Analysis and Knowledge Extraction.

Current Research:

- Compiling methods;
- Techniques and tools;
- Program Comprehension;
- Visualization and Animation;
- Data Cleaning (Data Mining preparation).

ADDITIONAL INFORMATION   ▬

Scientific societies   Portuguese Association for Artificial Intelligence (APPIA): President of the Coordination Board for 2006-2007, 2008-2009, and 2010-2011.

Publications (Books)
- Pereira, MJ, Berón, M, Perez, N & Henriques, P 2015, *Comprensión de sistemas paralelos*, RedUNCI-Universidade Nacional de Salta
- Fonseca, J, Pereira, MJ & Henriques, P 2014, *Converting ontologies into DSLs*, Maria João Varanda Pereira, José Paulo Leal e Alberto Simões
- Azcurra, J, Berón, M, Henriques, P & Pereira, MJ 2012, *Análisis de información informal para facilitar la comprensión de programas*, RedUNCI
- Oliveira, N, Pereira, MJ, Gancarski, A & Henriques, P 2012, *Learning spaces for knowledge generation*, Alberto Simões and Ricardo Queirós and Daniela da Cruz
- Pereira, MJ, Berón, M, Cruz, D, Oliveira, N & Henriques, P 2012, *Problem domain oriented approach for program comprehension*, Alberto Simões and Ricardo Queirós and Daniela da Cruz
- Carvalho, N, Almeida, JJ, Pereira, MJ & Henriques, P 2012, *Probabilistic synset based concept location*, Alberto Simões and Ricardo Queirós and Daniela da Cruz
- Bernardis, H, Berón, M, Riesco, D, Henriques, P & Pereira, MJ 2011, *Extracción de información dinámica en programación orientada a objetos (Java)*, Red de Universidades Nacionales con Carreras de Informática (RedUNCI)
- Miranda, E, Berón, M, Montejano, G, Riesco, D, Henriques, P & Pereira, MJ 2011, *Visualización de software: conceptos, métodos y técnicas para facilitar la comprensión de programas*, Red de Universidades Nacionales con Carreras de Informática (RedUNCI)
- Kadre, I, Berón, M, Salgado, C, Peralta, M, Henriques, P & Pereira, MJ 2011, *Construcción de representaciones del dominio del programa para facilitar la interconexión de dominios*, Red de Universidades Nacionales con Carreras de Informática (RedUNCI)
- Albanes, JL, Berón, M, Henriques, P & Pereira, MJ 2011, *Estrategias para relacionar el dominio del problema con el dominio del programa para la comprensión de programas*, Red de Universidades Nacionales con Carreras de Informática (RedUNCI)
- Aristiaran, M, Berón, M, Henriques, P & Pereira, MJ 2011, *Clasificaciones: un mecanismo de herencia múltiple para la construcción de modelos fáciles de comprender y mantener*, Red de Universidades Nacionales con Carreras de Informática (RedUNCI)

Figure 38.: Professor Henriques's final CV in the PDF format (2/2)

Figure 39.: "Basic Information" section of Professor Henriques's final CV in the HTML format

Figure 40.: "Work Experiences" section of Professor Henriques's final CV in the HTML format

Figure 41.: "Skills" section of Professor Henriques's final CV in the HTML format

Figure 42.: "Publications" section of Professor Henriques's final CV in the HTML format
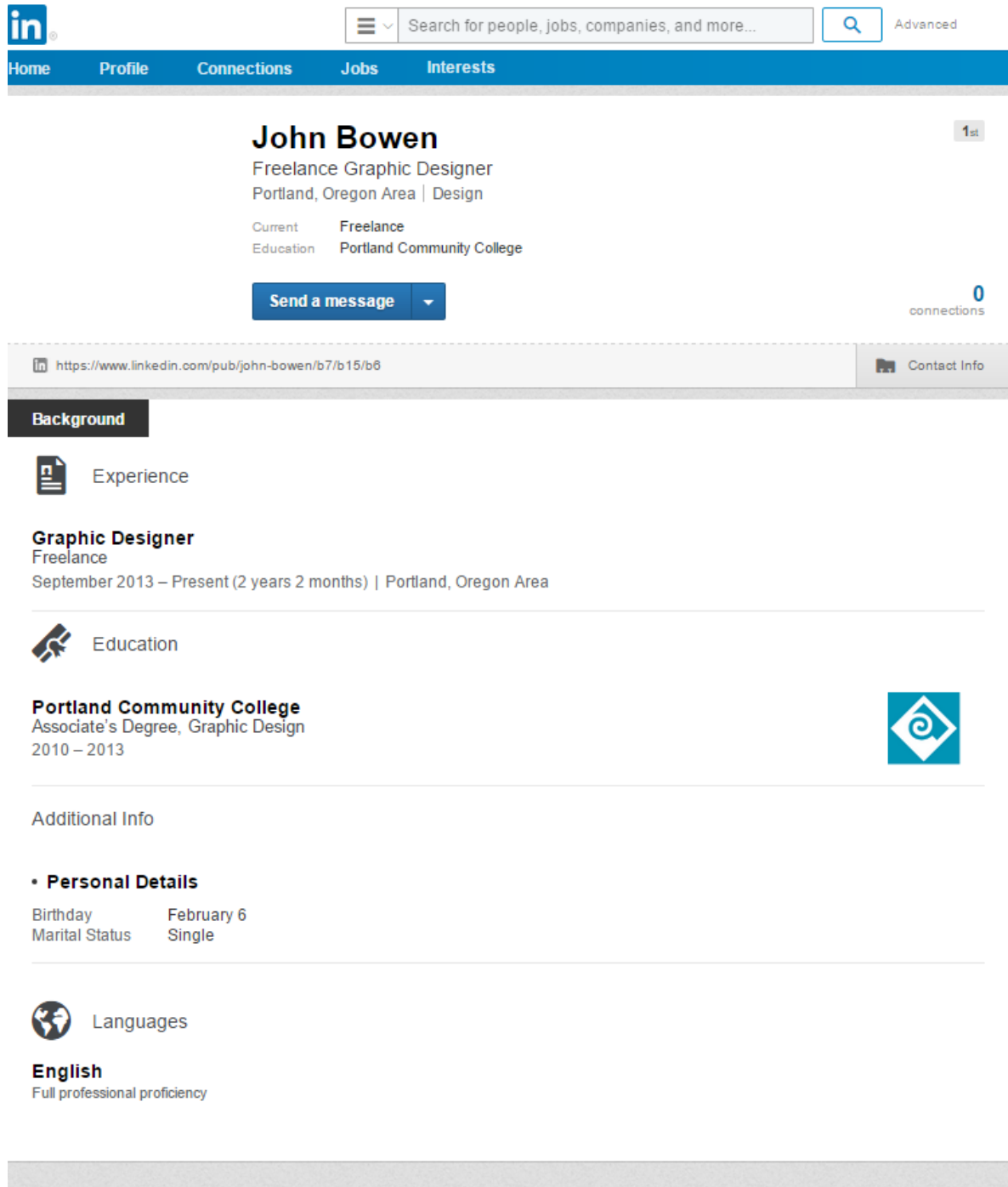
5.2.2 *Case Study 2*

Let us consider an hypothetical artist named John Bowen. John is applying for a designer position at a company, and he has recently completed his university degree. Besides his standard education, he also collaborated in several projects that are hosted on Behance.

Since he never created his own CV previously, he starts of with an empty CV in CVGenie.

To fill in his basic information, he imports it from his LinkedIn profile. Figure 43 shows his LinkedIn profile, and fig. 44 and listing 24 show the resulting CV (in PDF and XML Europass formats resp.).

Figure 43.: Jack Bowen's LinkedIn profile

Figure 44.: Jack Bowen's resulting CV after importing from LinkedIn

```xml
 1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
 2 <SkillsPassport schemaLocation="http://europass.cedefop.europa.eu/Europass http://europass.
       cedefop.europa.eu/xml/v3.3.0/EuropassSchema.xsd" locale="en">
 3   <DocumentInfo>
 4     <DocumentType>ECV_ESP</DocumentType>
 5     <Bundle/>
 6     <CreationDate>2015-09-16T15:36:47.960Z</CreationDate>
 7     <LastUpdateDate>2015-09-16T15:37:19.936Z</LastUpdateDate>
 8     <XSDVersion>V3.3</XSDVersion>
 9     <Generator>EWA</Generator>
10     <EuropassLogo>true</EuropassLogo>
11   </DocumentInfo>
12   <PrintingPreferences>
13     <Document type="ECV">
14       <Field name="LearnerInfo.Identification.PersonName" show="true" order="FirstName
             Surname"/>
15       <Field name="LearnerInfo.Identification.Demographics.Birthdate" show="false" format="
             text/long"/>
16       <Field name="LearnerInfo.WorkExperience[0].Period" show="false" format="text/long"/>
17       <Field name="LearnerInfo.Education[0].Period" show="false" format="text/long"/>
18       <Field name="LearnerInfo" show="false" order="Identification Headline WorkExperience
             Education Skills Achievement ReferenceTo"/>
19     </Document>
20   </PrintingPreferences>
21   <LearnerInfo>
22     <Identification>
23       <PersonName>
24         <FirstName>John</FirstName>
25         <Surname>Bowen</Surname>
26       </PersonName>
27       <ContactInfo>
28         <Address>
29           <Contact>
30             <Municipality>Portland</Municipality>
31             <Country>
32               <Code>US</Code>
```

```
33              <Label>Oregon Area</Label>
34            </Country>
35          </Contact>
36        </Address>
37      </ContactInfo>
38    </Identification>
39    <WorkExperienceList>
40      <WorkExperience>
41        <Period>
42          <From year="2013" month="--09"/>
43          <Current>false</Current>
44        </Period>
45        <Position>
46          <Label>Graphic Designer</Label>
47        </Position>
48        <Employer>
49          <Name>Freelance</Name>
50        </Employer>
51      </WorkExperience>
52    </WorkExperienceList>
53  </LearnerInfo>
54  <CoverLetter/>
55 </SkillsPassport>
```

Listing 24: Jack Bowen's resulting CV after importing from LinkedIn in XML

As shown above, it was only possible to import Jack's name, address and work experience.

John then loads up his Behance profile. He tries to identify which of his designs are more popular, and the ones he is most proud of. When he is ready, he opens the "Import from Behance" page on CVGenie.

The page asks him to paste the URLs to the projects he wishes to include in his CV, and how many images should be included. John copies over the URLs to the page, and selects the option to only import the first image of each project. Figures 45 to 48 show John's Behance profile and the projects he selected, as well as the resulting CV. Listing 25 shows the CV's Europass XML representation.

Figure 45.: Jack Bowen's Behance profile
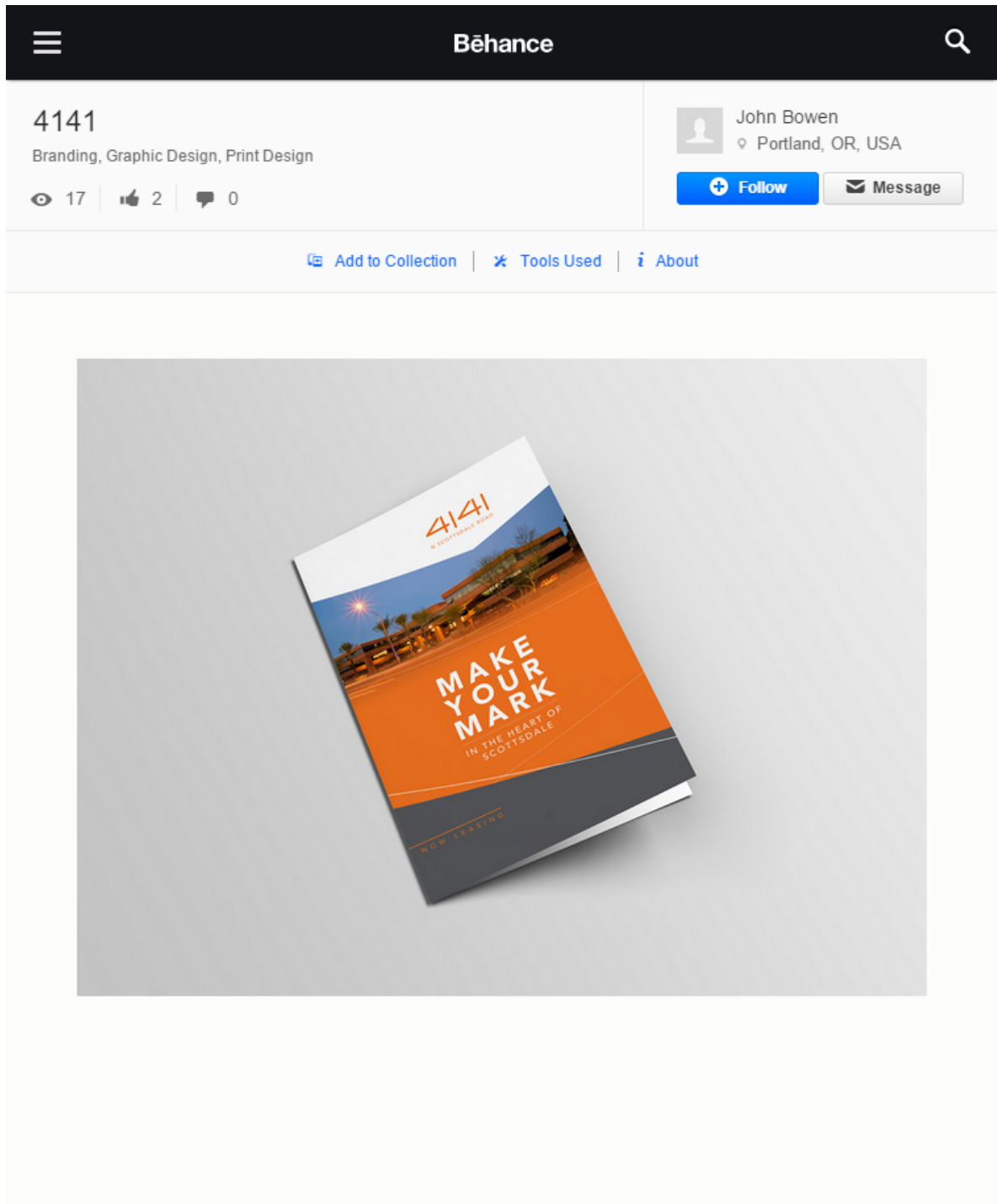
Figure 46.: Jack Bowen's Behance project

europass Curriculum vitae

PERSONAL INFORMATION	John Bowen
	📍 Portland (United States)

WORK EXPERIENCE	
09/2013	Graphic Designer
	Freelance

ADDITIONAL INFORMATION	
Projects	▪ **The Saw Man Artwork**
	October 21, 2015
	album artwork and merch for a musical sawist
	https://www.behance.net/gallery/13524154/The-Saw-Man-Artwork
	▪ **Airport Branding**
	October 17, 2015
	The goal was to create a brand supported by a print and web campaign that would stand out among the many industrial commercial real estate bi…
	https://www.behance.net/gallery/546321545/Airport-Branding
	▪ **4141**
	September 29, 2015
	The concept was to create an iconic, modern and timeless design that highlighted the adaptable building framework, distinctive architecture and v…
	https://www.behance.net/gallery/54315156/4141

Related document(s): The Saw Man Artwork (1) , Airport Branding (1) , 4141 (1)

ANNEXES	
	▪ The Saw Man Artwork (1)
	▪ Airport Branding (1)
	▪ 4141 (1)

Figure 47.: Jack Bowen's resulting CV after importing from Behance (1/2)

5.2. Case Studies

4141 (1) 🔗

Figure 48.: Jack Bowen's resulting CV after importing from Behance (2/2)

110

```
 1  <?xml version='1.0' encoding='UTF-8'?>
 2  <SkillsPassport xmlns="http://europass.cedefop.europa.eu/Europass" xmlns:xsi="http://www.w3.
        org/2001/XMLSchema-instance" xsi:schemaLocation="http://europass.cedefop.europa.eu/
        Europass http://europass.cedefop.europa.eu/xml/v3.3.0/EuropassSchema.xsd" locale="en">
 3    <DocumentInfo>
 4      <DocumentType>ECV_ESP</DocumentType>
 5      <CreationDate>2015-09-16T15:36:47.960Z</CreationDate>
 6      <LastUpdateDate>2015-10-23T16:11:39.638Z</LastUpdateDate>
 7      <XSDVersion>V3.3</XSDVersion>
 8      <Generator>EWA</Generator>
 9      <EuropassLogo>true</EuropassLogo>
10    </DocumentInfo>
11    <PrintingPreferences>
12      <Document type="ECV">
13        <Field name="LearnerInfo.Identification.PersonName" show="true" order="FirstName
            Surname"/>
14        <Field name="LearnerInfo.Identification.Demographics.Birthdate" show="false" format="
            numeric/long"/>
15        <Field name="LearnerInfo.WorkExperience[0].Period" show="false" format="numeric/long"/
            >
16        <Field name="LearnerInfo.Education[0].Period" show="false" format="numeric/long"/>
17        <Field name="LearnerInfo" show="false" order="Identification Headline WorkExperience
            Education Skills Achievement ReferenceTo"/>
18      </Document>
19    </PrintingPreferences>
20    <LearnerInfo>
21      <Identification>
22        <PersonName>
23          <FirstName>John</FirstName>
24          <Surname>Bowen</Surname>
25        </PersonName>
26        <ContactInfo>
27          <Address>
28            <Contact>
29              <Municipality>Portland</Municipality>
30              <Country>
31                <Code>US</Code>
32                <Label>United States</Label>
33              </Country>
34            </Contact>
35          </Address>
36        </ContactInfo>
37      </Identification>
38      <WorkExperienceList>
39        <WorkExperience>
40          <Period>
41            <From year="2013" month="--09"/>
42            <Current>false</Current>
43          </Period>
44          <Position>
45            <Label>Graphic Designer</Label>
46          </Position>
47          <Employer>
48            <Name>Freelance</Name>
```

```
49              </Employer>
50            </WorkExperience>
51        </WorkExperienceList>
52        <AchievementList>
53          <Achievement>
54            <Title>
55              <Label>Projects</Label>
56            </Title>
57            <Description>&lt;ul>&lt;li>&lt;b>The Saw Man Artwork&lt;/b>&lt;br />October 21,
                  2015&lt;br />album artwork and merch for a musical sawist&lt;br />&lt;a href="
                  https://www.behance.net/gallery/13524154/The-Saw-Man-Artwork" rel="nofollow">
                  https://www.behance.net/gallery/13524154/The-Saw-Man-Artwork&lt;/a>&lt;/li>&lt;
                  li>&lt;b>Airport Branding&lt;/b>&lt;br />October 17, 2015&lt;br />The goal was
                  to create a brand supported by a print and web campaign that would stand out
                  among the many industrial commercial real estate  b i &lt;br />&lt;a href="https
                  ://www.behance.net/gallery/546321545/Airport-Branding" rel="nofollow">https://
                  www.behance.net/gallery/546321545/Airport-Branding&lt;/a>&lt;/li>&lt;li>&lt;b
                  >4141&lt;/b>&lt;br />September 29, 2015&lt;br />The concept was to create an
                  iconic, modern and timeless design that highlighted the adaptable building
                  framework, distinctive architecture and  v &lt;br />&lt;a href="https://www.
                  behance.net/gallery/54315156/4141" rel="nofollow">https://www.behance.net/
                  gallery/54315156/4141&lt;/a>&lt;/li>&lt;/ul></Description>
58            <Documentation>
59              <ReferenceTo idref="ATT_1288005400310"/>
60              <ReferenceTo idref="ATT_6618367962519"/>
61              <ReferenceTo idref="ATT_0297209533751"/>
62            </Documentation>
63          </Achievement>
64        </AchievementList>
65        <Documentation>
66          <ReferenceTo idref="ATT_1288005400310"/>
67          <ReferenceTo idref="ATT_6618367962519"/>
68          <ReferenceTo idref="ATT_0297209533751"/>
69        </Documentation>
70    </LearnerInfo>
71    <AttachmentList>
72      <Attachment id="ATT_1288005400310">
73        <Name>The Saw Man Artwork (1)</Name>
74        <MimeType>image/jpeg</MimeType>
75        <Data>/9j/4gxYSUNDX1BST0ZJTEUAAQEAAAxITGlubwIQAABtbnRyUkdCIFhZWiAHzgACAAkABgAxAABh
76          <!-- (...) -->
77        </Data>
78        <Description>The Saw Man Artwork (1)</Description>
79      </Attachment>
80      <Attachment id="ATT_6618367962519">
81        <Name>Airport Branding (1)</Name>
82        <MimeType>image/jpeg</MimeType>
83        <Data>/9j/4gxYSUNDX1BST0ZJTEUAAQEAAAxITGlubwIQAABtbnRyUkdCIFhZWiAHzgACAAkABgAxAABh
84          <!-- (...) -->
85        </Data>
86        <Description>Airport Branding (1)</Description>
87      </Attachment>
88      <Attachment id="ATT_0297209533751">
89        <Name>4141 (1)</Name>
90        <MimeType>image/jpeg</MimeType>
```

```
91        <Data>/9j/2wDFAAMCAgMCAgMDAwMEAwMEBQgFBQQEBQoHBwYIDAoMDAsKCwsNDhIQDQ4RDgsLEBYQERMU
92          <!-- (...) -->
93        </Data>
94        <Description>4141 (1)</Description>
95      </Attachment>
96    </AttachmentList>
97 </SkillsPassport>
```

Listing 25: Jack Bowen's resulting CV after importing from Behance in XML

<div align="right">

6

</div>

CONCLUSION

This masters' project is entitled "Automatic Creation of CVs from Online Social Networks". The underlying research hypothesis is: *Is it possible, in a proper and effective way, to create a curriculum vitae automatically, using existing information available online in various social networks*?

To prove that hypothesis, several tasks, described below, were accomplished, giving rise to the project contributions:

1. Compare several existing CV formats, analyzing aspects such as the popularity of the formats and the information and areas of expertise that they support, with the objective of choosing the most appropriate format to serve as the base format for this master work. We chose the Europass CV format.

2. Study existing online social networks: what social networks are the best sources for professional information about individuals, either it being more generic or specific, and what information do they provide.

3. Compare existing solutions able to retrieve information online and subsequently creating CVs, while identifying their weak points.

4. Study how the Europass CV format could support information about specific areas of expertise, exemplifying with two case studies.

5. Detail how the information from the chosen online social networks could be integrated into the Europass CV format.

6. Describe the methodology for retrieving information from online social networks.

7. Apply the conducted research to implement a system capable of satisfying the objective of this master work, and detail the implementation of that system.

In the next paragraphs the seven tasks above are detailed and the thesis contributions will be emphasized. The present state of the project is reported and the main conclusions of each phase are remembered. After that some directions for future work are sketched.

CV FORMAT ANALYSIS

This master work begun with a thorough research phase.

In this phase, several CV formats were analyzed. One of the goals of this analysis was to identify a single format that would become the base format behind the proposed solution. For this purpose, the format would have to satisfy several requirements:

- Be relatively popular and have widespread use;
- Be generic, in the sense that it could include the common information about an individual;
- Be extensible, so that it could also include information relative to specific areas of expertise;
- Be open, to guarantee that it can be easily integrated into any system.

The Europass CV format revealed itself to be the ideal format, since it satisfies all of these requirements.

The other goal behind the analysis of CV formats was to identify case studies in which other types of CV formats were required. In these case studies, the areas of expertise required CV formats that could support more specific information that usually isn't supported by more generic formats. With this analysis, two case studies were defined that would continue to be analyzed through the master work: individuals with academic backgrounds, and individuals with academic and artistic backgrounds.

After analyzing the various CV formats, a comparison was made to explore their similarities and differences. This was followed by a in-depth analysis of the Europass CV format and its structure.

ONLINE SOCIAL NETWORKS ANALYSIS

The research continued by investigating sources of professional information, which would be online social networks (OSNs). These OSNs were categorized by their popularity and the information that they could provide. LinkedIn appeared to be the best source of common information, since its popularity is unmatched and the information it offers can be very detailed. Other social networks and websites were also identified and analyzed, to satisfy the recurrent case studies: Academia and Research Gate satisfy the need for most academic information; Google Scholar is the primary source for publications; and Behance is a good source for information about individuals' artistic endeavors.

EXISTING SOLUTIONS ANALYSIS

Existing solutions that shared the objective of this master work were also studied. The purpose of this analysis was to categorize the existing solutions by their features, and also identify their shortcomings, so that they could be addressed afterwards, in order to propose a solution that would overcome those

weak points and arise as objectively better. The major flaws that are common between these solutions are:

- Not using a standard format: most solutions have their own CV format, each with a different layout and presentation, and therefore none of them use a standard format;

- Not being able to export CVs in an editable format: they obtain external information and create CVs from it, but the CVs are then kept inside these systems, since it is not possible to export them in an editable (and ideally standard) format. This makes the users more dependant of the system, because once they start using it, they can't withdraw their CVs from it, and tend to continue using it for that reason. Also, because of this, these systems are less inter-operational.

- Being limited in the data sources that they use: most of the examined systems only import information from LinkedIn.

- Manage common information only: none of the analyzed solutions allow importing and managing information from specific areas of expertise, limiting themselves to common information only.

INTEGRATION OF INFORMATION IN THE EUROPASS CV FORMAT

A comprehensive analysis of the Europass CV XML format was made to clarify exactly what information was supported in the analyzed CV formats and online social networks and where that information could be integrated into Europass CV XML documents, from a technical standpoint.

In this analysis, every item of information that was supported by the CV formats respective to the case studies was deemed supported as well by the Europass CV format; in the same way, all of the information available in the selected online social networks was also considered integrable in this format.

Furthermore, through this analysis, it was possible to create a strategy to include any type of information in this format, regardless of the area of expertise.

This proved that the Europass CV format is as versatile as initially suspected, and justified the decision to use it as the base format of the system.

CONCEPTION OF THE PROPOSED SOLUTION

The proposed solution improves upon all of the alternatives that were examined.

It uses the Europass CV format, which is an open standard, as a base format. This allows the system to easily import and export CV documents in a format that is popular and open; those CV documents can then be easily introduced into other systems that support this format, and vice-versa.

It also uses more data sources than simply LinkedIn. Research Gate, Google Scholar and Behance are all supported data sources, making the system able to support both common information as well as more specific information from the academic and artistic areas of expertise.

The interoperability of the system is further explored by interacting with the official Europass web services, that complement the proposed system by allowing it to export CV documents in multiple formats as well as enabling the user to edit those documents online.

Formally, the proposed system's requirements were specified as using a standard CV format, being able to import exiting CV documents as files, import information from online media, manage information for specific areas of expertise and export CVs in both read-only formats and in an editable format. The proposed system's architecture was specified, with a thorough description of all the components it would require; its data model was also defined.

DATA RETRIEVAL METHODOLOGY

The methodology that should be used to retrieve information from the online social networks that were selected was also subject of analysis.

From the several methods available, two methods were thoroughly examined and described: the first is to interact with an API that is purposefully designed by the company in charge of the website; the other is "Web Scraping", in which the websites' web pages are retrieved and parsed.

Both methods have their advantages and disadvantages, which were explicitly stated. Nevertheless, both proved viable and were applied in the proposed solution's implementation.

IMPLEMENTATION OF THE PROPOSED SOLUTION

With the research completed, the proposed system was implemented, and it was entitled "CVGenie".

The proposed solution came to fruition in a multi-featured system that is capable of everything that was proposed and, due to Europass's inherent interoperability, was also capable of extra features, such as exporting CVs in additional formats and seamlessly use Europass's online editor.

To illustrate how the implemented solution functions, the most important features were presented via a series of instructions and figures. With that purpose as well, two case studies were defined and presented. These represent two realistic uses of the application, and featured most of the features that the CVGenie has to offer.

The first case study was about a university professor's CV. It displayed how the professor could import his common and academic information from Research Gate, and his publications from Google Scholar; it also showed an instance where conflict management was required, and the result of exporting the final CV in several formats, like a PDF document and a HTML web page.

The second case study showed how an artist could create his first CV, including common information from his LinkedIn profile and his most prized art projects from Behance as well.

FUTURE WORK

This master work proved that it is possible and viable to produce CV documents using information from online social networks. It also explored the possibilities of the Europass CV format, by trying to integrate not only common information into a CV, but also information from specific areas of expertise, which is done for two selected case studies. Possible future work could entail:

- In this master work, the Europass CV format was selected as a base format. One of the reasons for this is that this format is versatile, and can easily support information from any area of expertise. With CVGenie, two case studies were successfully implemented, and a strategy to support any area of expertise was defined.

  The key of Europass's versatility is its lack of strictness in its structure, in some sections of the format. One possible next step would be to extend this format, and further define its structure for items relative to various areas of expertise.

  Additional work could also be done with the intent of supporting any area of expertise: for example, consider creating a system that would interact with other systems, such as ontology management systems, to obtain meta-data about the structure of the document. The ontology would supply the fields that should exist in a CV included in a specific area of expertise, and that meta-data could then be used to define the structure of the CV document, augmenting the default Europass CV structure.

- Two different methodologies were used in CVGenie to extract information from four online social networks. The implementation of the data retrieval process was different for each online social network, but the methodologies that were used to do so were the same.

  Another system could be implemented to surpass CVGenie, in respect of the online social networks that could be supported. To do so, it would be feasible to further refine the data retrieval process, so that the same partial implementation could be used to interact with any online social network, and only specific details would require further implementation. This generic approach to data retrieval would allow creating a much more versatile system.

- In CVGenie, the sources of information that were selected were online social networks. Another approach would be to obtain information from alternative sources. The Semantic Web defines a set of standards that websites can use to include information in a standard manner.

  It could be feasible to build a system that would autonomously navigate and collect information from websites based on the Semantic Web standards.

  Other sources of information could be considered as well.

A

FEATURE DISPLAY

A.1   IMPORT FROM LINKEDIN - BEFORE SIGN IN

A.2   IMPORT FROM LINKEDIN - SIGNING IN
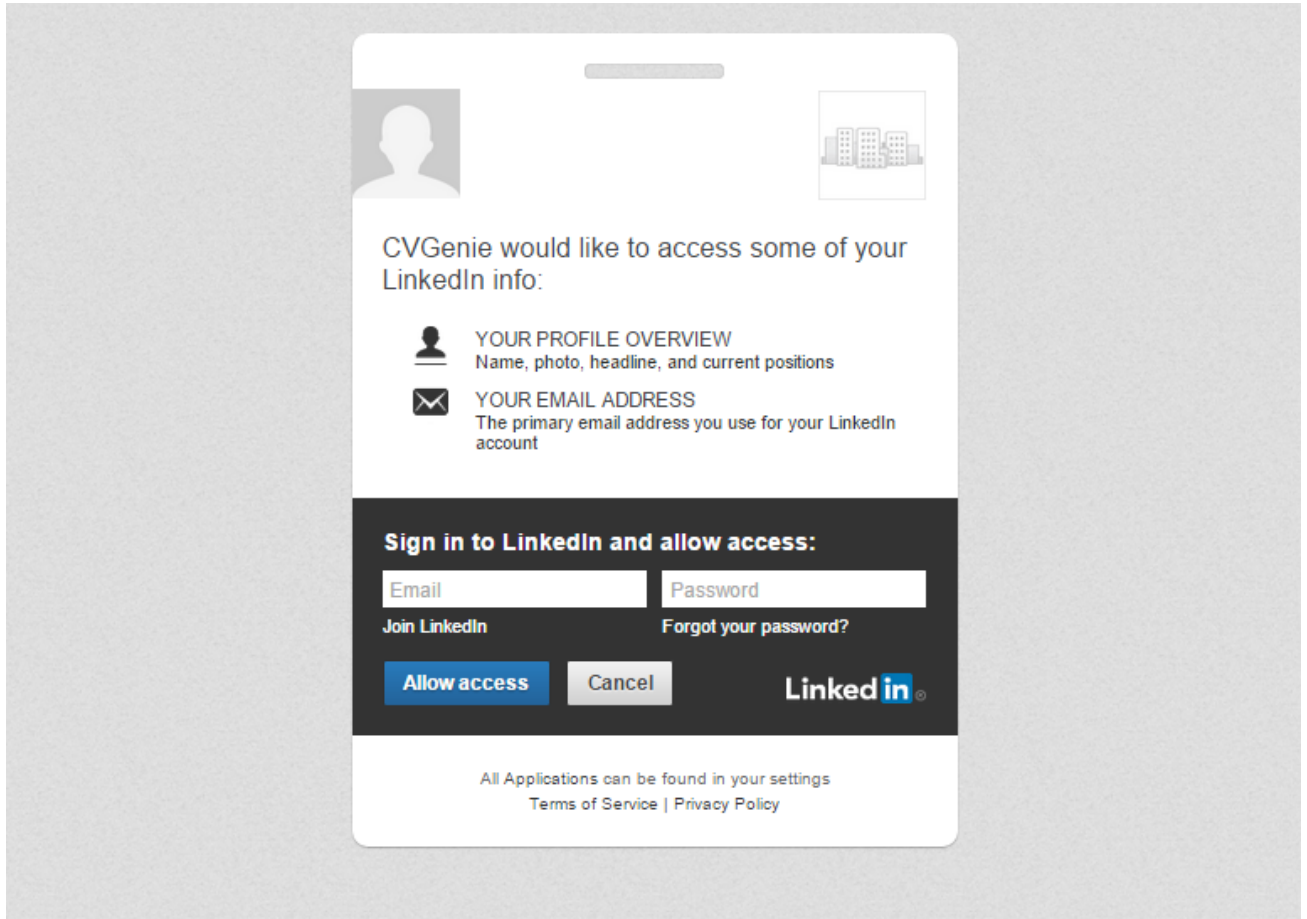
A.3   IMPORT FROM LINKEDIN - AFTER SIGN IN

A.4   IMPORT FROM GOOGLE SCHOLAR - FIRST TIME IMPORTING

CVGenie   My CVs   sergio@cvgenie.com ▾

Import Data From Google Scholar   Back

You can import data from Google Scholar here.
To do so, we need to identify your Google Scholar profile.

Warning: Any existing publications on your CV will be replaced. This process may take a few minutes.

Import from a new Google Scholar profile

You need to find your Google Scholar profile, copy the URL, and paste it here.

First type your name and press "Search". A new tab/window will be opened.

Full name:

| Thomas Evans Smith | Search |

Then open the right profile, copy the URL and paste it below.

Profile URL:

https://scholar.google.com/citations?user=123456789123=en

Import

A.5 EXPORT CV AS PDF

europass Curriculum vitae

PERSONAL INFORMATION	Pedro Rangel Henriques

WORK EXPERIENCE

01/1994–Present **Researching (Group Leader)**
University of Minho, Departamento de Informática (DI), Language Processing Group (gEPL), Guimarães (Portugal)

01/1996–Present **Teaching (Associate Professor)**
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

01/1993–01/1996 **Teaching (Auxiliary Professor)**
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

01/1981–01/1993 **Teaching (Assistant Lecturer)**
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

EDUCATION AND TRAINING

01/2002–01/2012 **Computer Science, Habilitation**
University of Minho, Guimarães (Portugal)

01/1988–01/1993 **Computer Science, Doctor of Philosophy**
University of Minho, Guimarães (Portugal)

01/1981–01/1984 **Computer Science, Master of Science**
University of Minho, Guimarães (Portugal)

01/1972–01/1977 **Electro-technical Engineering, Bachelor of Engineering**
University of Porto, Porto (Portugal)

PERSONAL SKILLS

Other language(s)

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	
English				
Portuguese				

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

Job-related skills
- Data Analysis and Knowledge Extraction
- Programming Paradigms
- Methods and Languages

A.6 EXPORT CV AS WORD DOCUMENT

A.7 EXPORT CV AS OPEN DOCUMENT

CV 1.ott | − | + | Automatic | ⇕ | Choose Fil | 🖥 | ⛶ | **ViewerJ**

PERSONAL INFORMATION   Pedro Rangel Henriques

WORK EXPERIENCE

01/1994–Present   Researching (Group Leader)
University of Minho, Departamento de Informática (DI), Language Processing Group (gEPL), Guimarães (Portugal)

01/1996–Present   Teaching (Associate Professor)
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

01/1993–01/1996   Teaching (Auxiliary Professor)
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

01/1981–01/1993   Teaching (Assistant Lecturer)
University of Minho, Departamento de Informática (DI), Guimarães (Portugal)

EDUCATION AND TRAINING

01/2002–01/2012   Computer Science, Habilitation
University of Minho, Guimarães (Portugal)

01/1988–01/1993   Computer Science, Doctor of Philosophy
University of Minho, Guimarães (Portugal)

01/1981–01/1984   Computer Science, Master of Science
University of Minho, Guimarães (Portugal)

01/1972–01/1977   Electro-technical Engineering, Bachelor of Engineering
University of Porto, Porto (Portugal)

PERSONAL SKILLS

Other language(s)

UNDERSTANDING		SPEAKING		WRITING
Listening	Reading	Spoken interaction	Spoken production	

English
Portuguese

Levels: A1 and A2: Basic user - B1 and B2: Independent user - C1 and C2: Proficient user
Common European Framework of Reference for Languages

126

A.8   EXPORT CV AS HTML

A.9   EDIT IN EXTERNAL EDITOR

BIBLIOGRAPHY

[1] *Academia.edu - Share research*. 2014. URL: http://www.academia.edu.

[2] *Academic Careers Curriculum Vitae - The Career Center of the University of Washington*. 2014. URL: http://careers.washington.edu/ifiles/all/files/docs/gradstudents/pdfs/AcademicCareers-Curriculum_Vitae_07-08.pdf.

[3] *Create my CV Online for free*. 2015. URL: http://Create-CV.com.

[4] *Developer Program Transition | LinkedIn Developer Network*. 2015. URL: http://developer.linkedin.com/support/developer-program-transition.

[5] *DoYouBuzz: Your best resume*. 2015. URL: http://DoYouBuzz.com.

[6] *Europass: Curriculum Vitae*. 2014. URL: http://europass.cedefop.europa.eu/en/documents/curriculum-vitae.

[7] *Getting Started: REST API | LinkedIn Developer Network*. 2015. URL: http://developer.linkedin.com/docs/rest-api.

[8] *Getting Started with Rails - Ruby on Rails Guides*. 2015. URL: http://guides.rubyonrails.org/getting_started.html.

[9] *Google Scholar*. 2014. URL: http://scholar.google.pt.

[10] *Home | Europass Interoperability*. 2015. URL: http://interop.europass.cedefop.europa.eu.

[11] *LinkedIn: World's Largest Professional Network*. 2014. URL: http://www.linkedin.com.

[12] *Online Editor: Curriculum Vitae*. 2015. URL: http://europass.cedefop.europa.eu/editors/en/cv/compose.

[13] *Online Portfolios on Behance*. 2014. URL: http://www.behance.net.

[14] *ResearchGate*. 2014. URL: http://www.researchgate.net.

[15] *Ruby on Rails*. 2015. URL: http://rubyonrails.org.

[16] *Ruby Programming Language*. 2015. URL: http://www.ruby-lang.org/en.

[17] *Software to Import Resumes*. 2015. URL: http://www.egrabber.com/resumegrabbersuite.

[18] *Standards and Guidelines | College Art Association | CAA*. 2014. URL: http://www.collegeart.org/guidelines/visartcv.

[19]   *Turn your LinkedIn Profile into a Resume | Resume Builder*. 2015. URL: http://resume.linkedinlabs.com.

[20]   *VisualCV - Online CV Builder and Professional Resume CV Maker*. 2015. URL: http://VisualCV.com.