



GRADO DE INGENIERÍA EN TECNOLOGÍAS DE LA
TELECOMUNICACIÓN

Curso Académico 2017/2018

Trabajo Fin de Grado

CONFERENCE BOOK GENERATOR

Autora : Seila Oliva Muñoz

Tutor : Dr. Gregorio Robles

Trabajo Fin de Grado

Conference Book Generator

Autora : Seila Oliva Muñoz

Tutor : Dr. Gregorio Robles Martínez

La defensa del presente Proyecto Fin de Carrera se realizó el día de
de 2018, siendo calificada por el siguiente tribunal:

Presidente:

Secretario:

Vocal:

y habiendo obtenido la siguiente calificación:

Calificación:

Fuenlabrada, a de de 2018

*Dedicado a
mi familia*

Agradecimientos

Quiero dar las gracias en primer lugar a mis padres, Gustavo y Yolanda, por apoyarme en todo momento, desde el primer día que decidí hacer esta carrera, y por todo lo que hacen día a día por mí. Y aunque no se lo diga a menudo, no sé que haría sin ellos.

También quiero agradecerse a mi hermano, Kevin, ya que gracias a él descubrí esta carrera y, aunque no se de cuenta, me ha ayudado muchísimo con ella. Espero poder lograr los mismos éxitos que él está logrando en su vida, o al menos lo intentaré.

Y por último, pero no por ello menos importante, tengo que dar las gracias a mi pareja, Adrián, por soportarme estos cuatro años de carrera, por todos los días sin salir y los planes que no podíamos hacer porque yo tenía que estudiar, por darme apoyo y fuerzas cuando más lo necesitaba y, sobre todo, por estar a mi lado durante esta etapa de mi vida.

Quiero acordarme también del resto de mi familia, en especial de mis abuelos, ya que estoy segura que estarán muy orgullosos y contentos por mí.

Resumen

En la actualidad es muy frecuente la realización de seminarios y conferencias en las que se tratan temas concretos de diferentes índoles, a las que suelen ir personas especializadas en el tema a tratar dispuestas a conocer otras personas que compartan sus mismos intereses, incluso buscando personas con las que trabajar.

Es por esto que surge la idea de tener la información de todos los asistentes para conocer sus intereses, así como sus datos de contacto por si interesa una posible colaboración/contratación.

Mi proyecto consiste en la elaboración de un *book* en el que aparezcan estos datos de todos los participantes. Para ello, previamente los participantes tendrán que rellenar un formulario de *Google Forms*. Una vez se hayan recopilado los datos de todos los participantes, se trasladarán a una *Google Sheet*, cuya URL se introducirá en el programa que he elaborado con *Python*.

Se extraerán los datos de los participantes de esta hoja de cálculo, y se manipularán en caso de que sea necesario para, con ellos, generar un documento en \LaTeX . Este documento se compilará para obtener un PDF en el que aparecerán los datos de todos los asistentes y estará listo para ser impreso en forma de revista.

Summary

At present, it is very common to make seminars and conferences that address specific issues of different kinds, where usually go people specialized in the issue who be willing to meet other people who share their same interests, even looking for people with those that work.

This is why the idea of having the data of all the attendees to know their interests and contact data in case a possible collaboration/contracting interests.

My project consists in the elaboration of a *book* in which these data of all the participants appear. To do this, previously the participants will have to fill a *Google Forms* form. Once the data of all the participants has been collected, they will be moved to a *Google Sheet*, whose URL will be introduced in the program that I have created with **Python**.

The data of the participants of this spreadsheet will be extracted, and they will be manipulated in case it is necessary to generate a document in \LaTeX with them. This document will be compiled to obtain a PDF in which the data of all the attendees will appear and it will be ready to be printed in the form of a magazine.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Estructura de la memoria	2
2. Objetivos	3
2.1. Objetivo general	3
2.2. Objetivos específicos	3
2.3. Planificación temporal	4
3. Estado del arte	5
3.1. Python	5
3.1.1. Python 2 y Python 3	6
3.2. L ^A T _E X	7
3.2.1. Creación de un documento L ^A T _E X	8
3.2.2. XeTeX	10
3.3. Google Forms	10
3.4. Google Sheets API	10
3.4.1. Método <i>spreadsheets.value.get</i>	11
3.4.2. OAuth 2.0	12
3.5. Pycountry	13
3.6. Pillow	13
3.7. Pandas	13
3.8. Matplotlib	14

4. Diseño e implementación	15
4.1. Arquitectura general	15
4.2. Creación de un formulario (Google Forms)	16
4.3. Traspaso de respuestas a hoja de cálculo	16
4.4. Extracción de los valores de la hoja de cálculo	17
4.4.1. Volcado de valores a NamedTuple	18
4.5. Fichero de errores	19
4.6. Proceso de tratamiento de información	19
4.6.1. Ordenar alfabéticamente	20
4.6.2. Convertir caracteres especiales	21
4.6.3. Descarga de imágenes personales	21
4.6.4. Obtener banderas de nacionalidad	23
4.6.5. Acortar presentaciones	24
4.7. Creación de la sección de participantes en \LaTeX	25
4.8. Apartados opcionales	25
4.8.1. Diagramas de sectores	26
4.8.2. Listados de personas	26
4.9. Guardar el fichero \LaTeX con los datos	26
4.10. Unir automáticamente documentos \LaTeX	27
4.11. Compilado del documento \LaTeX	27
4.12. Personalización	28
4.12.1. Personalización avanzada	29
5. Resultados	31
6. Conclusiones	35
6.1. Consecución de objetivos	35
6.2. Aplicación de lo aprendido	35
6.3. Lecciones aprendidas	36
6.4. Trabajos futuros	37
Bibliografía	39

<i>ÍNDICE GENERAL</i>	XI
A. Manual de usuario	41
A.1. Dónde descargar la aplicación	42
B. Ejemplo de uso	43
B.1. 15 th International Conference on Mining Software Repositories (MSR) 2018 . .	44
B.2. 14 th International Conference on Open Source Systems (OSS) 2018	72
B.3. 11 th Seminar on Advanced Techniques & Tools for Software Evoluton SATToSE 2018	80

Índice de figuras

3.1. Ejemplo de preámbulo	9
3.2. Ejemplo de cuerpo del documento	9
3.3. Esquema solicitud token para OAuth 2.0	12
4.1. Esquema de funcionamiento	15
4.2. Tipos de respuesta de Google Forms	17
4.3. Ejemplo hoja de cálculo a array	18
4.4. Ejemplo NamedTuple	18
4.5. Estructura información participante	25
5.1. Página ConfBook MSR 2018	32
5.2. Página ConfBook OSS 2018	33
5.3. Página ConfBook SATToSE 2018	34

Capítulo 1

Introducción

1.1. Motivación

Hoy en día es muy frecuente la realización de seminarios o conferencias de muy diversos temas. A ellos acude gente de diferentes partes del mundo, con perfiles similares o diferentes, dispuestos a aumentar su lista de contactos.

Normalmente, la única manera de conocer gente en estos lugares es presentándose personalmente a cada persona, y en una conferencia con un número grande de personas esta labor puede resultar tediosa, por no decir prácticamente imposible.

Por esto, mi tutor, Gregorio, me contó la idea de crear un “Book” o revista en el que aparecieran los datos de todos los asistentes a la conferencia, para repartirlo en el lugar, y así tener el contacto de todos. En este libro cada asistente podrá incluir sus datos de contacto, su centro de trabajo, sus intereses, etc.

Este proyecto pretende automatizar la labor de hacer este Book, de manera que cada asistente rellene un formulario previamente con sus datos, y que el documento se genere automáticamente.

1.2. Estructura de la memoria

Con el fin de facilitar la lectura de la memoria, en esta sección se detalla la estructura de la misma y el contenido que se trata en cada una de las secciones:

- **Capítulo 1: Introducción.** Contexto en el que surge la idea del proyecto y breve explicación de la estructura de la memoria.
- **Capítulo 2: Objetivos.** En esta sección se explican los objetivos marcados durante la realización del proyecto.
- **Capítulo 3: Estado del arte.** Presentación de las tecnologías con las que se ha desarrollado el proyecto.
- **Capítulo 4: Diseño e implementación.** Explicación del desarrollo del proyecto paso a paso.
- **Capítulo 5: Resultados.** Análisis de los resultados obtenidos.
- **Capítulo 6: Conclusiones.** Conclusiones finales tras la realización del trabajo, como los objetivos conseguidos, posibles mejoras futuras del proyecto, cómo he aplicado lo aprendido en la carrera o lo que he aprendido durante la realización del proyecto.

Capítulo 2

Objetivos

2.1. Objetivo general

El objetivo de este proyecto es crear un recurso para que todas las personas que asistan a una conferencia/seminario puedan conocer al resto de asistentes, así como darse a conocer ellos mismos, y de esta forma poder obtener información sobre cada uno (datos de contacto, situación laboral...).

Para ello, se persigue otro objetivo que es automatizar esta labor: crear un programa que sea capaz de tomar como entrada los datos de cada asistente y generar un documento con toda la información.

2.2. Objetivos específicos

Los objetivos específicos que llevé a cabo a lo largo del proyecto fueron los siguientes:

- **Usar la aplicación para casos reales.** Utilizar la aplicación para generar *books* para conferencias reales, imprimirlos y que se repartan en ellas.
- **Sencillez de uso.** Intentar que la aplicación se pueda ejecutar tanto por personas que están familiarizadas con las tecnologías utilizadas, como por las que no las conocen o no saben mucho sobre ellas.

2.3. Planificación temporal

La realización de este proyecto me ha ocupado todo un curso académico (9 meses), y las fases que he seguido han sido las siguientes:

- **Idea de proyecto.** En primer lugar, me reuní con mi tutor del proyecto, Gregorio, para que me ayudase a elegir un proyecto que se adaptase a lo que estaba buscando. Una vez decidido el proyecto, tomé un tiempo para estudiarlo y saber qué tecnologías iba a necesitar para su desarrollo.
- **Aprendizaje de las tecnologías utilizadas.** Una vez me di cuenta de las tecnologías necesarias, comencé por aprender sobre ellas: para qué sirven, cómo utilizarlas... Este proceso es imprescindible para la continuación del proyecto.
- **Desarrollo del proyecto.** Desarrollo del código en Python para conseguir las ideas anteriormente expuestas.
- **Redacción de la memoria.** Como último paso del proyecto, comencé a escribir esta memoria en la que se recopilan en detalle todos los pasos que seguí para la consecución del proyecto.

Capítulo 3

Estado del arte

En esta sección, se describen las tecnologías utilizadas para la realización del proyecto:

3.1. Python

Python [1] es un lenguaje de programación que busca la simplicidad y facilidad, así como la legibilidad del código. Es un lenguaje interpretado, es decir, no necesita ser compilado para ejecutar.

Fue creado en los finales de los ochenta por Guido Van Rossum en el Centro para las Matemáticas y la Informática (CWI, Centrum Wiskunde & Informatica), en los Países Bajos. El nombre del lenguaje proviene de la afición de su creador por el grupo de humoristas británicos *Monty Python*.

Con Python se pueden hacer todo tipo de programas. Aunque no es un lenguaje específico para el desarrollo web, sí se pueden desarrollar páginas con él.

Es un lenguaje multiparadigma: permite más de un estilo de programación (orientada a objetos, imperativa, funcional...). También es multiplataforma, por lo que se puede usar con distintos sistemas informáticos.

Una ventaja de Python es que cuenta con un gran número de funciones incorporadas en el

lenguaje (por ejemplo, para tratar *strings*, ficheros...) y librerías para usos específicos que podemos importar a nuestro programa, lo que nos ayuda a realizar muchas tareas sin necesidad de programarlas desde cero.

Python es ideal para trabajar con grandes volúmenes de datos, porque favorece su extracción y procesamiento, siendo el elegido por las empresas de Big Data. A nivel científico, posee una amplia biblioteca de recursos con especial énfasis en las matemáticas.

3.1.1. Python 2 y Python 3

Hay dos versiones principales de Python: Python 2 y Python 3. Python 2 ha estado presente durante mucho tiempo, por lo que es la versión más extendida actualmente y la que tiene mayor disponibilidad de bibliotecas. La versión más popular de Python 2 es Python 2.7.

Sin embargo, Python 3 se considera el futuro de Python, ya que está pensado para añadir más características y corregir errores anteriores, y está en continuo desarrollo. Mientras, Python 2 no añadirá ya prácticamente nuevas características. Por ello, Python 3 comienza a adelantar a Python 2 en lo que a usuarios que la usan se refiere.

Las funcionalidades de Python 2 y 3 son, en su mayoría, iguales, pero hay algunas diferencias en la sintaxis y la manipulación. Las principales diferencias entre estas dos versiones son:

- **Print:** quizá es la diferencia más destacada a simple vista, y es que *print* ha pasado de ser una sentencia a ser una función. Esto se traduce, sobre todo, a que con Python 2 se puede escribir `print "Hello World"`, sin necesidad de paréntesis. En cambio, en Python 3 es obligatorio el uso de paréntesis para hacer un *print*: `print("Hello World")`.
- **División entera:** en Python 2 la división de dos números enteros da como resultado un número entero ($3/2 = 1$), y para que el resultado tenga decimales hay que escribir diviendo y divisor en forma de decimal ($3.0/2.0 = 1.5$). En Python 3 la división de enteros ya da como resultado un número decimal ($3/2 = 1.5$).

- **Soporte Unicode:** para hacer una cadena Unicode en Python 2, se debe añadir el caracter ‘u’ a la cadena: `u"Hello World"`, mientras que en Python 3 las cadenas son Unicode por omisión.
- **Next (iterables):** en Python 2, para obtener el siguiente elemento de un iterador, se pueden usar tanto el método `iter.next()` como la función `next(iter)`. Sin embargo, en Python 3 sólo se puede utilizar la función.
- **Input:** en Python 2 hay dos funciones para introducir datos por teclado: `raw_input()` (los datos introducidos los trata como una cadena de texto) e `input()` (los datos introducidos los trata por su tipo, es decir, si introducimos 15 lo trata como un int). En Python 3 se ha eliminado la función `raw_input()`, y en su lugar permanece la función `input()` con las características del `raw_input()` de Python 2.

3.2. \LaTeX

\LaTeX [2] es un sistema de composición de textos (de software libre), orientado especialmente a la creación de libros y documentos científicos y técnicos que contengan fórmulas matemáticas. Está formado por un gran conjunto de macros de \TeX , con la intención de facilitar el uso del lenguaje de composición tipográfica, \TeX .

Una de las ventajas de \LaTeX es la calidad profesional de los documentos que genera, así como su excelente calidad de imprenta. Otra ventaja es que permite separar claramente el contenido y el formato del documento.

\LaTeX presupone una filosofía de trabajo diferente a la de los procesadores de texto habituales (“lo que ves es lo que obtienes”) y se basa en instrucciones. Tradicionalmente, este aspecto se ha considerado una desventaja. Sin embargo, \LaTeX permite a quien escribe un documento centrarse exclusivamente en el contenido, sin tener que preocuparse de los detalles del formato.

El proceso de generación de un documento con \LaTeX consiste en 3 pasos:

- **Editar.** El primer paso es usar un editor para generar un archivo con extensión `.tex`, y en

él escribir el código \LaTeX que describa la estructura y contenido del documento.

- **Compilar.** Compilar es el proceso realizado por el motor de \LaTeX , que convierte los archivos *.tex* en documentos con formato *.pdf*, que se podrán imprimir y ver en pantalla. El proceso de compilación detectará e indicará los errores que tenga el código, que habrá que corregir y volver a compilar para obtener el documento *.pdf*.
- **Visualizar.** Una vez compilado, y sin errores, podremos visualizar el resultado. Hay editores específicos para \LaTeX que permiten compilar y visualizar el documento desde el propio editor; si no fuese el caso, habría que dirigirse al directorio para buscar y abrir el archivo *.pdf* que ha sido generado.

3.2.1. Creación de un documento \LaTeX

Un documento \LaTeX tiene dos partes principales: el preámbulo y el cuerpo del documento.

- **Preámbulo.** Comienza por la instrucción `\documentclass[.]{.}`. En todas las instrucciones, entre [] se incluyen los parámetros opcionales, que en el caso del *documentclass* son el tamaño de letra (11pt, 12pt...), si el texto va a dos columnas (*twocolumn*) o ajuste de los márgenes para imprimir a doble cara (*twoside*). Entre { } se incluyen los parámetros obligatorios, que en este caso es el tipo de documento que estamos elaborando (*article*, *report*, *book*...).

En el preámbulo se pueden incluir instrucciones para activar paquetes que agregan funciones adicionales a \LaTeX , así como datos generales sobre el documento que se está escribiendo. Y al final se incluyen los campos `\title`, `\author` y `\date`, que especifican los datos que irán en el encabezado del documento.

Se puede ver un ejemplo de preámbulo en la Figura 3.1.

- **Cuerpo del documento.** Está delimitado por los comandos `\begin{document}` y `\end{document}`. Si comenzamos el cuerpo con el comando `\maketitle` se escribirán los datos del título con la información que se indicó en el preámbulo (título, autor y fecha). Lo único que queda es empezar a escribir el texto, y \LaTeX se encargará del formato y apariencia. Algunas cosas importantes a tener en cuenta son:

```

1 \documentclass[llpt]{article}
2 \usepackage[utf8]{inputenc}
3 \usepackage[T1]{fontenc}
4 \usepackage{graphicx}
5 \usepackage{subfigure}
6 \usepackage{xcolor}
7 \usepackage{hyperref}
8 \usepackage{multicol}
9 \include{defs}
10
11 \usepackage{lipsum}
12 \usepackage[spanish]{babel}
13 \setcounter{page}{0}
14
15 *****
16 % Title Page
17 \title{}
18 \author{}
19 \date{}
20 *****

```

Figura 3.1: Ejemplo de preámbulo

- **Signos especiales:** Algunos símbolos no aparecen en el pdf tal cual los escribimos en el editor en el *.tex*, y hay que escribirlos de manera especial, como los signos de apertura de interrogación y exclamación (¿ y ¡) que habría que escribir ? y ! seguido de un acento grave para que se visualicen correctamente; o las comillas que, tanto para simples como para dobles, para abrirlas hay que escribir acento grave y para cerrarlas comilla simple.
- **Comandos para formato de texto:** hay comandos para títulos y secciones (`\chapter{}`, `\section{}`, `\subsection{}`...) y para enfatizar o cambiar el formato del texto (`\textbf{}`, `\textit{}`, `\TEXTSC{}`, `\textsf{}`, `\textsl{}`, `\texttt{}`...).
- **Índice:** L^AT_EX genera el índice del documento automáticamente si añadimos la instrucción `\tableofcontents`.

Se puede ver un ejemplo del cuerpo del documento en la Figura 3.2.

```

22 \begin{document}
23 \maketitle
24 \tableofcontents
25
26 \section{Name_Section}
27 \subsection{Name_Subsection}
28 Esto es un \textbf{ejemplo} del cuerpo de un documento \LaTeX.
29
30
31 \end{document}

```

Figura 3.2: Ejemplo de cuerpo del documento

3.2.2. XeTeX

Aunque \LaTeX proporciona un amplio conjunto de fuentes, es posible que se desee utilizar una fuente externa que nos guste más, y que esté instalada previamente en el sistema. Esto es posible gracias a XeTeX [3], que es un motor de tipografía \TeX que utiliza Unicode y es compatible con tecnologías de fuentes modernas como OpenType (OTF), TrueType (TTF), Graphite y Apple Advanced Typography (AAT). El compilador necesario es XeTeX o XeLaTeX.

En este proyecto, estamos compilando con **XeLaTeX**.

3.3. Google Forms

Google Forms [4] es una herramienta útil que permite planificar eventos, enviar una encuesta, hacer preguntas o recopilar otro tipo de información de forma fácil y sencilla.

Crear un formulario es tan fácil como añadir las preguntas seleccionando en cada una el tipo que mejor se adapte a nuestras necesidades. Los tipos de pregunta que nos podemos encontrar son, por ejemplo, texto, tipo test, elegir de listas, desplegables, etc. También permite añadir imágenes, vídeos o saltos de página para dividir la encuesta en secciones.

Una vez enviado el formulario a los destinatarios, se irán recopilando las respuestas recibidas. Éstas se pueden guardar en el propio Google Forms, de manera que se mostrarán como un resumen de las respuestas, o se pueden exportar a una hoja de cálculo en la que cada línea estarán las respuestas de cada persona (cada respuesta en una columna).

3.4. Google Sheets API

Google Sheets API [5] permite leer y modificar cualquier aspecto de una hoja de cálculo. La API ofrece dos formas principales de interactuar con la hoja de cálculo:

- Leer o escribir solo valores de celdas (con la colección *spreadsheets.values*).
- Leer o escribir cualquier aspecto de la hoja de cálculo (con la colección *spreadsheets*).

En nuestro caso, es suficiente con la lectura de los valores de las celdas, por lo que usaremos la colección `spreadsheets.values`¹.

Todos los métodos de la API requieren un parámetro `spreadsheetId` que se utiliza para identificar la hoja de cálculo a la que se accederá. Este ID es el valor entre `/d/` y `/edit` de la URL de la hoja de cálculo. Esta URL también contiene el `sheetId`, que es el ID de la hoja concreta con la que queremos trabajar. Este valor aparece como el valor del parámetro `gid`. Por tanto, la URL de una hoja de cálculo tendrá el siguiente aspecto:

```
https://docs.google.com/spreadsheets/d/spreadsheetId/edit#gid=sheetId
```

Para indicar el rango de celdas que queremos leer (o escribir), habrá que escribir ese rango como parámetro del método utilizado, en notación A1.

Para utilizar esta herramienta en Python, tendremos que descargar la biblioteca de Google Client (`pip install -upgrade google-api-python-client`), así como las credenciales necesarias.

3.4.1. Método `spreadsheets.value.get`

Este método lee y devuelve los valores de un rango de celdas de una hoja de cálculo. Se encarga de enviar un HTTP Request con la siguiente forma:

```
GET https://sheets.googleapis.com/v4/spreadsheets/{spreadsheetId}/values/{range}
```

La respuesta de esta petición tendrá la siguiente estructura:

```
{
  "range": ...,
  "majorDimension": ...,
  "values": [
    ...
  ]
}
```

¹<https://forward2.herokuapp.com/developers/sheets/api/reference/rest/v4/spreadsheets.values?hl=es-419>

range es el rango de celdas que hemos extraído, *majorDimension* indica si se ha leído por filas o por columnas (por defecto lee por filas) y *values* es un array con el valor de las celdas: cada elemento del array se corresponde con una fila (o con una columna si *majorDimension* = COLUMNS) y, a su vez, cada fila se describe con otro array.

Para este caso, necesitaremos sólo el campo *values*.

3.4.2. OAuth 2.0

Las APIs de Google usan el protocolo OAuth 2.0 [6] para autenticación y autorización. Para utilizar una API utilizando OAuth 2.0 hay que seguir los siguientes pasos:

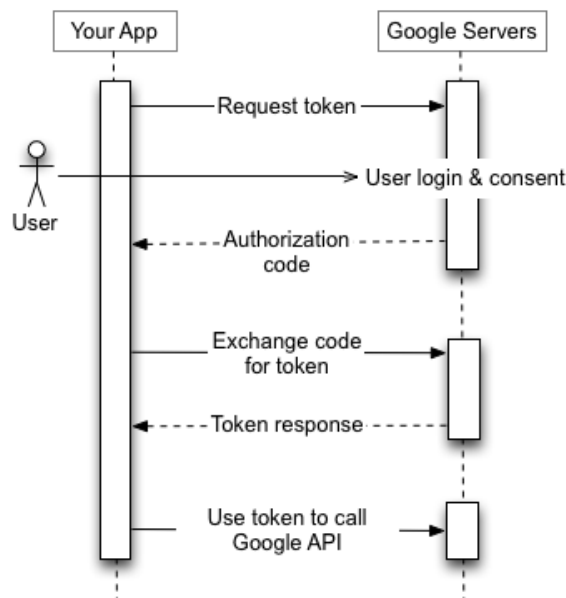


Figura 3.3: Esquema solicitud token para OAuth 2.0

1. **Obtener las credenciales de OAuth 2.0 desde la Consola de la API de Google**, que serán una ID y un *client secret* conocidos tanto por Google como por la aplicación. Estas credenciales hay que descargarlas en un JSON, cuyo nombre necesitará la aplicación para los pasos posteriores.
2. **Obtener un token de acceso del Servidor de Autorización de Google**. La aplicación realiza la solicitud del token al Servidor, que le devolverá el token preciso.

3. **Enviar el token de acceso a la API a la que se desea acceder.** Una vez que la aplicación obtiene un token de acceso, lo envía a la API en un encabezado de autorización HTTP.
4. **Actualiza el token de acceso, si es necesario.** Los tokens de acceso tienen vidas limitadas, por lo que se puede obtener un token de actualización, que permitirá a la aplicación obtener nuevos tokens de acceso.

3.5. Pycountry

pycountry [7] es un paquete que proporciona idioma, territorio, moneda y códigos de script de ISO de los países, así como sus traducciones entre ellos, tomados del paquete *iso-codes*.

Se puede acceder a los países a través de un objeto de base de datos que ya está configurado al importar *pycountry*: *countries*. Para acceder a un país específico, habrá que hacerlo con *get()* (*pycountry.countries.get(name=...)*).

3.6. Pillow

pillow [8] es un *fork* de PIL ². PIL es la Biblioteca de Imágenes de Python (Python Imaging Library) que agrega capacidades de procesamiento de imágenes a su intérprete de Python. Esta biblioteca es compatible con muchos formatos de archivo y proporciona un potente procesamiento de imágenes y capacidades gráficas.

En este proyecto se utiliza el método *Image* de PIL ³, que se utiliza para representar una imagen. Proporciona funciones como cargar imágenes desde archivos, o crear nuevas imágenes.

3.7. Pandas

Pandas [9] es una biblioteca de código abierto de Python que proporciona estructuras de datos y herramientas de análisis de datos de alto rendimiento. *Pandas* lleva a cabo tareas como

²<https://pypi.org/project/PIL/>

³<https://pillow.readthedocs.io/en/3.1.x/reference/Image.html>

alinearse datos para su comparación, fusionar conjuntos de datos o calcular tablas de frecuencias, entre otros.

La estructura de datos básica de pandas es el *DataFrame*, que es una estructura de datos tabular que se compone de columnas y filas ordenadas.

3.8. Matplotlib

Matplotlib [10] es una biblioteca de dibujo 2D de Python que produce figuras. Puede generar gráficos, histogramas, espectros de potencia, gráficos de barras, diagramas de errores, diagramas de dispersión, etc. con solo unas pocas líneas de código.

El módulo *pyplot* [11] proporciona una interfaz parecida a MATLAB ⁴. Para el usuario avanzado, tiene un control total de los estilos de línea, las propiedades de fuente, las propiedades de los ejes, etc. a través de una interfaz orientada a objetos o mediante un conjunto de funciones familiares para los usuarios de MATLAB.

⁴<https://es.mathworks.com/products/matlab.html>

Capítulo 4

Diseño e implementación

4.1. Arquitectura general

En la figura 4.1 se puede observar un esquema a modo de resumen del proceso que sigue el proyecto. A continuación, describiré cada bloque de este esquema.

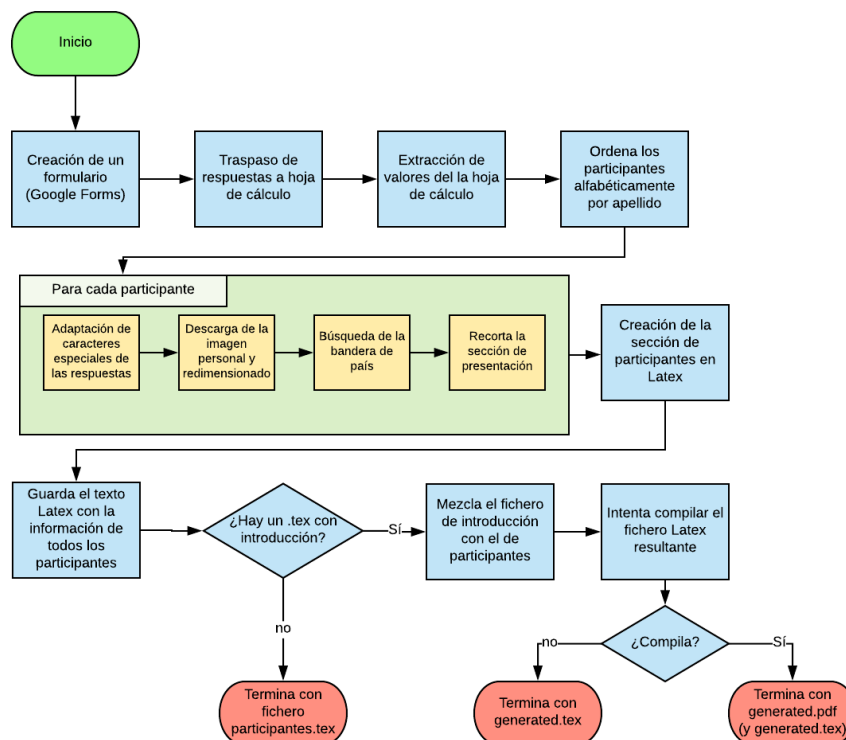


Figura 4.1: Esquema de funcionamiento

4.2. Creación de un formulario (Google Forms)

La base para obtener un buen resultado final, es crear desde el inicio un buen formulario. Para ello, hay que anticiparse a las respuestas de las personas que van a rellenar nuestra encuesta y, para esto, Google Forms nos ayuda mucho, ya que nos permite elegir entre múltiples tipos de respuestas posibles.

Por ejemplo, si preguntamos el nombre de la persona, elegiremos como tipo de respuesta la opción **Respuesta corta**, ya que el nombre consta de pocas palabras. En cambio, si pedimos una descripción, convendría elegir la opción **Párrafo** (incluso se puede limitar el número de caracteres de dicho párrafo). En los casos en los que conocemos de antemano las posibles respuestas de los asistentes, y éstas se vayan a repetir entre varios de ellos, podemos elegir las opciones de **Casillas de verificación**, **Selección múltiple** o **Desplegable**, según creamos conveniente. Algunos ejemplos de estos casos serían unas casillas de verificación para respuestas de Sí/No, o un desplegable para elegir la nacionalidad de la persona.

Con estos últimos casos, conseguiremos que todas las respuestas sean concisas y uniformes. Si no lo hacemos de esta forma y preguntamos, por ejemplo, “¿Cuál es tu mes de nacimiento?”, dos personas distintas que hayan nacido en el mismo mes pueden contestar “Yo nací en enero” y “Mi mes de nacimiento es enero”. Sin embargo, si aquí elegimos como tipo de respuesta un desplegable en el que simplemente haya que seleccionar el mes, obtendremos en ambas respuestas “Enero”. Esto nos puede ayudar en el futuro si queremos, por ejemplo, buscar a todas las personas que tengan la misma respuesta a una determinada pregunta.

Además de estas, hay varias opciones más para las respuestas, como se puede observar en la Figura 4.2.

4.3. Traspaso de respuestas a hoja de cálculo

Una vez que los participantes han rellenado el formulario, podemos acceder a sus respuestas de varias maneras: en la propia página de las preguntas, donde se pueden ver tanto las respues-

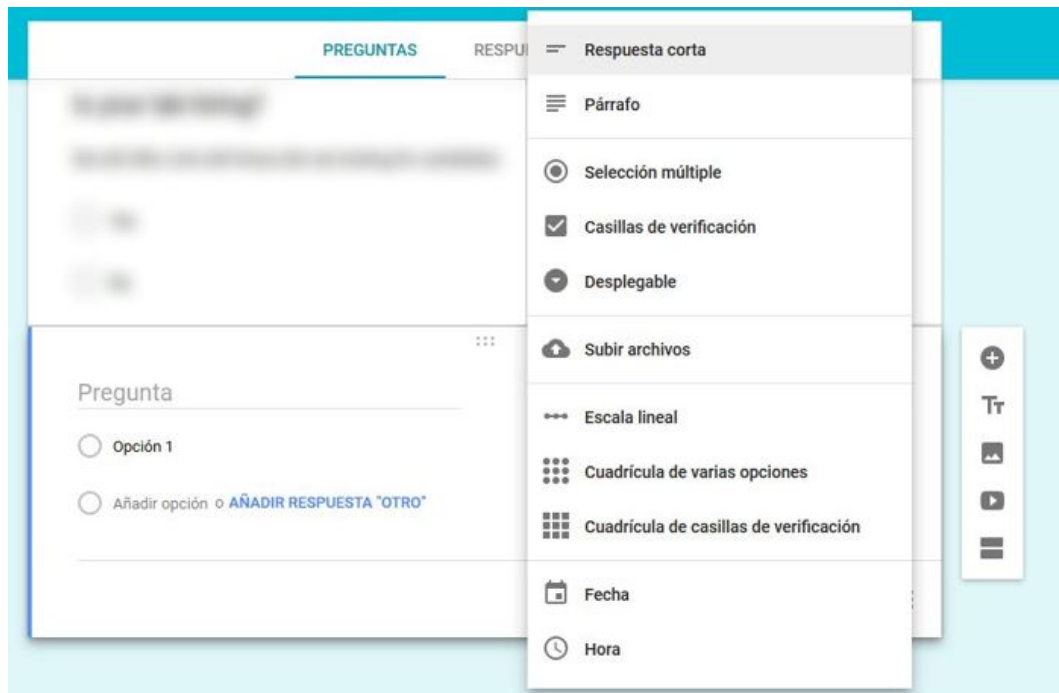


Figura 4.2: Tipos de respuesta de Google Forms

tas individuales como un resumen de las respuestas globales; y exportando las respuestas a una hoja de cálculo.

Realizaremos esta última opción, de manera que todos los datos quedarán volcados en una hoja de cálculo de Google. Esta hoja de cálculo tendrá la siguiente estructura:

- En la primera fila aparecen en cada celda las preguntas realizadas, a excepción de la primera celda, llamada “Timestamp”, que indicará la fecha y hora a la que ha contestado cada participante.
- En cada una de las siguientes filas aparecerá, bajo cada pregunta, la correspondiente respuesta de un participante.

4.4. Extracción de los valores de la hoja de cálculo

Esta es la primera parte que pertenece a mi proyecto. Lo primero que hace es obtener las credenciales necesarias para utilizar la API de Google Sheet. Una vez acreditado, con el *spreadsheetId* de la hoja de cálculo que hemos obtenido del formulario, procedemos a la ex-

tracción de todos los datos. En este caso, no necesitamos la primera fila, donde aparecen los enunciados de las preguntas, por lo que indicaremos en el programa que queremos la información a partir de la celda 'A2'.

El resultado de este proceso será un array en el que cada elemento será otro array. Es decir, cada fila de la hoja de cálculo, se convertirá en un array cuyos elementos serán cada respuesta (cada celda). Estas filas (arrays) se guardan a su vez en otro array, quedando este último como un array de filas. Esta descripción se puede entender mejor con el siguiente ejemplo:

	A	B	C
1	Marca temporal	Name	Position
2	18/05/2018 18:57:25	Gregorio Robles	Associate Professor
3	18/05/2018 22:15:12	Ahmed Zerouali	PhD Student
4	18/05/2018 22:18:46	Mathieu Nayrolles	I work in industry
5	18/05/2018 22:19:51	Foutse Khomh	Associate Professor

→

```
[ [18/05/2018 18:57:25, Gregorio Robles, Associate Professor, ...],
  [18/05/2018 22:15:12, Ahmed Zerouali, PhD Student, ...],
  [18/05/2018 22:18:46, Mathieu Nayrolles, I work in industry, ...],
  [18/05/2018 22:19:51, Foutse Khomh, Associate Professor, ...] ]
```

Figura 4.3: Ejemplo hoja de cálculo a array

4.4.1. Volcado de valores a NamedTuple

Una vez tenemos el array con todos los datos, convertiremos este *array de arrays* en un *array de NamedTuples*.

NamedTuple es una “*Tupla con nombre*”. Tiene una estructura parecida a la del array, pero cada elemento de éste se identifica con un nombre. Siguiendo el ejemplo anterior de la figura 4.3, el resultado sería:

```
[ Fields(date='18/05/2018 18:57:25', name='Gregorio Robles', position='Associate Professor', ...),
  Fields(date='18/05/2018 22:15:12', name='Ahmed Zerouali', position='PhD Student', ...),
  Fields(date='18/05/2018 22:18:46', name='Mathieu Nayrolles', position='I work in industry', ...),
  Fields(date='18/05/2018 22:19:51', name='Foutse Khomh', position='Associate Professor', ...) ]
```

Figura 4.4: Ejemplo NamedTuple

En este caso, el `NamedTuple` se llama *Fields*. Los nombres de los campos hay que definirlos antes de generar el `NamedTuple`. Una vez tenemos definido el `NamedTuple`, podemos referirnos a cada campo de éste con su nombre (si tuviésemos un array, tendríamos que saber en qué posición está el campo al que queremos referirnos para llamarlo). Por ejemplo, si del ejemplo anterior de la figura 4.4 tenemos guardado el primer `NamedTuple` en una variable llamada `row`, podremos acceder al nombre de la persona usando `row.name`.

4.5. Fichero de errores

Durante la ejecución del programa, puede haber algún aspecto que no se consiga realizar, y se pueda realizar manualmente. Por ello, es conveniente informar al usuario de qué cosas no ha podido realizar el programa.

Al final de la ejecución, se generará también un fichero llamado **errors.txt** en el que aparecerán todos los errores que han ocurrido en la ejecución, así como el proceso para solventarlo manualmente.

En las siguientes secciones, se desarrollará más sobre los errores que se pueden dar en cada parte del programa.

4.6. Proceso de tratamiento de información

Antes de generar el documento \LaTeX con la información que se ha extraído de la hoja de cálculo, hay que revisar toda esa información que tenemos, y ver si hay que hacer alguna modificación, o no usar literalmente el texto que tenemos en un campo.

Para esto, se utiliza un bucle que recorre los datos de todos los participantes. Este bucle tendría la siguiente estructura:

```
order_by_name

for row in fields:
    make_chars_latex_friendly
    download_image
    get_flag
    cut_presentation

text_Latex = ...
```

La variable *fields* es el array que contiene los datos de cada persona que ha rellenado la encuesta. Por tanto, en cada pasada del bucle la variable *row* se corresponde con un `NamedTuple`, por lo que podremos extraer los datos con los nombres de los campos (*row.name*, *row.affiliation...*).

Exceptuando el proceso de ordenar alfabéticamente que se hace una única vez antes de entrar al bucle, los demás procesos se llevan a cabo para cada persona. Una vez tratada toda la información que poseemos, se generará el texto en `LATEX`.

4.6.1. Ordenar alfabéticamente

Como en este caso tratamos con datos personales, hay que definir un orden de aparición, y para considerar a todos los participantes por igual, sin preferencias, ordenaremos a las personas alfabéticamente.

Para esto, en el método *order_by_name* se extrae el apellido de cada nombre, y se guardan en un array. Este array es ordenado alfabéticamente con el método de Python *sort*. Con la ayuda de un bucle y del array ordenado, se creará una nueva lista con todos los `NamedTuples` de datos ordenado alfabéticamente por el apellido del participante.

4.6.2. Convertir caracteres especiales

Como la intención de este proyecto es generar un documento en \LaTeX , hay que adaptarlo para ello. \LaTeX tiene una serie de caracteres que utiliza para funciones especiales, tales como $\$, $\$$, $\%$..., y para que al generar el PDF estos caracteres aparezcan correctamente, tienen que aparecer modificados en el documento \LaTeX .

Para solucionar este problema, se ha creado el método *make_chars_latex_friendly*, donde revisaremos todos los campos en busca de estos caracteres ($\$$, $\#$, $\%$, $\&$, $\^$, $_$, $\{$, $\}$, \sim , \backslash) y cuando los encuentre, añadirá delante de ellos un *backslash* (\backslash), o lo convertirá convenientemente para que aparezca el símbolo correctamente (ejemplo: \backslash = \backslashtextbackslash). De esta forma, aparecerá el carácter deseado, y no generará errores y confusiones a la hora de generar el PDF final.

Hay algunos casos especiales en los que no conviene hacer este proceso. Por ejemplo, si uno de los campos es una URL, y dicha URL no la vamos a escribir literalmente en el documento final, sino que la vamos a utilizar para obtener otro dato. En este caso tenemos, por ejemplo, un campo que corresponde a la URL de una imagen (se explicará a continuación), y esta URL será utilizada para descargar la imagen correspondiente, por tanto si hacemos el proceso anterior sobre la URL, cuando la utilicemos en el programa será incorrecta, por lo que no hay que modificarla.

4.6.3. Descarga de imágenes personales

Una de las preguntas de nuestro formulario es para que se incluya una imagen personal, para añadirla junto a la información de cada asistente. En esta parte, cada asistente incluye una URL donde se encuentra su imagen, que habrá que descargar para poder incluirla en nuestro documento.

El método *download_image* se encarga de esto. Lo primero que hace es comprobar si la URL es una imagen: abre la URL con la librería `urllib.request` (`resp = urllib.request.urlopen(url)`) y obtiene la información sobre ella (`resp.info()`), que suele tener la siguiente apariencia:

```

resp: Server: nginx/1.9.11
Date: Tue, 19 Jun 2018 16:48:45 GMT
Content-Type: image/jpeg;charset=UTF-8
Transfer-Encoding: chunked
Connection: close
Content-Disposition: attachment; filename="paola_accioly.jpg"
Cache-Control: max-age=604800
Expires: Tue, 26 Jun 2018 16:48:45 GMT
Strict-Transport-Security: max-age=31536000

```

De toda esta información, el campo que nos interesa es el *Content-Type*, ya que en él podremos comprobar si la URL es una imagen. Para extraer la información de este campo lo hacemos con `resp.info().get('Content-Type')`.

Una vez que comprueba que este campo empieza por “image”, se recoge también la extensión de la imagen, que aparece justo después de esta expresión (*imgType*). A continuación, se procede a descargar la imagen con `urllib.request.urlretrieve(url, nameImage)`. Esta expresión descarga la imagen de la URL que se indica en el primer parámetro de entrada, y la guarda con el nombre que aparece en el segundo parámetro.

El nombre de la imagen se compone de la ruta de la carpeta donde se va a guardar la imagen, seguido de la cadena “img” y un número identificador de cada persona, y acabando con la extensión que hemos almacenado anteriormente en *imgType*.

Normalmente, las extensiones de las imágenes suelen ser las comunes .jpg, .png, . . . , pero también pueden aparecer imágenes con extensión **.webp**. WebP es un formato de imagen creado por Google para reducir el tamaño de las imágenes y así disminuir el tiempo de carga de las mismas. Este formato no es reconocible por L^AT_EX, por lo que es necesario cambiar el formato de este tipo de imágenes, y para ello utilizamos `im = Image.open(nameImage).convert("RGB")`. Ahora tenemos la imagen en el formato que queremos en la variable *im*, por lo que hay que guardarla, por ejemplo, en .jpg, con `im.save(nameImage.jpg, "jpeg")`.

Además, se redimensionan las imágenes para que todas tengan el mismo tamaño y se genere un resultado uniforme.

En los casos en los que no se haya adjuntado URL de imagen, así como cuando haya algún problema para descargar la imagen (URL inexistente, URL no corresponde a una imagen, error en la descarga...), se asignará una imagen por defecto que indica que no poseemos la imagen de esa persona. Además, en estos casos, se incluirá un mensaje de error para que el encargado de generar este documento pueda revisar manualmente estos casos.

Los mensajes de error en estas situaciones podrían ser de la siguiente manera:

- Error in image of Alberto Bacchelli (pos. 4) ->It's impossible to download the image automatically. Please, try download it manually, and name it img4.jpg. Next, change the name of this person's image in the .tex
- Error in image of Safwat Hassan (pos. 26) ->This person hasn't provided image link.

Aquí, nos indica en la imagen de qué participante nos ha saltado el error y la posición del documento donde aparece. En el caso en que haya que descargar la imagen manualmente, hay que cambiar en el documento \LaTeX el nombre de la imagen correspondiente (ya que al no poder descargarse se ha adjuntado la imagen por defecto, llamada *img0.jpg*). Para seguir las pautas de todo el proceso, convendría guardar la imagen con el nombre que se indica en el mensaje de error.

Finalmente, el método devolverá el nombre con el que se ha guardado la imagen de la persona correspondiente o, en su defecto, la imagen de error.

4.6.4. Obtener banderas de nacionalidad

Cada persona tendrá que indicar también su país de procedencia, y en vez de poner literalmente el nombre del país, se incluye una bandera de éste. Para ello tenemos una carpeta con todas las banderas guardadas cuyos nombres son el nombre del país en código *ISO 3166-1 alpha-2*.

Para elegir la bandera correspondiente a cada persona, utilizamos el método *get_flag*, en el que se utiliza la librería *pycountry* para convertir el nombre del país a código *ISO 3166-1 alpha-2*. Para ello se obtiene la información de los códigos del país del cuál buscamos la bandera, con *pycountry.countries.get(name=nationality)*, y nos quedaremos con el código en alpha-2.

Por ejemplo, si buscamos España:

```
nac = pycountry.countries.get(name="Spain")
```

Obtendremos que *nac* es:

```
Country(alpha_2='ES', alpha_3='ESP', name='Spain', numeric='724',  
official_name='Kingdom of Spain')
```

Y nos interesaría quedarnos con el código alpha-2 en minúsculas:

```
nac.alpha_2.lower()
```

De esta forma, añadiendo la extensión y, si fuera necesario, el directorio en el que se almacenan las imágenes de las banderas, se generaría el nombre de imagen *flags/es.png*.

Al igual que con las imágenes, si ocurre algún error como que el país que buscamos no es correcto, se indicaría con un mensaje de error al usuario.

4.6.5. Acortar presentaciones

Una de las partes del formulario consiste en una descripción personal. Debido a que es una pregunta de respuesta libre, la extensión de estas respuestas puede variar mucho. Como buscamos un resultado ordenado, no podemos permitir que una presentación que ocupe demasiado desordene nuestro trabajo, al ocupar, por ejemplo, el espacio que debería ocupar la información de otra persona.

Por ello, en el método *cut_presentation*, fijamos una longitud máxima para este campo, y si el texto escrito por el participante sobrepasa este valor, lo cortaremos e indicaremos que ha sido recortado añadiendo puntos suspensivos al final.

4.7. Creación de la sección de participantes en L^AT_EX

Una vez tenemos todos los datos listos, procedemos a desarrollar el texto en L^AT_EX. El resultado final que buscamos tiene una estructura como la figura 4.5, y se consigue con el siguiente código L^AT_EX:



Figura 4.5: Estructura información participante

```
\begin{minipage}{0.3\textwidth}
  \centering
  \includegraphics[height=5cm]{images/img2.jpeg}
\end{minipage}
\hfill
\begin{minipage}{0.6\textwidth}\raggedright
  \color{color1}\uppercase{\textbf{Bram Adams}}
  \color{color2}\hspace{0.2cm}\includegraphics{flags/be.png}
  \hspace{0.1cm}{\scriptsize (PhD 2008)}
  \hspace{0.2cm}\textit{@mcis\_lab}\hspace{0.1cm}\includegraphics[height=0.5cm]
  {figs/hiring.png}\\
  Associate Professor at MCIS, Polytechnique Montreal\\
  {\footnotesize Heading the lab on Maintenance, Construction and Intelligence of
  Software (MCIS), which uses MSR techniques on a variety of research problems
  related to release engineering, human affect and the \ldots}\\
  \includegraphics[height=0.4cm]{figs/internet.png}\hspace{0.1cm}
  {\footnotesize \color{color1}\url{http://mcis.polymtl.ca/bram.html}}
\end{minipage}
```

4.8. Apartados opcionales

Opcionalmente, se pueden añadir dos apartados más al documento.

4.8.1. Diagramas de sectores

Con el método *generate_graph* podemos obtener diagramas de sectores que indiquen el porcentaje de repetición de las respuestas que deseemos. Para ello, le pasamos como argumento un array con los nombres de los campos de los que queremos que se genere el diagrama. Estos nombres de campos se corresponden con los que indicamos anteriormente para el *NamedTuple*.

4.8.2. Listados de personas

Usando el método *generate_list* se puede crear una lista de todas las personas que han dado una respuesta concreta a una pregunta. Como argumento a este método se pasará un array con todos los casos que queremos cubrir. Cada caso se indicará con un array de dos posiciones: pregunta de la que queremos hacer la lista, indicada con el nombre que se le ha asignado para el *NamedTuple*; y un *string* con la respuesta que buscamos.

4.9. Guardar el fichero \LaTeX con los datos

Cuando se ha formado el texto \LaTeX con los datos de todos los participantes, guardado en una variable, se procede a escribir este texto en un fichero *.tex* llamado **participants.tex**. En él aparecerá únicamente la sección de participantes. Si queremos incluir esta sección en otro documento \LaTeX (nos referiremos a él como *principal*) podremos hacerlo añadiendo en este último documento el comando `\input{participants}` en el lugar en el que queramos que aparezca esta sección.

Así, en el documento principal podremos añadir una portada, unas páginas de introducción, formatos para el texto, cabeceras o pies de página, etc.

Con esto, al compilar el documento principal, obtendremos un PDF con la unión del documento principal y el de los participantes, que es el resultado deseado.

4.10. Unir automáticamente documentos L^AT_EX

El proceso de la sección anterior también se puede automatizar, es decir, en vez de tener dos documentos y que uno llame al otro, podemos unir los `.tex`. Para esto, necesitamos tener listo el documento principal al que nos referíamos antes, y que en él aparezca el comando `\input{participants}`. Este documento deberá llamarse *intro.tex*

Nuestro programa buscará este fichero y leerá su contenido, buscando la sentencia `\input{participants}`, que será sustituida por el texto que hemos producido con los datos de los participantes. Una vez que tenemos la información de los participantes unida al texto del documento *intro.tex*, lo guardaremos en un nuevo fichero, llamado *ConfBook.tex*, quedando todo el contenido en un único documento L^AT_EX.

Si no se pueden unir los documentos, es decir, no se encuentra el fichero *intro.tex* o la sentencia donde se introduce la información de los participantes, se mostrará un mensaje indicando que se ha creado el fichero con la sección de los participantes, así como su nombre, listo para incluirse en otro `.tex`:

```
Participants section created (participants.tex). Include it in
your .tex
```

4.11. Compilado del documento L^AT_EX

En caso de que hayamos conseguido obtener el documento *ConfBook.tex*, y contemos en nuestro PC con el paquete de L^AT_EX instalado (en nuestro caso utilizamos XeLaTeX), el programa intentará compilarlo con el comando `xelatex ConfBook.tex` a través de una llamada al sistema:

```
os.system('xelatex ConfBook.tex')
```

Si se ha compilado correctamente, obtendremos por pantalla un mensaje informativo indicando que el PDF ha sido generado, y su nombre:

```
PDF has been generated ->ConfBook.pdf
```

En caso de que haya algún problema para compilar, nos mostrará un mensaje de error:
Impossible to generate PDF automatically. You must compile in Latex manually

4.12. Personalización

Antes de la ejecución del programa, tenemos que indicarle ciertos datos, como el *id* de la hoja de cálculo de la que queremos extraer los datos, el título que queremos dar a nuestro *book*, la imagen de la portada y del logo, etc.

Para ello, tenemos el fichero *settings.py*, en el que tendremos que completar todos estos datos:

- **SPREADSHEET_ID.** Id de la hoja de cálculo de la que queremos extraer los datos.
- **LONG_NAME.** Nombre completo de la conferencia.
- **SHORT_NAME.** Abreviatura del nombre de la conferencia.
- **PLACE.** Lugar de realización.
- **DATES.** Fechas de realización.
- **FRONT_IMAGE.** Imagen de portada.
- **FRONT_IMAGE_WIDTH.** Medida del ancho de la imagen de portada.
- **LOGO.** Imagen del logo.
- **LOGO_HEIGHT.** Medida del alto de la imagen del logo.
- **HEADER.** Array con los nombres asignados a cada campo del NamedTuple, y con los que se trabajará en el programa. Cada nombre corresponde, por orden, a cada columna (pregunta) de la hoja de cálculo. En este array, deberán aparecer siempre una serie de nombres, sin importar el orden. Estos nombres son: *name*, *position*, *affiliation*, *nationality*,

graduation, picture, homepage, twitter, presentation, looking y *hiring*. Además de estos, se pueden añadir cualesquiera.

- **GENERATE_GRAPHICS**. Indica si se quieren generar gráficas de estadísticas. Será *True* o *False*.
- **GRAPHICS_TO_GENERATE**. En el caso de que queramos que se generen los gráficos, aquí se indicará de qué apartados los queremos generar. Se asignarán con un array cuyos elementos son los nombres de cada campo del que queremos el gráfico.
- **GENERATE_LISTS**. Indica si se quieren generar listas de personas que hayan contestado de manera concreta a una pregunta. Será *True* o *False*.
- **LISTS_TO_GENERATE**. En el caso de que queramos que se generen las listas, aquí se indicará el apartado del que queremos la lista, y la respuesta que nos interesa. Se hará con un array en el que cada elemento es otro array de dos posiciones: nombre del campo y respuesta que buscamos.

Los datos que incluyamos aquí serán introducidos en la plantilla llamada *defs.tpl*, que contiene toda la información sobre la portada, los pies de página, el formato. . . , y toda esta información será adjuntada en nuestro *.tex* final.

Para que esto ocurra, debemos tener el fichero *intro.tex*, y en él debe aparecer la sentencia `\include{defs}` (al igual que hacíamos con la sección de participantes y la sentencia `\input{participants}`).

4.12.1. Personalización avanzada

Si queremos modificar el formato que se asigna por defecto, tenemos dos opciones:

- Acceder al fichero final tras la ejecución, y modificarlo.
- Antes de ejecutar, crear un nuevo fichero con la nueva personalización, llamado con un nombre distinto a *defs*, y crear el documento *intro.tex* con la sentencia `\include{NOMBRE_FICHERO}`. De esta forma, cuando ejecutemos no incluirá la plantilla por defecto, sino el fichero personalizado.

Capítulo 5

Resultados

La mejor manera de mostrar los resultados en este caso, es con ejemplos. Esta aplicación ya ha sido puesta a prueba en 3 ocasiones, en escenarios reales.

Aunque se muestran los ejemplos completos en el Apéndice B, las siguientes imágenes muestran algunos fragmentos del resultado final.

- La primera prueba se realizó para el *15th International Conference on Mining Software Repositories (MSR) 2018*¹, celebrada el 28 y 29 de mayo de 2018 en Gotemburgo (Suecia), y el resultado fue el mostrado en la figura 5.1.
- La segunda ocasión en la que se ha utilizado la aplicación ha sido para el *14th International Conference on Open Source Systems (OSS) 2018*²³, celebrada del 8 al 10 de junio de 2018 en Atenas (Grecia). Este caso se muestra en la figura 5.2.
- El último caso en el que se ha usado es en el *11th Seminar on Advanced Techniques & Tools for Software Evolution SATToSE 2018*, celebrado en Atenas (Grecia) del 4 al 6 de julio de 2018. Este ejemplo es el mostrado en el ejemplo 5.3.

¹<https://conf.researchr.org/home/msr-2018>

²<https://www.oss2018.org/>

³<http://sattose.org/2018>

PARTICIPANTS



PAOLA ACCIOLY 🇧🇷 (PhD 2018) @paola_accioly 📧

Associate Professor at Universidade Federal de Pernambuco
Just finished my PhD in software engineering and started teaching at the
Universidade Federal de Pernambuco, Brasil.

<https://github.com/prga>



BRAM ADAMS 🇨🇦 (PhD 2008) @mcis_lab 📧

Associate Professor at MCIS, Polytechnique Montreal
Heading the lab on Maintenance, Construction and Intelligence of
Software (MCIS), which uses MSR techniques on a variety of research
problems related to release engineering, human affect and the analysis of
multimedia repositories. Likes a nice bottle of water, and is fully convinced
that the MSR world needs more Perl programmers and Emacs users. Oh,
and don't be afraid to submit a paper to MSR 2019, Montreal has no snow
in May (with a significance level of 0.05 ;-)).

<http://mcis.polymtl.ca/bram.html>



AMRITANSHU AGRAWAL 🇮🇳 (PhD 2019) @amritbhanu 📧

PhD Student at North Carolina State University

I am a 4th year PhD student in Computer science at NC State University,
specializing in machine learning & software engineering. I have worked on
multiple problems, like topic modeling, class imbalance, GitHub code
development practices, and others. I am currently working as data scientist
intern at Lucidworks Inc, have worked at IBM US before, and also research
assistant intern at Durham University, UK. I am author to many published
papers which you can read up on my website: www.amritanshu.us



ALBERTO BACCHELLI 🇮🇹 (PhD 2013) @sback_ 📧

SNSF Professor at University of Zurich

My research field is empirical software engineering and I have been in the
MSR community since the start of my PhD in 2009 at the University of
Lugano with Prof. M. Lanza. Now I am heading ZEST (Zurich Empirical
Software engineering Team) at the University of Zurich, and I have the
pleasure of supervising 4 (soon to be 6!) amazing PhD students (V.
Kovalenko, L. Pascarella, D. Spadini, A.A. Sawant) and collaborating with a
fantastic postdoc (F. Palomba). We are all attending MSR 2018!

<http://www.ifl.uzh.ch/en/zest/team/bacchelli.html>



Figura 5.1: Página ConfBook MSR 2018

PARTICIPANTS



KULJIT KAUR CHahal  (PhD 2011) @kuljitkahal2 

Associate Professor at Dept of Computer science, Guru Nanak Dev university, Amritsar, India

Interested in research and teaching both. Topics of interest - open source software, and distributed systems.



KHALID ELBAAMRANI 

Associate Professor at Cadi Ayyad University

Khalid El Baamrani received his PhD in Telecommunication Engineering from the University of Cadi Ayyad, Marrakech, Morocco in 2005. He is presently working as an Associate Professor in the ENSA of Marrakech, University of Cadi Ayyad, Marrakech, Morocco. His research interests include packet switching architectures/algorithms, networked systems, software-defined networks



MARCO A. GEROSA  @gerosa_marco 

Associate Professor at Northern Arizona University (NAU)

I am an Associate Professor at the Northern Arizona University (NAU), USA. I am also a member of the Computer Science Graduate Program at the University of São Paulo (USP), Brazil. I research the intersections of Software Engineering (SE) and Social Computing (CSCW). Currently, I am working on the analysis and support of OSS communities, focusing on newcomers and casual contributors.

<https://www.imc.usp.br/~gerosa>



JESUS M. GONZALEZ-BARAHONA  @jgbarah

Professor at Rey Juan Carlos University

I teach and research in Universidad Rey Juan Carlos, Fuenlabrada (Spain). My research interests include the study of software development communities and processes, with a focus on quantitative analysis. I'm one of the founders of Bitergia, the software development analytics company. I like coding, taking photos, and drinking coffee.

<https://gysc.urjc.es/jgb>



Figura 5.2: Página ConfBook OSS 2018

PARTICIPANTS



MICHEL CHAUDRON   

Professor at Chalmers | Gothenburg University

I am Full Professor at the Software Engineering division part of the joint Dept of Computer Science of Chalmers and Gothenburg Univ. (SE). I obtained my PhD in the area of formal methods and programming calculi for parallel computing. My research interests are in software architecture, software design, software modeling with a special focus on UML, software composition. I have an interest in empirical studies in software engineering, preferably industrial industrial context.

http://dblp.uni-trier.de/pers/hd/c/Chaudron:Michel_R=_V=



ALEXANDRE DECAN  (PhD 2013) 

Postdoc at University of Mons, Belgium

Alexandre Decan obtained a Ph.D. degree in 2013 for the thesis entitled "Certain Query Answering in First-Order Languages". He is now postdoctoral researcher at the Software Engineering Lab of the University of Mons, where he has been actively involved in several research projects and has co-authored many publications related to the maintenance and evolution of software ecosystems.

<https://decan.lexpage.net>

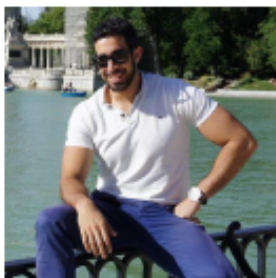


VASILIKI EFSTATHIOU 

Postdoc at Athens University of Economics and Business

I am a Postdoc at the Athens University of Economics and Business where I mainly do research on mining natural language artefacts from software repositories. My main focus is on mining developer communication data; besides this, my research interests lie in the wider area of knowledge representation, reasoning and natural language processing.

<https://www.balab.aueb.gr/vasiliki-efstathiou.html>



STEFANOS GEORGIU  (PhD 2019) @StefanosGeorgi1

PhD Student at Athens University of Economics and Business

In love with programming, sporting, and travelling. Addicted to good quality food but with no quantity limitations. Performing research on Energy Efficiency with focus on Software Engineering.

https://stefanos1316.github.io/my_portfolio/home.html

Figura 5.3: Página ConfBook SATToSE 2018

Capítulo 6

Conclusiones

6.1. Consecución de objetivos

Al principio de esta memoria enumeré una serie de objetivos que pretendía conseguir a lo largo del proyecto, los cuales se analizan a continuación para comprobar si se han conseguido o no.

- **Usar la aplicación para casos reales.** Como se ha indicado anteriormente en la sección de Resultados, esta aplicación se ha utilizado en tres casos reales: *MSR*, *OSS* y *SATT0SE*.
- **Sencillez de uso.** Para personas familiarizadas con este campo, así como las tecnologías utilizadas, resultará fácil ejecutar el programa para la obtención del *book*. Sin embargo, en el caso de las personas que no tengan ninguna relación con las tecnologías que se han utilizado, podría resultar difícil la ejecución.

6.2. Aplicación de lo aprendido

A lo largo de la carrera he adquirido muchas habilidades y aprendizajes en diferentes áreas, siempre relacionadas con mi carrera de Ingeniería de Tecnologías de la Telecomunicación. De todas ellas, las que he podido aplicar en el desarrollo de este proyecto han sido:

1. **La programación.** Aunque para mí era totalmente desconocida antes de llegar a la universidad, decidí hacer este Trabajo de Fin de Grado totalmente relacionado con ella, ya

que ha sido lo que más me ha interesado de todo el grado, y a lo que me gustaría dedicarme. Las asignaturas que me han introducido a la programación y me han enseñado programación a bajo nivel han sido *Fundamentos de la Programación*, *Programación de Sistemas de Telecomunicación* y *Sistemas Operativos*.

La programación a más alto nivel la he aprendido gracias a asignaturas como *Servicios y Aplicaciones Telemáticas* o *Ingeniería de Sistemas de Telecomunicación*.

- 2. Capacidad de trabajar bajo presión y enfrentamiento a los problemas.** A lo largo de la carrera nos enfrentamos a numerosos casos en los que tenemos que trabajar bajo presión, como en épocas de exámenes, en las que tenemos varios exámenes en muy pocos días, teniendo en nuestra mente la obligación de aprobarlos todos. Además, esta carrera me ha enseñado, y no en asignaturas, que debemos enfrentarnos a los problemas y buscar la manera de resolverlos.

6.3. Lecciones aprendidas

Una de las cosas más importantes que he aprendido con este trabajo ha sido planificar un proyecto de gran envergadura como es este, y organizarlo tanto temporal como estructuralmente.

También he adquirido conocimientos en aplicaciones y tecnologías antes desconocidas para mí, como pueden ser *Google Forms* o *Google Sheet API*, que me han sido de gran ayuda en este proyecto, ya que me han permitido poder trabajar con los datos de manera online, sin tener que descargar y tratar documentos en local.

También he podido aprender mucho más sobre el lenguaje \LaTeX , del cual sabía muy poco antes de este trabajo. Además de utilizarlo en el programa para generar el *book* final, lo he utilizado para realizar esta memoria, y me han sido muy útiles los conocimientos adquiridos para ello.

Por último, y no menos importante, la realización de este proyecto, que he llevado a cabo a lo largo del curso, lo que incluye las épocas de mayor carga de trabajo como son los exámenes,

unido a la realización de las prácticas curriculares, me ha enseñado a compaginar grandes cargas de trabajo para conseguir el mayor objetivo personal que tengo en este momento: la finalización del grado.

6.4. Trabajos futuros

Un posible trabajo futuro es crear una aplicación web en la que el gestor del seminario pueda crear de manera fácil un formulario y que cada persona lo rellene con sus datos. Así, automáticamente, se añadirían esos datos a los de los demás participantes y se generaría el pdf con la nueva incorporación.

También podría haber en esta misma web un apartado donde consultar de forma online los datos de los participantes que hayan rellenado el formulario, con enlaces a sus páginas personales.

De esta forma, se podría cumplir uno de los objetivos que no he conseguido cumplir en este proyecto: hacer que el proceso de generación del book lo puedan realizar personas que no tienen conocimientos sobre las tecnologías utilizadas, o no estén nada familiarizadas con este campo.

Otra idea es, además de la web, hacer una aplicación móvil, de manera que en vez de tener el documento en versión papel, se pueda tener toda la información en la app.

Bibliografía

- [1] Python. <https://www.python.org/>.
- [2] The LaTeX Project. <https://www.latex-project.org/>.
- [3] XeLaTeX - ShareLaTeX. <https://es.sharelatex.com/learn/XeLaTeX>.
- [4] Google Forms. <https://www.google.es/intl/es/forms/about/>.
- [5] Google Sheets API. <https://developers.google.com/sheets/guides/concepts?hl=es-419>.
- [6] Using OAuth 2.0 to Access Google APIs. <https://developers.google.com/identity/protocols/OAuth2?hl=es-419>.
- [7] Pycountry. <https://pypi.org/project/pycountry/>.
- [8] Pillow. <https://pypi.org/project/Pillow/>.
- [9] Pandas. <https://pandas.pydata.org/>.
- [10] Matplotlib. <https://matplotlib.org/>.
- [11] Pyplot tutorial. https://matplotlib.org/users/pyplot_tutorial.html.

Apéndice A

Manual de usuario

Para que la aplicación funcione correctamente, es necesario tener instalado *Python 3* o superior, además de los siguientes paquetes, en las versiones indicadas:

- *google-api-python-client*¹ versión 1.6.5 o superior.
- *pycountry*² versión 18.2.23 o superior.
- *pillow*³ versión 5.1.0 o superior.
- *oauth2client*⁴ versión 4.1.2 o superior.

Además, para conseguir el compilado en \LaTeX y que en la misma ejecución nos genere el documento en PDF, necesitamos tener instalado también *latex-xetex* en la versión 3 o superior.

Una vez instalado todo lo anterior, tenemos que ejecutar el programa. Para hacerlo desde una *Shell Unix* hay que ejecutar el comando `python3 creator.py`. Si todo va bien, al finalizar obtendremos en la carpeta de destino tanto los ficheros \LaTeX como el PDF con el resultado final.

¹<https://developers.google.com/api-client-library/python/start/installation>

²<https://pypi.org/project/pycountry/>

³<https://pypi.org/project/Pillow/>

⁴<https://pypi.org/project/oauth2client/>

A.1. Dónde descargar la aplicación

El programa está disponible en el repositorio de GitHub:

<https://github.com/SheilaOM/TFG>

Apéndice B

Ejemplo de uso

Como he indicado con anterioridad, este proyecto se ha utilizado en casos reales, y los resultados han sido los siguientes.

B.1. 15th International Conference on Mining Software Repositories (MSR) 2018

CONFERENCE YEARBOOK



15th International Conference on Mining Software Repositories
Gothenburg, Sweden, May 28th-29th 2018

PARTICIPANTS




PAOLA ACCIOLY 🇧🇷 (PhD 2018) @paola_accioly 

Associate Professor at Universidade Federal de Pernambuco
Just finished my PhD in software engineering and started teaching at the
Universidade Federal de Pernambuco, Brasil.

<https://github.com/prga>



BRAM ADAMS 🇧🇪 (PhD 2008) @mcis_lab 

Associate Professor at MCIS, Polytechnique Montreal
Heading the lab on Maintenance, Construction and Intelligence of
Software (MCIS), which uses MSR techniques on a variety of research
problems related to release engineering, human affect and the analysis of
multimedia repositories. Likes a nice bottle of water, and is fully convinced
that the MSR world needs more Perl programmers and Emacs users. Oh,
and don't be afraid to submit a paper to MSR 2019, Montreal has no snow
in May (with a significance level of 0.05 ;-)).

<http://mcis.polymtl.ca/bram.html>



AMRITANSHU AGRAWAL 🇮🇳 (PhD 2019) @amritbhanu 

PhD Student at North Carolina State University
I am a 4th year PhD student in Computer science at NC State University,
specializing in machine learning & software engineering. I have worked on
multiple problems, like topic modeling, class imbalance, GitHub code
development practices, and others. I am currently working as data scientist
intern at Lucidworks Inc, have worked at IBM US before, and also research
assistant intern at Durham University, UK. I am author to many published
papers which you can read up on my website www.amritanshu.us



ALBERTO BACCHELLI 🇮🇹 (PhD 2013) @sback_ 

SNSF Professor at University of Zurich
My research field is empirical software engineering and I have been in the
MSR community since the start of my PhD in 2009 at the University of
Lugano with Prof. M. Lanza. Now I am heading ZEST (Zurich Empirical
Software engineering Team) at the University of Zurich, and I have the
pleasure of supervising 4 (soon to be 6!) amazing PhD students (V.
Kovalenko, L. Pascarella, D. Spadini, A.A. Sawant) and collaborating with a
fantastic postdoc (F. Palomba). We are all attending MSR 2018!

<http://www.ifi.uzh.ch/en/zest/team/bacchelli.html>





SEBASTIAN BALTES  (PhD 2018) [@s_baltes](#) 

Researcher, PhD Student at University of Trier

In my research, I analyze software developers' work habits to identify requirements for new tools supporting developers; to build, evaluate, and compare new and existing tools; and to communicate developers' opinions, work habits, and needs to the software engineering research community. Most empirical studies I conduct follow a mixed-methods approach, combining qualitative and quantitative methods.

<http://empirical-software.engineering/>





JONATHAN BELL  (PhD 2016) [@_jon_bell_](#) 

Assistant Professor at George Mason University

Jon is an Assistant Professor directing research in Software Engineering and Software Systems at George Mason University. His research makes it easier for developers to create reliable software. Recently, his research has encompassed topics such as flaky test detection, security testing, and code coverage analysis.

<http://www.jonbell.net>



MORITZ BELLER  (PhD 2018) [@Inventitech](#) 

Postdoc, PhD Student at Delft University of Technology
Mr. WatchDog and TravisTorrent

<https://inventitech.com/>



ALEXANDRE BERGEL  (PhD 2005) [@AlexBergel](#) 

Associate Professor at University of Chile

Alexandre Bergel and his collaborators carry out research in software engineering. His effort is about designing tools and methodologies to improve the overall performance and internal quality of software systems, by employing profiling, visualization, and artificial intelligence techniques.

<http://bergel.eu>





COR-PAUL BEZEMER 🇳🇱 @corpaul 

Assistant Professor at University of Alberta

I will join the Electrical & Computer Engineering department of the University of Alberta as an assistant professor on July 1st. If you are interested in doing a Master's or PhD in the field of software engineering and/or performance engineering, don't hesitate to come talk to me!

<http://sailhome.cs.queensu.ca/~corpaul/>



TYSON BULMER 🇨🇦 @TysonBulmer 

New Graduate at University of Victoria

New graduate from the University of Victoria. Loves all things technology from web development to cryptocurrency. Pursued research under the supervision of Daniela Damian.

<http://tysonbulmer.ca>



PAOLO CALCIATI 🇮🇹 (PhD 2019)

PhD Student at IMDEA Software Institute

I am currently a PhD Student at IMDEA Software Institute in Madrid, working in the Software Engineering area, and in particular focusing on analysing the evolution of Android applications.



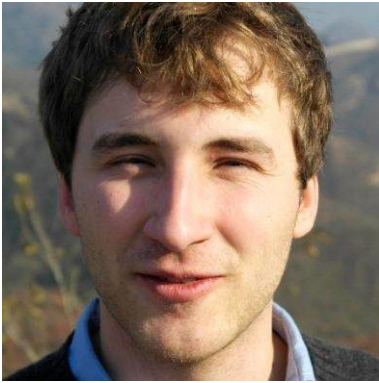
FRANCESC CAMPOY 🇪🇸 @francesc 

I work in industry at source{d}

I manage Developer Relations at source{d}, where we are working on industrializing ML on Source Code. I'm from Barcelona, but live in San Francisco now and if you want to know more you should just come ask. I'm a friendly guy!

<https://campoy.cat>





MAXIMILIAN CAPRARO 🇩🇪 (PhD 2019) @maxcapraro 

PhD Student at Friedrich-Alexander-Universität
Erlangen-Nürnberg

Maximilian Capraro is a doctoral student at the Open Source Research Group at Friedrich-Alexander-University Erlangen-Nürnberg. He received the BEng degree in information and communication technology from BA Eisenach and the MSc degree in computer science from FAU. Max worked on inner source in research projects with a variety of companies incl. Siemens and Continental. His interests include inner source, quantifying dev. collaboration, and software architecture, design and implementation.



MAËLICK CLAES 🇧🇪 (PhD 2016) @maelickclaes 

Postdoc at University of Oulu

I'm a postdoc at the University of Oulu, as part of the Empirical Software Engineering in Software, Systems and Services research unit (M3S). I got my PhD theses from the University of Mons in 2016 and its topic was maintainability issues in OSS package ecosystems. I am now conducting research as part the AutoTime project which goal is to determine the impact of time pressure on developers productivity and happiness.

<https://maelick.net>



ELENI CONSTANTINO 🇬🇷 (PhD 2015) @eleni_const

Postdoc at University of Mons

I am a postdoctoral researcher at the University of Mons. My research interests include software ecosystems, their evolution and software health. I am currently involved in two projects, SECOHealth and SECO-ASSIST, aiming to understand and measure software health within software ecosystems, as well as predict and recommend concrete solutions to problems related to ecosystem evolution.

<http://www.econst.eu/>



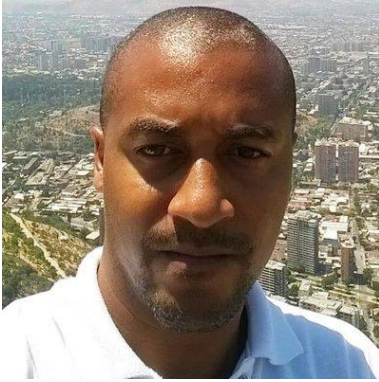
DANIEL ALENCAR DA COSTA 🇧🇷 (PhD 2017) 

Postdoc, I am also excited to start my faculty position as a Lecturer at the University of Adelaide thi year! :) at Queen's University

I like people, and many other things beyond research.

<http://danielalencar.github.io/>





GUILLERMO DE LA TORRE CÁRDENAS 🇨🇺 (PhD 2019)

@jwilliamtc

PhD Student at University of Chile

I got a major degree in Computer Science at Universidad de la Habana (Cuba) and more than 15 years of software development and leadership (at Centro de Neurociencias de Cuba (6 years), Striking Systems (1.5 year), and freelancing too). I am currently PhD student at Universidad de Chile (Chile) - expected end august, 2019; researching in source code change detection by exploiting more the underlying programming language (so far proving concepts in C#).



ALEXANDRE DECAN 🇧🇪 (PhD 2013)

Postdoc at University of Mons, Belgium

Alexandre Decan conducted his doctoral studies at the University of Mons on the subject of data quality in relational databases. He obtained a Ph.D. degree in 2013 for the thesis entitled "Certain Query Answering in First-Order Languages". He is postdoctoral researcher at the Software Engineering Lab of the University of Mons, where he has been actively involved in several research projects and has co-authored many publications related to the maintenance and evolution of software ecosystems.

<https://decan.lexpage.net>



ROBERT DYER 🇺🇸 (PhD 2013)

Associate Professor at Bowling Green State University

Want to mine lots of data easily? See here: <http://boa.cs.iastate.edu/>

<http://www.cs.bgsu.edu/rdyer/>



JOHAN FABRY 🇧🇪 (PhD 2005) @johanfabry

I work in industry at Raincode Labs

I work at Raincode Labs on a variety of projects across different languages. These projects aim to migrate client software to less niche or more recent languages. Hence I mainly do a lot of source code transformation. Until recently I was a software engineering researcher with an overall focus on easing programmers' jobs by providing them with the right abstractions for the task at hand.

<https://scholar.google.be/citations?user=Kw-9h4AAAAAJ&hl=en>





MARCO AURELIO GEROSA 🇧🇷 (PhD 2006) @gerosa_marco



Associate Professor at Northern Arizona University (NAU)

I research the intersections of Software Engineering (SE) and Social Computing (CSCW). Currently, I am working on the analysis and support of open-source software (OSS) development communities and collaborative systems in general.

<http://www.marcoagerosa.com/>



MIKE GODFREY 🇨🇦 @migod



Associate Professor at University of Waterloo

Old guy, used to do research sometimes.

<http://plg.uwaterloo.ca/~migod>



JESUS M. GONZALEZ-BARAHONA 🇪🇸 @jgbarah

Professor at Rey Juan Carlos University

I teach and research in Universidad Rey Juan Carlos, Fuenlabrada (Spain).

My research interests include the study of software development communities and processes, with a focus on quantitative analysis. I'm one of the founders of Bitergia, the software development analytics company. I like coding, taking photos and drinking coffee.

<https://gsyc.urjc.es/jgb>



GEORGIOS GOUSIOS 🇬🇷 (PhD 2009) @gousiosg



Assistant Professor at TU Delft

I am assistant professor at the Software Engineering group, Delft University of Technology. My research interests include software engineering, software analytics and programming languages. I works in the fields of distributed software development processes, software ecosystems, software quality, software testing, developer productivity assessment, research infrastructures and software security.

<http://gousios.org>





MICHAEL S HARVEY 🇨🇦 @MHarvito 🔍

I work in industry, Researcher, Software Engineering
Undergrad Student at McGill University

I'm interested in studying how the software development process can be improved for managers, users, and coders alike! I'm always looking for work opportunities to challenge my knowledge and practice my data analysis skills!

<https://linkedin.com/in/harvito>



SAFWAT HASSAN 🇸🇩 (PhD 2018)

PhD Student at Queen's University



CARINA HAUPT 🇩🇪 @caha42 📌 NOW HIRING

Researcher at German Aerospace Center (DLR)

Carina Haupt is the head of the Software Engineering group at the German Aerospace Center's (DLR) Simulation and Software Technology division. Her research fields include software engineering, open source, and knowledge management. Thereby her interest is in supporting domain scientist from all over the world in their work by creating software and process based solutions for their day to day problems.



SHINPEI HAYASHI 🇯🇵 (PhD 2008) @sh5i

Associate Professor at Tokyo Institute of Technology

Shinpei Hayashi is an associate professor of School of Computing, Tokyo Institute of Technology. His research interests include software maintenance and evolution, and software development environment, in particular building actionable techniques for supporting refactoring activities.

<https://www.se.cs.titech.ac.jp/~hayashi/>





STEFFEN HERBOLD 🇩🇪 @HerboldSteffen

Postdoc at University of Goettingen

I am a data scientist among software engineers, or vice versa. That is sometimes hard to distinguish.

<http://cs.ugoe.de/sherbold>



ABRAM HINDLE 🇨🇦 (PhD 2010) @abramh 

Associate Professor at University of Alberta

Abram Hindle is an associate professor of Computing Science at the University of Alberta. Abram works on problems relating to MSR, improving SE-oriented information retrieval with contextual information, the impact of software maintenance on software energy consumption (Green Mining), and how SE informs computer music—he makes terrible music. He has program co-chaired MSR and SCAM. Abram's PhD is from the University of Waterloo, and his MSc and BSc are from the University of Victoria.

<http://softwareprocess.ca/>



TRUONG HO-QUANG 🇻🇳 (PhD 2019) 

PhD Student at Gothenburg and Chalmers University

My research interests lay in software architecture and practices of using software architecture artifacts (e.g. UML models, software architecture documentation) in OSS. For that I have collected more than 93k+ UML models from 24k+ Github projects and made some observations/surveys on the way developers of those projects use UML models. Take a look here to keep up with recent findings: <http://oss.models-db.com>



AKINORI IHARA 🇯🇵 @redpig_ihara

Associate Professor at Wakayama University

Hi, this is Akinori Ihara who is a lecturer at Wakayama University in Japan. I'm just starting new position from this April. I'm interested in Empirical Software Engineering, Open Source Software Engineering, Social Software Engineering and Mining Software Repositories.

<http://www.wakayama-u.ac.jp/~ihara/>





KATSURO INOUE 🇯🇵 (PhD 1984) @katsuroinoue

Professor at Osaka University

Software Engineering Lab in Department of Computer Science, Osaka University.

<http://sel.ist.osaka-u.ac.jp/people/inoue/>



RIDHI JAIN 🇮🇳 (PhD 2020) 🔍

PhD Student at Indraprastha Institute of Information Technology, Delhi

I am a PhD student at Indraprastha Institute of Information Technology, Delhi, working under the supervision of Dr. Rahul Purandare. My research interests are in software binary comprehension and fault localization. My recent work is about a search system for mathematical expressions on software binaries. Currently, I am studying the impact of incorporating third party libraries on the stability of a multi-threaded application in weaker memory models.



JOHN JENKINS 🇺🇸 @JohnJen11057723

I work in industry, Master Student at Washington State University

I am a master's student focusing on developing program analysis tools, specifically tools using data-flow analysis techniques. I also work full-time as a software engineer for Schweitzer Engineering Laboratories.



XIANHAO JIN 🇨🇳 (PhD 2022) 📄

PhD Student at Virginia Tech

My name is Xianhao Jin and I am currently pursuing my Ph.D. degree in Computer Science at Department of Computer Science at Virginia Tech. My advisor is professor Francisco Servant and my research focuses on software development productivity and software quality.

<http://www.xianhaoj.com>





FOUTSE KHOMH 🇨🇦 (PhD 2010) @SWATLab 

Associate Professor at Polytechnique Montréal

Foutse Khomh is an associate professor at Polytechnique Montréal, where he heads the SWAT Lab on software analytics and cloud engineering research (<http://swat.polymtl.ca/>). His research interests include software maintenance and evolution, cloud engineering, service-centric software engineering, empirical software engineering, and software analytic. His work has received a ten-year Most Influential Paper (MIP) Award, three Best Paper Awards, and multiple nominations for Best paper Awards.

<http://swat.polymtl.ca/~foutsekh/>



DONGSUN KIM 🇰🇷 (PhD 2010) @darkrsw 

Researcher at University of Luxembourg

Dongsun Kim received the BEng, MS, and PhD degrees in computer science and engineering from Sogang University, Seoul, Korea, in 2003, 2005, and 2010, respectively. He is currently a research associate at the University of Luxembourg. His research interests include mining software repositories, automatic patch generation, static analysis, search-based software engineering (SBSE).

<https://www.darkrsw.net/>



JACQUES KLEIN 🇫🇷 (PhD 2006) @JacquesKlein14 

Associate Professor at University of Luxembourg

Jacques Klein is a Senior Research Scientist at the interdisciplinary centre for security, reliability and trust (SnT) of the University of Luxembourg. He received his PhD in Computer Science from University of Rennes 1 in 2006. His research interests include Security, mainly Mobile Security and Software Security; Analytics; Model-Driven Engineering and Model@runtime for Data Analytics; Repository Mining; Software Testing; Software Product Lines.

<https://sites.google.com/site/jacqueskleinwebpage/Home>



TAKASHI KOBAYASHI 🇯🇵 (PhD 2004)

Associate Professor at Tokyo Institute of Technology

Takashi KOBAYASHI is an associate professor in Department of Computer Science, School of Computing, Tokyo Institute of Technology. His research interests include software reuse, software design, software maintenance, program analysis, software configuration management, Web-services compositions, workflow, multimedia information retrieval and data mining.

<http://www.sa.cs.titech.ac.jp/~tkobaya/index.html.en>





DIMITRIS KOLOVOS 🇬🇷 @kolovos 

Professor at University of York

<https://www-users.cs.york.ac.uk/~dkolovos/>



KONSTANTIN KUZNETSOV 🇷🇺 (PhD 2018) 

PhD Student at Saarland University / CISPA

I am a PhD student in Software Engineering at Saarland University and a research assistant at CISPA. My advisor is Andreas Zeller. My research interests lie at the intersection of program analysis, machine learning and natural language processing. In my work I identify anomalies in the behavior of Android apps.

<https://cispa.saarland/people/kuznetsov/>



CHRISTOPH LAABER 🇸🇪 (PhD 2020) @ChristophLaaber

PhD Student at University of Zurich

I have been a PhD student since 2016 at University of Zurich in Harald Gall's group. Mainly supervised by Philipp Leitner, my main research interest is on performance testing in continuous-integration settings and cloud environments. In my work, I aim at bridging the gap between activities performed by software developers and performance engineers and therefore evolving performance-related tasks into first-class citizens of a programmer's work day.

<http://www.ifi.uzh.ch/en/seal/people/laaber.html>



MAXIME LAMOTHE 🇨🇦 (PhD 2020) 

PhD Student at Concordia University

I am currently in the second year of my PhD at Concordia University, doing research on API evolution and Migration.

https://users.encs.concordia.ca/~max_lam/





PHILIPP LEITNER 🇦🇹 @xLeitix 

Assistant Professor at Chalmers | University of Gothenburg
I am currently Assistant Professor of Software Engineering at Chalmers and University of Gothenburg, Sweden. I hold a PhD degree from Vienna University of Technology (in business informatics, focussing on cost-aware service composition). My current research interests are software performance engineering, especially in a cloud computing context.

<http://philippleitner.net>



DAYI LIN 🇨🇳 (PhD 2019) @lindayi45 

PhD Student at SAIL Lab, Queen's University, Canada
Dayi Lin is a Ph.D. student in the Software Analysis and Intelligence Lab (SAIL) at Queen's University, Canada. His research interests include mining software repositories and empirical software engineering, in particular, Software Engineering related aspects on PC games. His research aims at providing a better understanding of Software Engineering aspects on PC games, to help game developers produce games with better quality and player satisfaction.

<http://lindayi.me>



YUZHAN MA 🇨🇳 

I work in industry at Amazon.com Services, Inc.
I recently graduated from Washington State University, School of Computer Science and Electrical Engineering, with Master of Science degree in Computer Science. I performed research in the Software Engineering Lab under the supervision of Dr. Venera Arnaoudova. I received and accepted a position as Software Development Engineer at Amazon.

<https://yuzhanmaggiema.wordpress.com/>



VADIM MARKOVTSEV 🇷🇺 @vadimlearning 

I work in industry, Master Student at source{d}
Vadim is a Google Developer Expert in Machine Learning and a Lead Machine Learning Engineer at source{d} (sourced.tech) where he works with "big code" and natural languages. His academic background is compiler technologies and system programming. He is an open source zealot and an open data knight. Vadim worked at Samsung Research Center in Moscow, fought with email spam at Mail.Ru and also was a visiting associate professor at Moscow Institute of Physics and Technology.

<http://linkedin.com/in/vmarkovtsev>





FABIO MASSACCI  (PhD 1998) 

Professor at University of Trento, Italy

Chief Paper Shuffler at the University of Trento. Ph.D. in Computing from University of Rome La Sapienza. Been in Cambridge (UK), Toulouse (FR) and Siena (IT). Started in automated reasoning and verification for security protocols. Now on security requirements engineering, empirical methods for risk and security of FOSS. Participate to the FIRST Standard body on CVSS and was the European Coordinator of the multi-disciplinary research project SECONOMICS on socio-economic aspects of security.

<http://www.massacci.org>




FABIO MASSACCI  (PhD 1998)  

Professor at University of Trento

Fabio Massacci started in Rome, been in Cambridge, Toulouse and Siena. Now he is in Trento. Visited KULeuven, Durham Business School and USC. In 2015 he won the Ten Years Most Influential Paper by the IEEE RE Conference for security requirements engineering. He participates to the FIRST standard on Common Vulnerability Scoring System. Recently coordinated the SECONOMICS project on socio-economic aspects of security.

<http://disi.unitn.it/~massacci/>



SHANE MCINTOSH  (PhD 2015) *@shane_mcintosh* 

Professor at McGill University

Leader of the Software REBELs: <http://rebels.ece.mcgill.ca/>

<http://shanemcintosh.org/>



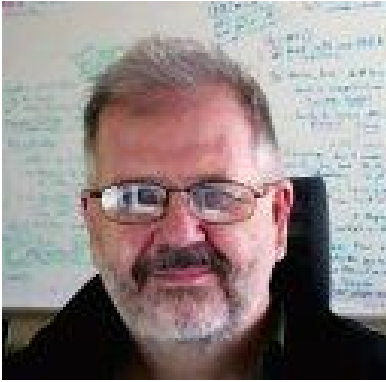
TOM MENS  *@tom_mens* 

Professor at Software Engineering Lab, University of Mons, Belgium

I am full professor at the University of Mons, directing the Software Engineering Lab since 2003. My main research interests are software evolution, empirical software engineering, software ecosystems, and model-driven software engineering. I have numerous publications, including two Springer books “Software Evolution” and “Evolving Software Systems”. I am involved in two interuniversity research projects around the topics of health and evolution of evolving software ecosystems

<http://staff.umons.ac.be/tom.mens/>






TIM MENZIES 🇺🇸 (PhD 1995) @timmenzies

Professor at NC State

Full prof, ex-nurse, rocketman, taxi-driver, journalist (it all made sense at the time).

<http://menzies.us>



ANDRE MEYER 🇨🇭 (PhD 2019) @casaout 

PhD Student at University of Zurich

I am a PhD student at the University of Zurich in Switzerland, and currently a visiting researcher at the University of British Columbia, Vancouver. I am interested in developer productivity and work, and in creating tools that foster productivity at work by using persuasive technologies such as self-monitoring and goal-setting. I've interned with Microsoft Research twice and also work in the information technology industry as an application developer and consultant.

<https://www.andre-meyer.ch>



LEANDRO L. MINKU 🇧🇷 (PhD 2010)

Lecturer (Assistant Professor) at University of Leicester

I am a Lecturer in Computer Science at the Department of Informatics, University of Leicester (UK). Prior to that, I was a research fellow at the University of Birmingham (UK). I received my PhD degree from the University of Birmingham (UK) in 2010. Among other roles, I am a steering committee member for PROMISE, an associate editor for the Journal of Systems and Software, a conference correspondent for IEEE Software, and was the co-chair for the challenge track of SSBSE'2015-2016.

<http://www.cs.le.ac.uk/people/l1m11>



SARAH NADI 🇨🇦 @sarahnadi 

Assistant Professor at University of Alberta

I, finally, decided on a name for my research group at the University of Alberta: the Software Maintenance and Reuse (SMR) group, pronounced as "summer" – pun intended for our "warm" Edmonton weather :-). SMR now has 5 awesome graduate students & thanks to funding from the Canada Research Chairs Program, I'm looking for new students to join our group. Come talk to me if you are interested in helping developers use SW libraries and APIs more effectively or in supporting software product ...

<http://sarahnadi.org/>





MEIYAPPAN (MEI) NAGAPPAN  (PhD 2011)

@MeiNagappan 

Assistant Professor at University of Waterloo

I am an Assistant Professor in the SWAG lab at David R. Cheriton School of Computer Science in the University of Waterloo. I am interested in the use of large-scale Software Engineering (SE) data to address the concerns of the various stakeholders (e.g., developers, operators, and managers).

<http://mei-nagappan.com/>



KANAGARAJ NACHIMUTHU NALLASAMY  

Master Student at Virginia Tech

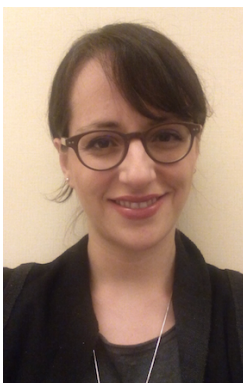
I love teaching, research of Software Engineering.



MATHIEU NAYROLLES  @mathieunls 

I work in industry at Ubisoft

I am Technical Architect dedicated to Research & Development on software quality and productivity at Ubisoft. I started several collaboration with other companies such as Mozilla and presented at various international scientific conferences such as SANER (Software Analysis, Evolution, and Reengineering), MSR (Mining Software repositories) or WCRE (Working Conference on Reverse Engineering). I also wrote several books on open-source technologies such as Angular, Solr or Magento.



NICOLE NOVIELLI  (PhD 2010) @NicoleNovielli

Assistant Professor at University of Bari, Italy

I am an Assistant Professor at the University of Bari. My expertise is in mining cognitive and affective states from natural language. I am a member of the COLLAB research group, whose research is in Software Engineering and Computer-Supported Cooperative Work, with focus on collaborative software development. I am the PI of the project 'EmoQuest: Investigating the Role of Emotions in Online Question & Answer Sites'.

<http://collab.di.uniba.it/nicole/>





DARIO DI NUCCI  (PhD 2018) @dardin88 

Postdoc at Vrije Universiteit Brussel

I obtained my PhD in Management and Information Technology at the University of Salerno, advised by Prof. Andrea De Lucia this year. Currently I am a research fellow at the Software Languages Lab of the Vrije Universiteit Brussel. My research interests are within the Software Engineering topic and include software maintenance and evolution, software testing, search based software engineering, green mining, mining software repositories, and empirical software engineering.

<https://dardin88.github.io/>



LINA OCHOA  (PhD 2021) @lina_8a

PhD Student at CWI

PhD student at the SWAT group. Working on code and metadata analysis.



SEBASTIANO PANICHELLA  (PhD 2014) @spanichella

Researcher, Postdoc at University of Zurich

Currently I'm a Research Associate at University of Zurich. Currently I'm mainly working on the interesting research problems that are collocated in the intersection of the following SE topics: Continuous Delivery, Continuous integration, Software maintenance and evolution, Mobile Computing, Summarization Techniques for Code, Changes and Testing. Another topic that is also of my interest is Code Review.

<http://www.ifi.uzh.ch/en/seal/people/panichella.html>



MIKE PAPADAKIS  

Researcher at Luxembourg University

Mike Papadakis is a research scientist at the Interdisciplinary Centre for Security, Reliability and Trust (SnT) at the University of Luxembourg. He received a Ph.D. diploma in Computer Science from the Athens University of Economics and Business. His research interests include software testing, static analysis, prediction modelling, mutation analysis and search-based software engineering.

<https://sites.google.com/site/mikepapadakis/home>





CHRIS PARNIN 🇺🇸 @chrisparnin 

Assistant Professor at North Carolina State University

There is no greater lens into the theory of computation than through the study of human cognition.

<http://chrisparnin.me>



MARTIN PINZGER 🇦🇹

Professor at Alpen-Adria-Universität Klagenfurt

Martin Pinzger is a Full Professor and Chair of Software Engineering at the Alpen-Adria-Universität Klagenfurt, Austria where he is the head of the Software Engineering Research Group and the Department of Informatics Systems.

<https://serg.aau.at/bin/view/MartinPinzger/WebHome>



SEBASTIAN PROKSCH 🇩🇪 @prksch 

Postdoc at University of Zurich, Switzerland

I'm trying to better understand developer needs and to support their daily work with tools that ease their lives. So far, I've been focussing my research on the development and the proper evaluation of code completion engines. However, I recently also started to be super excited about tools that go beyond the individual developer. For example, tools that facilitate team collaboration or that improve team awareness. – Come talk to me, I'd love to get to know you!

<http://proks.ch/>



GUILHERME B. DE PÁDUA 🇧🇷 @Gui_Padua_io 

I work in industry, Master Student at Concordia University, Montreal

<https://guipadua.github.io/>







NOAM RABBANI  @NoamRabbani

Undergraduate Student at McGill University

I am a front end web developer, online marketer and technology enthusiast. I also have interest in software engineering research.

<https://github.com/NoamRabbani>



AKOND RAHMAN  (PhD 2019) @akondrahman 

PhD Student at North Carolina State University

Fourth year PhD Student. Looking for research-based job opportunities in industry or academia. Looking forward to apply my internship and research experience to solve real-world software engineering research problems.

<https://akondrahman.github.io/>

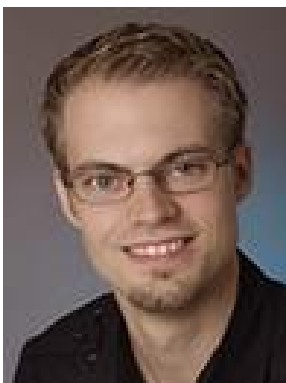


MD MASUDUR RAHMAN  (PhD 2020) @masud99r 

PhD Student at University of Virginia

I am a Ph.D. student in Computer Science department at the University of Virginia. I am also working as a Graduate Research Assistant. My supervisor is Baishakhi Ray. My research interest focuses on applying Information Retrieval, Natural Language Processing, and Machine Learning techniques on large-scale software repositories to analyze and automate software engineering systems.

<http://mmasud.me/>



MICHAEL RATH  (PhD 2020) 

Researcher, PhD Student at Technical University Ilmenau





GREGORIO ROBLES 🇪🇸 (PhD 2006) @gregoriorobles

Associate Professor at Universidad Rey Juan Carlos

I am Associate Professor at the Universidad Rey Juan Carlos, a public university in Madrid. I mainly do research in following two fields: Free/Open Source Software and Computational Thinking. I co-founded Bitergia, a software analytics company. I am getting old; I remember to have taken part in the first MSR back then in 2004 in Edinburgh!

<https://gsyc.urjc.es/~grex>



GEMA RODRIGUEZ-PEREZ 🇪🇸 (PhD 2018) @gerope90

PhD Student at TuDelft & TuEindhoven

I started my Ph.D. in 2015, in the LibreSoft Group at the University of Rey Juan Carlos, Spain. Then, at the beginning of 2018, I moved to the Netherlands to visit TuDelft and TuE in my last year of Ph.D. My research work focuses on understanding the Bug Seeding process. Specifically, I analyze the information from code review process and bug tracking systems to understand how defects are introduced and manifested in the source code.

<https://gemarodri.github.io/Presentacion/>



COEN DE ROOVER 🇧🇪 @oniroi

Professor at Vrije Universiteit Brussel

I work on the design of program analysis and transformation techniques, and their application in development tools that vary in scope from a single program, over the programs' development history, to the evolution of an entire ecosystem of programs over time. My group and I currently try to embrace the software ecosystem itself as a source of already-proven solutions for recommendations about library selection, library maintenance, and library migration.

<http://soft.vub.ac.be/~cderoove/>



GUENTHER RUHE 🇩🇪

Professor at University of Calgary

Guenther Ruhe is the Industrial Research Chair in Software Engineering at the University of Calgary. His research focuses on product release planning, app store mining, software project management, decision support, data analytics, empirical software engineering, and search-based software engineering. He's the editor in chief of Information and General Chair of RE'18. Contact him at ruhe@ucalgary.ca.

<http://ruhe.cpsc.ucalgary.ca/>





DAVIDE DI RUSCIO  (PhD 2007) @ddiruscio 

Researcher at University of L'Aquila (Italy)

Davide Di Ruscio is Assistant Professor at the DISIM department of the University of L'Aquila (Italy). Over the last decade, he has applied Model Driven Engineering (MDE) techniques in different application domains including service-based software systems, autonomous systems, and open source software (OSS). Currently, he is the technical director of the EU H2020 CROSSMINER project.

<http://www.di.univaq.it/diruscio/>



BARBARA RUSSO  

Associate Professor at Free University of Bolzano, Italy

My main research is on data mining in software and systems engineering. In particular, I am interested in discovering patterns in software evolution and in system behaviour that can pinpoint anomalies or specific characteristics of the SUTs.

<http://www.inf.unibz.it/~russo>

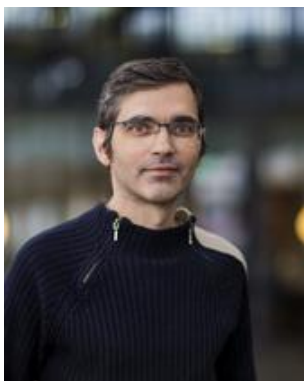


ANDREAS SCHREIBER  @onyame 

Researcher at German Aerospace Center (DLR)

Rocket & data scientist at the German Aerospace Center (DLR), head of the Department for Intelligent and Distributed Systems located in Berlin, Braunschweig and Cologne. Loves Python, dogs, and coffee!

<https://andreas-schreiber.net>



ALEXANDER SEREBRENİK  (PhD 2003) @aserebrenik

Associate Professor at Eindhoven University of Technology

I study how developers' communities change with time. In particular, I'm interested in understanding emotions of developers as well as how interaction between developers affects and is affected by development practices.

<http://www.win.tue.nl/~aserebre/>





FRANCISCO SERVANT @fjservant

Assistant Professor at Virginia Tech

My research focuses on software development productivity, software development expertise, and software quality. I use software evolution analysis and program analysis to create practical, efficient, and human-friendly techniques and tools that provide automatic support for all stages of software development. My research interests also include mining of software repositories, program comprehension, and software visualization.

<http://www.fservant.com>



WEIYI SHANG 🇨🇳 (PhD 2014)

Assistant Professor at Concordia University

I am an Assistant Professor and Concordia University Research Chair in Ultra-large-scale Systems at Concordia University. I am interested in mining various sources of data from both the development and operation of large software systems.

<https://users.encs.concordia.ca/~shang/>

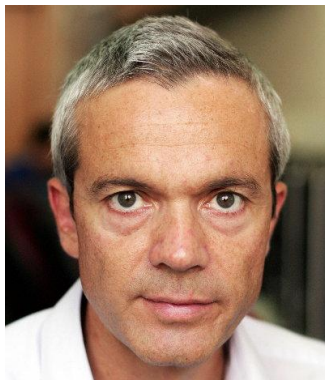


MAURICIO SOTO 🇨🇷 (PhD 2020)

PhD Student at Carnegie Mellon University

I am a Software Engineering PhD Student at the Institute for Software Research, School of Computer Science, at Carnegie Mellon University. I am advised by Dr. Claire Le Goues. I come from Costa Rica, the land of happiness, coffee and sloths. Before joining CMU, I graduated from the University of Costa Rica with a Bachelor's in Computer Science. I also worked in several companies and universities as a software engineer. I am a very happy person, I play the saxophone, and I love salsa.

<http://www.cs.cmu.edu/~msotogon/>



DIOMIDIS SPINELLIS 🇬🇷 @CoolSWEng

Professor at Athens University of Economics and Business

Diomidis Spinellis is a Professor and Head in the Department of Management Science and Technology at AUEB. He's the author of two award-winning books, Code Reading and Code Quality, as well as more than 300 papers. His most recent book is Effective Debugging. He's written the UMLGraph and CScout tools as well as code that ships with Apple's macOS and BSD Unix. He's a senior member of the ACM and the IEEE. From January 2015 he is serving as the Editor-in-Chief for IEEE Software.

<https://www.dmst.aueb.gr/dds/>





KATE STEWART 🇨🇦 @_kate_stewart

I work in industry at Linux Foundation

Kate Stewart is a Sr. Director of Strategic Programs at the Linux Foundation. With 30 years of experience in the software industry, she has held roles from hands-on developer thru to senior director. She has worked in Canada, Australia, and US. She set up CHAOSS, Real-Time Linux and Zephyr projects in LF, and contributes to SPDX, FOSSology, OpenChain as well as analysis of linux kernel licensing with cregit.

<https://www.linkedin.com/in/katestewartaustrin/>



MARGARET-ANNE (PEGGY) STOREY 🇨🇦 @margaretstorey



Professor at University of Victoria

I am a prof of Computer Science at the University of Victoria, Canada. I am also proud to hold the Lise Meitner Guest Professorship at Lund University in Sweden from 2016-2018. My goal is to improve developer tools and processes used in industry.

<http://margaretstorey.com/>



MATÚŠ SULÍR 🇸🇰 (PhD 2018) @MatusSulir

PhD Student at Technical University of Košice

I am a final (4th) year PhD student at the Department of Computers and Informatics, FEEI TU Košice. My research is focused on the integration of runtime data with source code to facilitate software understanding and maintenance. In the future, I would like to continue in this direction - particularly, I think concrete examples are the best way to explain abstract source code. If you share the same opinion, do not hesitate to contact me about collaboration.

<http://sulir.github.io>



HAILONG SUN 🇨🇳 @HailongSun  

Associate Professor at Beihang University

I am working on applying AI technologies to software engineering mainly through big data analysis.

<http://act.buaa.edu.cn/hsun>





ANDREAS SYMEONIDIS 🇬🇷 @asymeon 

Assistant Professor at Aristotle University of Thessaloniki

I am an Assistant Professor with the Department of Electrical and Computer Engineering at the Aristotle University of Thessaloniki. My research interests include Software Engineering (SE) Processes, Data Mining and Knowledge Extraction from Software repositories, Model-driven Engineering, Software Quality and Software Analytics. I have worked on many EU-funded projects (S-CASE, RAPP, Mobile Age, SEAF, CASSANDRA etc).

<http://users.auth.gr/symeonid>



SIMON URLI 🇫🇷 (PhD 2015) @Neo_nderthalis 

Postdoc at INRIA & University of Lille (France)

PhD in software product lines now working on software repair topic



ENRIQUE LARIOS VARGAS 🇨🇷

Researcher, Postdoc at Software Improvement Group

I am an enthusiastic software engineer who after my PhD studies is doing research in software analytics, mining software repositories and streaming analytics.

<https://www.linkedin.com/in/enrique-larios-vargas>



BOGDAN VASILESCU 🇷🇴 (PhD 2014) @b_vasilescu 

Assistant Professor at Carnegie Mellon University

I study how to make open-source communities more sustainable along multiple dimensions, from better tooling support to higher diversity and inclusiveness.

<https://cmustrudel.github.io>





CARMINE VASSALLO 🇮🇹 (PhD 2020) @ccvassallo

PhD Student at University of Zurich

Carmine Vassallo is a research assistant in the Software Evolution and Architecture Lab at University of Zurich. His main research goal is to support developers while dealing with Continuous Delivery breaks (e.g., CI build failures) and develop new approaches to prevent them.

<http://www.ifi.uzh.ch/en/seal/people/vassallo.html>



HAOYU WANG 🇨🇳 (PhD 2016)

Assistant Professor at Beijing University of Posts and Telecommunications

I'm an assistant professor of Beijing University of Posts and Telecommunications. My current research interest lies at the intersection of mobile system, privacy and security, and program analysis.

<http://sei.pku.edu.cn/~wanghy11/>



JIM WHITEHEAD 🇺🇸 @TheJimWhitehead

Professor at Univ. of California, Santa Cruz

Jim is Professor and Chair of Computational Media at the University of California, Santa Cruz. He directs the Software Introspection Laboratory and the Augmented Design Laboratory.

<https://users.soe.ucsc.edu/~ejw/>



DAVID WIDDER 🇺🇸 (PhD 2022) @davidthewid  


PhD Student at Carnegie Mellon University

Hi! I study tool use and choice in Open Source communities. I'm always looking for new collaborations and research internships!

<http://www.davidwidder.me>





ERIK WITTERN 🇩🇪 (PhD 2014) @erikwittern 

I work in industry at IBM Research

I work as a Research Staff Member for IBM at the T.J. Watson Research Lab. In the past few years, I was mostly concerned with software engineering-related aspects of consuming and providing web APIs. Lately, this research has been evolving around GraphQL.

<http://www.wittern.net/>



AIKO YAMASHITA 🇺🇸 @aikochama 

I work in industry, Associate Professor at DnB / Oslo Metropolitan University

Data Scientist at DnB (the largest financial institution in Norway) and adjunct associate professor at Oslo Metropolitan University

<https://www.hioa.no/eng/employee/aikyam>



NORIIHIRO YOSHIDA 🇯🇵 (PhD 2009) @NorihiroYoshida

Associate Professor at Nagoya University

I have been working for the Graduate School of Informatics at Nagoya University as an associate professor. I received my Ph.D. in Information Science and Technology from Osaka University under the supervision of Prof. Katsuro Inoue in 2009. Before joining Nagoya University in April 2014, I was an assistant professor in the Graduate School of Information Science at the Nara Institute of Science and Technology (NAIST) from June 2010 to March 2014.

<https://sites.google.com/site/yoshidaatnu/>



ANDY ZAIDMAN 🇧🇪 (PhD 2006) @azaidman

Associate Professor at Delft University of Technology

Andy Zaidman is an associate professor at Delft University of Technology, the Netherlands. His research interests include software testing, software evolution and software analytics. In 2013 he received a prestigious NWO Vidi career grant from NWO, the Dutch science foundation, for his work in the area of software testing and software evolution. In 2015 he gave a talk a TEDxDelft "Making Testing Fun".

<https://azaidman.github.io>






AHMED ZEROUALI  (PhD 2020) @a_zerou

PhD Student at University of Mons; Rey Juan Carlos
Universidad; Bitergia

Ahmed is currently working within the European SENECA project as a Software Researcher at Bitergia. He is a PhD candidate at the University of Mons/Universidad Rey Juan Carlos. The goal of his research is to empirically study how software libraries are used in open source projects, and to recommend improvements in such library usage by providing automated tools to support developers of software projects and software libraries.



THOMAS ZIMMERMANN (PhD 2008) @tomzimmermann 

I work in industry at Microsoft Research

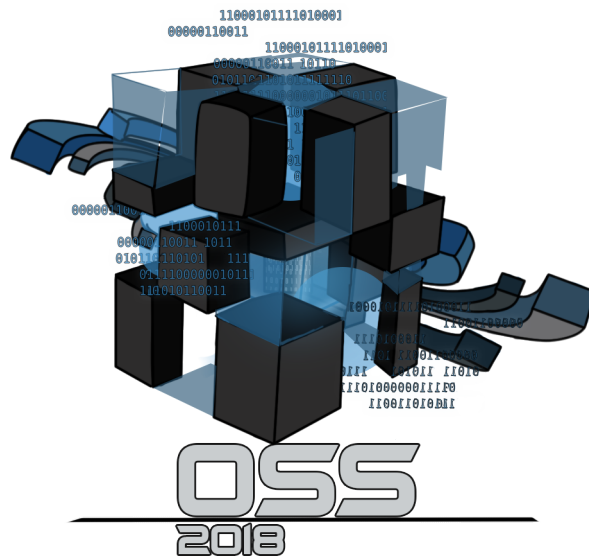
I'm currently working on productivity of software engineers and data scientists. Love to talk to you about this topic. I'm chair of the Steering Committee for MSR and Co-Editor Chief of the Empirical Software Engineering journal. I'm also running for Chair of SIGSOFT (please vote for me). Any ideas on how to make MSR, EMSE, or SIGSOFT better? Let me know.

<http://thomas-zimmermann.com/>



B.2. 14th International Conference on Open Source Systems (OSS) 2018

CONFERENCE YEARBOOK



The 14th International Conference on Open Source Systems

Athens, Greece, June 8th–10th 2018

www.oss2018.org

PARTICIPANTS



KULJIT KAUR CHAHAL  (PhD 2011) @kuljitcahal2 

Associate Professor at Dept of Computer science, Guru Nanak Dev university, Amritsar, India

Interested in research and teaching both. Topics of interest - open source software, and distributed systems.



KHALID ELBAAMRANI 

Associate Professor at Cadi Ayyad University

Khalid El Baamrani received his PhD in Telecommunication Engineering from the University of Cadi Ayyad, Marrakech, Morocco in 2005. He is presently working as an Associate Professor in the ENSA of Marrakech, University of Cadi Ayyad, Marrakech, Morocco. His research interests include packet switching architectures/algorithms, networked systems, software-defined networks



MARCO A. GEROSA  @gerosa_marco 

Associate Professor at Northern Arizona University (NAU)

I am an Associate Professor at the Northern Arizona University (NAU), USA. I am also a member of the Computer Science Graduate Program at the University of São Paulo (USP), Brazil. I research the intersections of Software Engineering (SE) and Social Computing (CSCW). Currently, I am working on the analysis and support of OSS communities, focusing on newcomers and casual contributors.

<https://www.ime.usp.br/~gerosa>



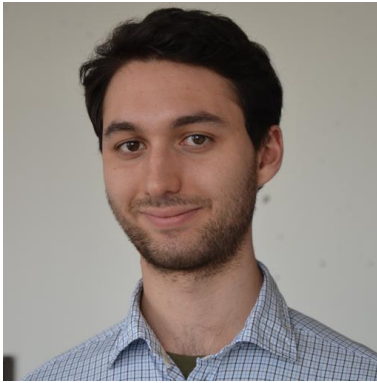
JESUS M. GONZALEZ-BARAHONA  @jgbarah

Professor at Rey Juan Carlos University

I teach and research in Universidad Rey Juan Carlos, Fuenlabrada (Spain). My research interests include the study of software development communities and processes, with a focus on quantitative analysis. I'm one of the founders of Bitergia, the software development analytics company. I like coding, taking photos, and drinking coffee.

<https://gsyc.urjc.es/jgb>





NIKOLAY HARUTYUNYAN 🇷🇺 (PhD 2019) 

PhD Student at Friedrich-Alexander University
Erlangen-Nürnberg

Nikolay Harutyunyan is a researcher at the Professorship for Open Source Software at Friedrich-Alexander University Erlangen-Nürnberg. Nikolay's research focuses on open source governance and supply chain management. Before joining the research group of the Professorship for Open Source Software, he studied Economics, Management, and Information Systems. You can reach him at nikolay.harutyunyan@fau.de.

<http://osr.cs.fau.de/people/members/harutyunyan>



PATRICIA ILIN 🇸🇪 

Master Student at Gothenburg University

Patricia is a MSc student in the Digital Leadership program of Gothenburg University. She has a keen interest in emerging technologies, startups and novel ways of funding niche products. When she is not busy with her studies she volunteers in Food2Change to fight food waste from grocery stores as well as is the chairwoman of DataTjej, an organization to attract more women into IT.

<https://www.linkedin.com/in/ilinpatricia>



STAVROS KOLONIARIS 🇬🇷

Master Student at Harokopio University of Athens

Holding a degree in "Computer Science" from Hellenic Open University and a MSc in "Advanced Information Systems in Business" from Harokopio University. Employed for over 10 years in the computer industry, mostly as IT supporter, and since 2008 in the Municipality of Kallithea, Greece. Main interests include Governmental Digitization and Transformation, Digital Citizenship and Governance of Smart Cities and Regions.

<http://www.e-tic.gr>



ELIA KOUZARI 🇨🇾 (PhD 2019)

I work in industry, PhD Student, Computer Lecturer at Dept. of Informatics, Aristotle Univ. of Thessaloniki

I am currently a PhD candidate in Software Engineering at Aristotle University of Thessaloniki focusing on process mining. Since 2011, I have been teaching in the Higher Education sector in Cyprus. I have been working as a Business Analyst at a Statistics and Informatics firm since 2016 applying my gained knowledge in the industry sector. Research interests include: OSS, agile practices and process mining.

<https://www.linkedin.com/in/eliakouzari/>





APOSTOLOS KRITIKOS 🇬🇷 (PhD 2021) @akritiko 

I work in industry, Researcher, PhD Student at Aristotle University of Thessaloniki

I am a researcher and software engineer from Greece. I enjoy working with open technologies. I do research on the fields of Open Source Software Engineering and Software Resilience with the Informatics Dept., AUTH. In 2014, I co-founded Social Mind, a 360 degrees digital marketing agency, with the aim of designing affordable, high-quality software and digital marketing solutions for the Web.

<http://softwareseilience.com>



PAULO MEIRELLES 🇧🇷 (PhD 2013) @paulormm

Professor, Postdoc at Federal University of São Paulo

Paulo Meirelles is a Professor at the Federal University of São Paulo (UNIFESP) and researcher at the Free/Libre/Open Source Software Competence Center (CCSL) at the Institute of Mathematics and Statistics of the University of São Paulo (IME-USP), in Brazil. His research interests include Free Software development, Agile Software Development, and Source code metrics. Contact him at paulo.meirelles@unifesp.br.

<http://softwarelivre.org/paulormm>



NIKOLAOS PAPANPYROU 🇬🇷 (PhD 1998) @nickie_42

Associate Professor at National Technical University of Athens

Associate professor at the School of Electrical and Computer Engineering of the National Technical University of Athens and a member of the Software Engineering Laboratory. I am interested in the theory and implementation of programming languages: semantics, type systems, compilers, static analysis, formal verification.

<https://www.softlab.ntua.gr/~nickie/>



DIMITRIOS PLATIS 🇬🇷 @PlatisSolutions 

I work in industry at Aptiv PLC

Dimitrios is a software engineer at Aptiv in Gothenburg, Sweden. During the day he is working on bringing Android to the automotive industry. In the evenings, he develops open source software and hardware projects that he likes to blog about. He is enthused about robotics, IoT, autonomous driving and DIY gadgets.

<https://platis.solutions/>





ATHOS COIMBRA RIBEIRO 🇧🇷

Master Student at Institute of Mathematics and Statistics -
University of São Paulo

FLOSS enthusiast, software Engineer, and Graduate Student in the
Institute of Mathematics and Statistics at University of São Paulo, persuing
a master degree under the advisory of Fabio Kon.

<http://www.ime.usp.br/~athoscr>



DIRK RIEHLE 🇩🇪 

Professor at FAU

Professorship of Open Source Software at Friedrich-Alexander University
Erlangen-Nürnberg in Germany



GREGORIO ROBLES 🇪🇸 (PhD 2006) @gregoriorobles

Associate Professor at Universidad Rey Juan Carlos

I am Associate Professor at the Universidad Rey Juan Carlos, a public
university in Madrid. I mainly do research in following two fields:
Free/Open Source Software, esp. by Mining Software Repositories, and
Computational Thinking. I co-founded Bitergia, a software analytics
company. I am getting old – I remember to have taken part in the first OSS
back then in 2005 in Genoa!

<http://gsyc.urjc.es/~grex>



IOANNIS ROUTIS 🇬🇷 (PhD 2019) 

PhD Student at Harokopio University

<https://www.linkedin.com/in/ioannis-routis-a58312a5/>





CHRISTOS SARDIANOS 🇬🇷 @ChSardianos 

PhD Student at HAROKOPIO UNIVERSITY OF ATHENS

I am a PhD Student at the Department of Informatics & Telematics at Harokopio University of Athens. My research area is “Knowledge Mining from Large Scale Social Network Data”, under the supervision of Assistant Prof. Iraklis Varlamis. and my main topics of interest are Data Mining, Recommender Systems, Social Network Analysis & Mining, Big Graph Analysis.

<http://dit.hua.gr/~sardianos/>

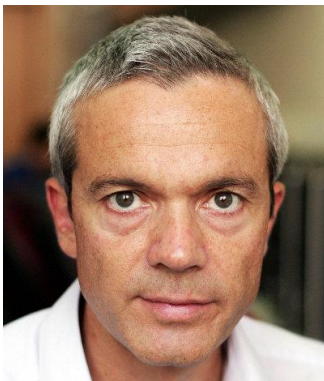


IOANNIS STAMELOS 🇬🇷

Professor at Aristotle University of Thessaloniki

-

<https://www.csd.auth.gr/el/staff/faculty?view=user&ro=1&id=14>

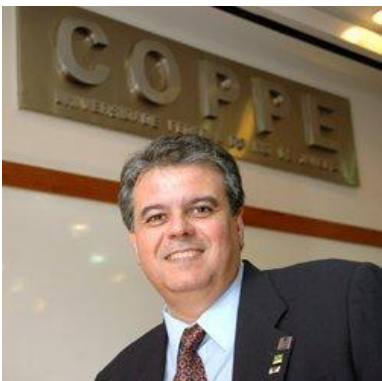


DIOMIDIS SPINELLIS 🇬🇷 (PhD 1993) @CoolSWEng 

Professor at Athens University of Economics and Business

Diomidis Spinellis is a Professor and Head in the Department of Management Science and Technology at AUEB. He's the author of two award-winning books, Code Reading and Code Quality, as well as more than 300 papers. His most recent book is Effective Debugging. He's a senior member of the ACM and the IEEE. From January 2015 he is serving as the Editor-in-Chief for IEEE Software.

<https://www.dmst.aueb.gr/dds/>



GUILHERME HORTA TRAVASSOS 🇧🇷 (PhD 1994)

Professor at COPPE/UFRJ

Professor of Software Engineering and a CNPq Researcher in Brazil. Holds a D.Sc. in Systems Engineering and Computer Science from COPPE/UFRJ, with a postdoc in Experimental Software Engineering at UMCP/USA. Leads the Experimental Software Engineering Group at COPPE/UFRJ. ISERN, SBC, and ACM member. Associate editor of Information and Software Technology, and editorial board member of IJSEKE, JSERD, and e-Informatica.

<http://www.cos.ufrj.br/~ght>





IRAKLIS VARLAMIS 🇬🇷 

Assistant Professor at Harokopio University of Athens

I am an Assistant Professor at the Department of Informatics and Telematics of Harokopio University of Athens. My research interests vary from data-mining and the use of semantics in web mining to social network analytics and knowledge extraction from social media and the news. I have published several articles in international journals and conferences on the above topics.

<https://www.dit.hua.gr/~varlamis/>



TONY WASSERMAN 🇺🇸 @twasserman

Professor at Carnegie Mellon University - Silicon Valley

Tony Wasserman works on evaluation, adoption, and use of FOSS by organizations (osspal.org). He was CEO of Interactive Development Environments (IDE), developer of Software through Pictures, one of the first 100 dot-coms, then VP at Bluestone Software. Before that, he was a Professor at UC San Francisco, where his User Software Engineering project built tools for rapid interactive systems design.



JOSEPH WILLIAMS 🇬🇧 (PhD 2019)

University Instructor at Canterbury Christ Church University

-

B.3. 11th Seminar on Advanced Techniques & Tools for Software Evoluton SATToSE 2018

CONFERENCE YEARBOOK



11th Seminar on Advanced Techniques & Tools for Software Evolution

SATToSE 2018

Athens, Greece, July 4th–6th 2018

#sattose18

PARTICIPANTS



MICHEL CHAUDRON 🇸🇪 📄 NOW HIRING 🔍

Professor at Chalmers | Gothenburg University

I am Full Professor at the Software Engineering division part of the joint Dept of Computer Science of Chalmers and Gothenburg Univ. (SE). I obtained my PhD in the area of formal methods and programming calculi for parallel computing. My research interests are in software architecture, software design, software modeling with a special focus on UML, software composition. I have an interest in empirical studies in software engineering, preferably industrial industrial context.

http://dblp.uni-trier.de/pers/hd/c/Chaudron:Michel_R=_V=



ALEXANDRE DECAN 🇧🇪 (PhD 2013) 📄

Postdoc at University of Mons, Belgium

Alexandre Decan obtained a Ph.D. degree in 2013 for the thesis entitled "Certain Query Answering in First-Order Languages". He is now postdoctoral researcher at the Software Engineering Lab of the University of Mons, where he has been actively involved in several research projects and has co-authored many publications related to the maintenance and evolution of software ecosystems.

<https://decan.lexpage.net>



VASILIKI EFSTATHIOU 🇬🇷

Postdoc at Athens University of Economics and Business

I am a Postdoc at the Athens University of Economics and Business where I mainly do research on mining natural language artefacts from software repositories. My main focus is on mining developer communication data; besides this, my research interests lie in the wider area of knowledge representation, reasoning and natural language processing.

<https://www.balab.aueb.gr/vasiliki-efsthioiu.html>



STEFANOS GEORGIU 🇬🇷 (PhD 2019) @StefanosGeorgi1

PhD Student at Athens University of Economics and Business

In love with programming, sporting, and travelling. Addicted to good quality food but with no quantity limitations. Performing research on Energy Efficiency with focus on Software Engineering.

https://stefanos1316.github.io/my_portfolio/home.html



TRUONG HO-QUANG 🇻🇳 (PhD 2019) @HoQuangTruong



PhD Student at Chalmers and Gothenburg University

My research interests lay in software architecture and practices of using software architecture artifacts (e.g. UML models, software architecture documentation) in OSS. For that I have collected more than 93k+ UML models from 24k+ Github projects and made some observations/surveys on the way developers of those projects use UML models. Take a look here to keep up with recent findings: <http://oss.models-db.com>

<https://www.linkedin.com/in/truonghoquang/>



DAMIEN LEGAY 🇧🇪

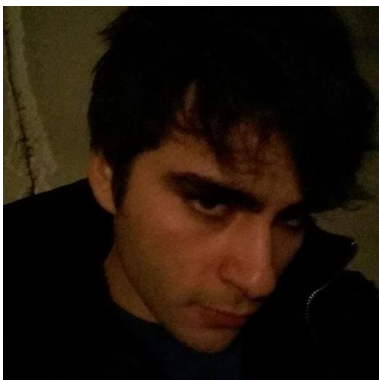
PhD Student at University of Mons



JOSE JAVIER MERCHANTE 🇪🇸 @JJ_Merchante

Researcher, Master Student at Universidad Rey Juan Carlos

My research interests are focused on Open Source Systems and mining repositories. My main focus is the Python language and the importance of "Pythonic" term in the community.



IASONAS NISSOPOULOS 🇬🇷 @IasonasNisso

Undergraduate at Athens University of Economics and Business

I find joy in life through programming, playing and listening to music, playing sports and several other activities.

<https://www.linkedin.com/in/iason-nissopoulos-a7b282120>



MICHAIL PAPAMICHAIL  (PhD 2020)

PhD Student at Intelligent Systems and Software Engineering Labgroup (ISSEL), Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki
Michail Papamichail is a Ph.D. candidate in the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece. His Ph.D. topic lies in the area of Software Analytics focusing on Software Quality from a User-perceived perspective, Software Lifecycle Analysis and Software Maintainability.

<https://issel.ee.auth.gr/staff/papamichail/>





GREGORIO ROBLES  (PhD 2006) @gregoriorobles

Associate Professor at Universidad Rey Juan Carlos
I am Associate Professor at the Universidad Rey Juan Carlos, a public university in Madrid. I mainly do research in following two fields: Free/Open Source Software, esp. by Mining Software Repositories, and Computational Thinking. I co-founded Bitergia, a software analytics company. I attended SATToSE for the first time in Mons in 2015 and I fell in love; since then I haven't missed one!

<http://gsyc.urjc.es/~grex>



TUSHAR SHARMA  (PhD 2018) @Sharma_Tushar 

PhD Student at Athens University of Economics and Business

I am a developer, researcher, author of three books including "Refactoring for software design smells" and creator of Designite. I like to talk about software in general and software quality, refactoring, and tools in particular.

<http://www.tusharma.in>



YANNIS SMARAGDAKIS 

Professor at University of Athens

Interests: Program analysis (static analysis, test generation, invariant inference, symbolic execution) Language mechanisms for abstraction (declarative languages, program generation, DSLs, modules and components, generics, extensible languages, multi-paradigm programming) Languages and tools for systems (concurrency models, distributed computing, memory management and program locality)

<http://smaragd.org>