

## Automatic annotation of Earnings Conference Calls

An Earnings Conference Call (ECC) is a telephone conference, or webcast, in which a public company discusses the financial results of a reporting period. It is a crucial means of communication of listed companies with the markets and, particularly with opinion leaders such as financial analysts. Text mining of ECC transcripts has become an important source of insight for high frequency traders and investors at large. In order to achieve a more relevant understanding of the potential impact of the communications, a deep analysis of ECC transcripts needs to be carried out, requiring the identification of the identity of the speakers, of the peculiar Question & Answer (Q&A) structure, and of the style of the answers.

The project aims at providing:

1. an automatic annotation of the *structure* of these transcribed documents, including the breakdown of major sections, the identification of questions and answers, and the identification of the speaker names;
2. an automatic annotation of the *style* of the questions and answers, including the identification of potential markers of the psychological status of the speakers;
3. a classification of the *questioning strategy* according to a provided genre-specific typology;

The final goal of such a tool is to map question and answer patterns on a larger scale to see which ones are relevant to predict stock price and other market variables. As an example of such documents, [here](#) is the fourth quarter 2016 conference call of Apple Inc.

Required skills are: an inquisitive mind, interest in a new and original problem, decent programming skills, and some knowledge of basic machine learning techniques.

The work is part of a research collaboration between the faculties of Informatics (INF) and Communication Sciences (COM) of USI. This project is a very unique opportunity for the applicant to gain a very valuable experience in software development and data analysis that will give him/her a taste of a real research project.

For more details please contact Prof Fabio Crestani (INF) and/or Prof Andrea Rocci (COM). The project will be supervised by Prof Crestani with the help of Dr Carlo Raimondo (postdoc in COM) and Esteban Rissola (PhD student in INF).

## **Sharding BitCoin for fun and profit**

Bitcoin is a highly hyped cryptocurrency used worldwide as a payment system. Although BitCoin is fully decentralized, in that there is no central repository or single administrator, each participating peer must store the complete sequence of transactions in what is known as the blockchain. Moreover, standard peers must also execute every single transaction that is integrated in the blockchain. This essentially means that each peer must store BitCoin's complete state, mostly composed of the various user accounts.

The goal of this project is to assess the potential performance benefit that sharding could bring to BitCoin. Sharding is a standard technique used to scale performance: Instead of storing the complete application state, each partition keeps only part of the state. If the partitioning can be such that most operations involve one shard only and single-shard operations are evenly distributed among shards, then performance can increase with the number of shards.

In the project, we intend to analyze how different sharding techniques would perform if BitCoin's blockchain could be sharded. We intend to use the actual transactions publicly available in the blockchain to perform the study by replaying the transactions and determining for each technique the ratio of single-shard and multi-shard transactions.

Some understanding of how BitCoin works is clearly an advantage. Good programming skills are a must.

## Accurate Profiling of Computations on the JVM using Bytecode-level Metrics

Program analysis and optimizations often need to identify underutilized or overutilized computing entities, such as actors, tasks, or threads. To measure the work performed by them (and thus their utilization), performance analysts have proposed different metrics. In applications running on the Java Virtual Machine (JVM), work is typically measured through either bytecode-level metrics or hardware-level metrics. On the one hand, bytecode-level metrics like bytecode count (i.e., the number of bytecodes executed by a computing entity) can be profiled in any environment with little effort, as they are easy to collect, platform independent, and do not require any hardware support. Unfortunately, the accuracy of such metrics is generally limited, as they cannot account native code, dynamic optimizations, and fall short in differentiating between simple and complex operations. On the other hand, hardware-level metrics like machine-instruction or reference-cycle count (i.e., the number of machine instructions or clock cycles executed by a computing entity) can quantify work with a higher accuracy, but are more difficult to collect, require ad-hoc support, and demand careful instrumentation to be profiled with little perturbations.

This project aims at 1) studying whether and to which extent bytecode-level metrics can approximate hardware-level metrics when measuring work in applications running on the JVM, and 2) finding new means of measuring work through bytecode-level metrics without significant loss of accuracy wrt. hardware-level metrics, two aspects that have been little explored in the current literature. This would significantly ease the effort of profiling work with high accuracy, and would enable more accurate measurements and analyses in applications that cannot benefit from hardware-level metrics, such as those running on old hardware or in the cloud.

The student will be involved in a number of activities:

1. Development of profilers for measuring bytecode-level and hardware-level metrics with high accuracy.
2. Analysis of the relationship between bytecode-level and hardware-level metrics in different workloads, environments, and under different system loads.
3. Investigation of new methodologies towards increasing the accuracy of bytecode-level metrics.
4. Evaluation of the proposed methodologies in different use cases and scenarios.

The project is a unique opportunity for the student to deepen his/her knowledge in the domains of concurrent programming, dynamic program analysis, and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen his/her abilities as a computer scientist. Applicants interested in this project should have a good knowledge of Java, C/C++ and UNIX-based operating systems, excellent programming skills, and deep interest in the fields of concurrent programming and dynamic analysis.

Teacher: Walter Binder <[walter.binder@usi.ch](mailto:walter.binder@usi.ch)>

Assistants: Andrea Rosà <[andrea.rosa@usi.ch](mailto:andrea.rosa@usi.ch)>, Eduardo Rosales <[rosale@usi.ch](mailto:rosale@usi.ch)>

## Automatic Benchmark Synthesis for Specific Evaluation Needs

Benchmarking is crucial for the experimental validation of new algorithms and systems. Researchers rely on benchmarks to demonstrate the feasibility or efficiency of their contributions. However, finding the right benchmark suite can be a daunting task, because existing benchmark suites may be outdated, known to be flawed, or simply irrelevant for the experimental evaluation at hand. Creating a proper benchmark suite is challenging, extremely time consuming, and also - unless it becomes widely popular - a thankless endeavor.

This project aims at the development of a new software infrastructure to automatically synthesize benchmark suites from large-scale open-source code repositories. The infrastructure will be able to automatically process huge code repositories such as GitHub, applying complex program analyses to the open-source projects in the repositories. Such analyses will allow one to synthesize custom benchmark suites targeting specific evaluation needs. The analyses will be represented as workflows of user-defined static and dynamic analyses, expressed in a new domain-specific language.

The student will be involved in a number of activities:

1. Development of a novel software infrastructure to automatically execute static and dynamic analyses on large-scale open-source code repositories.
2. Identification of scenarios and communities where there is a need for new benchmark suites (e.g., runtime verification, parallel computing).
3. Development of new program analyses to identify workloads suitable to be used as representative benchmarks for specific communities.
4. Generation and evaluation of novel benchmark suites matching specific user needs.

The project is a unique opportunity for the student to deepen his/her knowledge in the domains of repository mining, workload discovery, program analysis, and empirical evaluation, all being important skills for a software engineer. The student will work side-by-side with the members of the Dynamic Analysis Group at USI, and will receive support in learning advanced topics that will strengthen his/her abilities as a computer scientist. Applicants interested in this project should have a good knowledge of Java and UNIX-based operating systems, excellent programming skills, and deep interest in the field of program analysis.

Teacher: Walter Binder <[walter.binder@usi.ch](mailto:walter.binder@usi.ch)>

Assistants: Andrea Rosà <[andrea.rosa@usi.ch](mailto:andrea.rosa@usi.ch)>, Eduardo Rosales <[rosale@usi.ch](mailto:rosale@usi.ch)>

UROP 2018 – Project Proposal:

### **Fiabot!Plus: Bigger, Better and Stronger**

“Fiabot!” is the name of a mobile application designed, prototyped, and tested in 2014 with and for primary school children in local schools. Currently it is a simple iPad application, based on Keynote, that enables the creation of interactive and multimedia stories by letting children store and combine in a structured manner into one presentation images, text, video and audio. It fits with class activities where Fiabot! provides children with the opportunity to increase their ability to create a specific story genre by giving indications of the structure and “ingredients” needed for each story type as well as reinterpreting and formalizing the workflow of the story creation.

Fiabot! proposes a different structure for each literary genre and a “list of ingredients” to inspire the children while they are creating the plot of a story and drafting the characters. Children are then guided through the following stages: definition of story structure and plot; media creation and editing, and finally sharing within the class and publication of the story.

In the past few years we have extensively studied it from the user point of view. We have explored how Fiabot! helped creating stories and introduced it to older (intermedia school) and younger (pre-school) children. Now it is time to focus on its system side, and consider how it could be better engineered.

This short project aims at delivering a new, more stable and sophisticated version of Fiabot!, to be used by children both at school and at home for creating their multimedia stories. Fiabot!Plus will revisit the original Fiabot! app in order to overcome its limitations in terms of portability and use. We expect the candidate student to start from the analysis of the limitations of the current app, from a system perspective as opposed to the user one. He/she will then proceed with the design, implementation and testing of the new one. We envisage it to be a Web app with a tablet interface and a focus on the complexity of exchanging data to and from the two platforms.

Fiabot!Plus will offer a student the chance to engage with the complexities of working with real users without having to go through the ordeal of running a user study from scratch. The chosen student will also enjoy working in a truly multidisciplinary team.

It would be an ideal experience for a student looking forward developing apps for Web and tablets while working for real users.

# UROP Project Proposal



## MoviePlus – A Platform for Sharing Digital Content For Augmented Reality Displays

Nowadays people are using tablets, laptops and various digital devices to work collaboratively on the same project while being present physically or remotely. Imagine to work on the same

design or analytical task, or just simply watch an entertaining video using Google Glass. Annotating and interacting with virtual content became a daily activity of many architects, engineers, designers, artists, etc. We offer a student to join on exciting opportunity to design a prototype, which enables digital content sharing and interaction in Augmented Reality. The student design and develop the shared movie viewing system using VideoPlus, a platform for augmented video content (<https://search.usi.ch/en/projects/953/videoplus-a-platform-for-augmented-video-content>).

We propose to use the combination of a head-mounted display (e.g., iPhone X/ Pixel 2 with Daydream or Microsoft HoloLens) and a wrist-worn controller (e.g., Apple Watch) to develop an interactive system, which would enable tagging favorite scenes, sharing subtitles, sending ephemeral messages among viewers and/or taking collaborative trivia quizzes. We envision prototyping on the mobile device as part of the project work, so we would encourage students with iOS/Android or C#/C++ development skills to apply. Ultimately, we would like to conduct controlled experiments to evaluate a designed system.

This project is hosted by Faculty of Informatics and Institute of Computational Science. Thus, the student will be co-supervised by the responsible investigators from both sides.

# Estimation of Large-Scale Gaussian Graphical Models

Recovering the (inverse) covariance matrix is a fundamental task in the modern multivariate analysis. In this project, we are concerned with a sparse approximation of the (inverse) covariance matrix. This task is challenging due to the inherently high degree of uncertainty in the estimate when the number of samples is limited. A sparse approximation of the covariance matrix is essential in high-dimensional settings, and can also have a significant impact on the accuracy of the approximation.

In this project, a new method for approximating the sparse structure of the covariance matrix is explored. The algorithm operates by representing the matrix elements as the vertices of an undirected graph. The goal of this project is to implement a parallelized variate of the algorithm which is fit for modern compute architectures. The algorithm will then be coupled with the Sparse Quadratic Inverse Covariance Approximation (SQUIC) code base and tested against synthetic data. Finally, the developed framework is applied to optimization problems in finance (portfolio optimization, regression, etc.).

Professor: Olaf Schenk, Advanced Computing Laboratory, Faculty of Informatics.

PhD Student: Aryan Eftekhari, Advanced Computing Laboratory, Faculty of Informatics.

**Prerequisites** - The candidate should have working experience with Matlab, C++, and MPI and have completed coursework in linear algebra and statistics.

# Automatic Formative Feedback for Java Programming Assignments

Software and Programmer Efficiency Research Group



We use Informa in our Programming Fundamentals 2 course, and in some master-level courses. Informa currently supports automatically graded multiple-choice questions (in so-called "practice problems"), and it supports lab assignments graded by human teaching assistants.

The goal of this UROP project is to develop an autograder for Java programming assignments. The goal of the autograder is to provide immediate feedback to students, the kind of feedback traditionally provided by human teaching assistants. The learning sciences distinguish between two kinds of feedback: **summative** feedback (a score, a grade, used for a decision: e.g., pass or fail a course, rank students, hire employees) and **formative** feedback (information that helps the learner to improve their understanding).

The first step in building an autograder is to feed the submitted assignment through a compiler and then run it through a set of unit tests to check whether it passes. The second step is to provide formative feedback by mapping compiler errors and test failures to specific skills and misconceptions defined for the course.

If successful, this project will allow future generations of students using Informa to receive immediate formative feedback on their programming problem submissions in a way that makes learning more interesting and more effective.

If you are potentially interested in this project and have any questions, please do not hesitate to contact Matthias Hauswirth.

## Technologies

### Informa

Asynchronous Blended Mastery Learning Platform

A full-stack web application written in JavaScript on top of Meteor with MongoDB



### INGInious

The Docker-based backend of Ingenious for sandboxing compilation and unit test tasks

[http://inginius.readthedocs.io/en/latest/dev\\_doc/understand\\_inginius.html](http://inginius.readthedocs.io/en/latest/dev_doc/understand_inginius.html)

# Automatic Identification of Student Misconceptions with Machine Learning

Software and Programmer Efficiency Research Group



In Programming Fundamentals 2, you read a textbook and had to submit so-called "recall statements", short texts recalling the key points you read in each textbook section (we asked you to do this because recalling from memory what you just read has been shown to be an effective learning technique).

Now imagine the following scenario: Right after entering your recall statement, the system responds that you seem to have misunderstood a given concept, and it points you to specific information that helps you clarify your understanding.

In this UROP project you will use machine learning to try to realize this idea. Given the recall statements from a past PF2 course, the set of skills and misconceptions defined for the course, and the text of several open textbooks as training data, you will develop an approach to automatically identify misconceptions in newly entered recall statements.

You will evaluate your approach on recall statements from a second PF2 course. If the evaluation shows that the identification approach is effective, you can integrate the approach into the Informa platform, so that in future instances of PF2, Informa will automatically point out and explain misconceptions whenever a student enters a new recall statement. Moreover, a successful project can lead to a scientific publication.

If you are potentially interested in this project and have any questions, please do not hesitate to contact Matthias Hauswirth.

## Technologies



Deep Learning 4 Java, and its **word2vec** implementation

<https://deeplearning4j.org/word2vec.html>

## Informa

Asynchronous Blended Mastery Learning Platform

A full-stack web application written in JavaScript on top of Meteor with MongoDB

## Detection of Social Interactions using Smartphones

Today's wearable cameras (e.g., Google Clip) allow us to seamlessly capture various aspects of our daily activities in digital format, e.g. family moments, workout routines, work meetings etc. Such data can serve as "memory cues" to help individuals recall memories of certain past events. To create even richer memory cues, pictures taken by co-located people can often be helpful. For example, when giving a presentation, the data captured by the cameras of talk attendees may constitute a better memory cue than our own pictures (which would capture the audience). In a previous project, we implemented a smartphone application that supports proximity-based sharing of images and other captured data, based on the detection of nearby Bluetooth beacons of other users. In this summer project, we plan to extend the existing app to actually detect *social engagement*, e.g., when we directly engage with others, rather than simply detecting proximity of nearby phone (which may wrongly detect, e.g., people behind us or in a room next door).

At the outset, the project will investigate different approaches for detecting social engagement using a mobile device and conduct a comparative evaluation of few such techniques. Subsequently, the project will aim to provide a proof-of-concept prototype of the most promising approach based from the evaluation study.

The project requires mobile programming skills (preferably Android) and willingness to learn about the literature in social engagement detection.

## BIO-CAPTCHA

### **Introduction.**

This interdisciplinary project mixes web technologies and biomedical imaging to enhance computer vision algorithms. It will be done as a joint collaboration between the research groups led by Prof. Dr. Rolf Krause – High Performance Methods for Numerical Simulation in Science, Medicine and Engineering - and by Dr. Santiago Fernandez Gonzalez – Infection and immunity –.

The student will be tutored by Diego Ulisse Pizzagalli.

Multiphoton intravital microscopy (MP-IVM) is a technique for capturing, in 3D images and videos, the behavior of cells inside organs and tissues. Focusing on immunological research, in the last decade MP-IVM revealed unprecedented cell-to-cell interaction patterns, opening the possibility to discover new treatments for infective, autoimmune and neoplastic pathologies among others.

The first step required for analyzing microscopy data is segmentation (i.e. outlining the contour of objects). Despite recent advances of Computer Vision methods, this task is still challenging for machines and the human eye revealed with significantly superior performances.

CAPTCHA allowed discriminating between humans and bots by asking the user to solve tasks easily achieved by humans but still hard for machines.

This methodology further resulted suitable for obtaining human-generated datasets such as for digitalizing degraded books, not interpretable by classical OCR methods (reCAPTCHA).

Therefore, this project aims at developing a CAPTCHA-like web service and user interface that requires user to perform segmentation on MP-IVM data. This project will be beneficial for creating a large dataset of manually segmented cells, which in turn is required to validate and developing novel computer vision methods.

### **Outline.**

Milestone 1: Graphical interface to draw the contour of objects in biomedical images (July 2018)

The student is asked to design and develop a web based tool to draw the contour of objects in biomedical images. Skills on client-side web programming languages such as HTML5/JS/CSS will be required and fostered.

Milestone 2: Web service (August 2018)

The student is asked to develop a web service that provides data to the previously developed UI and allows the integration with 3rd parties web applications. It will be developed using common server-side programming languages such as node.js or PHP.

# UROF 2018 - Sampling and Color Voronoi Diagrams

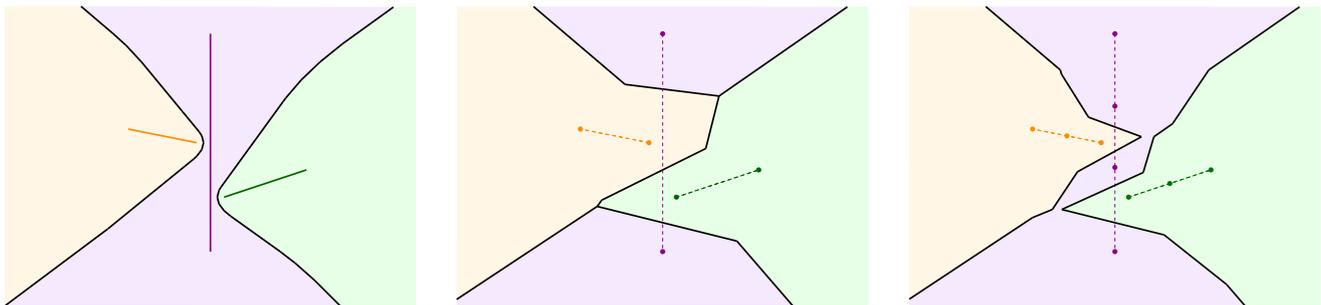
## Approximate Voronoi diagrams of different geometric objects

---

The **Voronoi Diagram** is a *geometric data-structure* encoding proximity information among different objects. It finds applications in diverse areas of *Computer Science*, such as *Robotics*, *Computer Graphics*, *etc.*, as well as in other sciences, such as *Biology*, *Chemistry*, *etc.* Despite having been widely studied over the last decades, *Voronoi diagrams* still attract the interest of the scientific community, with their numerous generalisations, variations and applications.

The (*nearest*) **Voronoi diagram** (*VD*) is a partition of the plane into regions. Each region corresponds to one *site*, such that all points of  $\mathbb{R}^2$  belonging to that region are closer to that site than any other. We are interested in diagrams where the sites are cluster of points and in particular, **Color Voronoi Diagrams** (*CVD*). In *CVDs* the distance of a point to a *site-cluster* is realised by the nearest point of that cluster. More specifically, the distance of a point  $x$  to a cluster of points  $P$  is  $d(x, P) = \min_{p_i \in P} d(x, p_i)$ .

When the sites are complex geometric objects computing a *Voronoi diagram* can be a complicated task. Depending on the application, many times approximate *VDs* can be sufficient. In that direction, we propose a method of approximating *VDs* based on (1) sampling the sites and then (2) constructing the *CVDs* of the sampled point clusters.



(a) Nearest line-segment *VD* of three sites.

(b) Nearest *CVD* of the segment endpoints sampled.

(c) Nearest *CVD* of an average-sized sample.

During this project the student is expected to devise efficient techniques for sampling geometric objects in order to construct approximate *VDs*. We will focus on diagrams where the region of each site is connected, see (a). The resulting approximate *Voronoi diagrams* should maintain the connected regions of the sites. Observe (c) in contrast to (b) where a region of one cluster is disconnected. At the same time, the size of the sampled points should be kept to a minimum. Moreover, the quality of the approximated diagram and its relation to the size of the sample could be examined.

### Prerequisites:

- ▷ Good background in **Algorithms and Data Structures**.
- ▷ **Eagerness** to discover new things and tackle potential difficulties.
- ▷ Knowledge of a **programming** language, preferably C++ or Python.

### Benefits:

- ▷ Work within a **research group** on an **intriguing problem**.
- ▷ Acquire profound knowledge in several **algorithmic aspects**.

### Contact:

- ▷ **Prof. Evanthia Papadopoulou**, SI-209, [evanthia.papadopoulou@usi.ch](mailto:evanthia.papadopoulou@usi.ch)
- ▷ **Ioannis Mantas**, Open Space 2<sup>nd</sup> floor (SI-2 VIII), [ioannis.mantas@usi.ch](mailto:ioannis.mantas@usi.ch)

## Patent Data Analysis

According to Wikipedia, “A *patent* is a set of exclusive rights granted by a sovereign state or intergovernmental organization to an inventor or assignee for a limited period of time in exchange for detailed public disclosure of an invention. An invention is a solution to a specific technological problem and is a product or a process. Patents are a form of intellectual property.”

The objective of this project is to expand, curate and analyse the patents registered with the various patent offices across the globe. As a first step towards the data collection we have already obtained patents registered in US (USPTO) and in Europe (EPO) for the period 1978-2016. This equate to over 200 million patent documents. The first step would be to unify their mark up, to be able to recognise the specific fields of each patent. A second step would involve linking one to another all patents according to their citations data, including expanding the citation network to non-patent literature (e.g. scientific articles, reports, etc.) to obtain a comprehensive network of the evolution of science and technology as perceived by innovators. At this stage we will be able to start the last and most interesting part of the project which will involve the full-text mining of the network to analyse and forecast innovation trends expanding over 50 years in different technological areas and for different countries.

The student responsibilities will therefore be as follows:

1. To curate and pre-process the patent collections in a uniform file format (favourably XML). Fill out any missing information or edit any incorrect ones after collection.
2. To collect non-patent publications and merge them appropriately with the existing patent collections.
3. To create a database and populate them with patent information for easier access and retrieval later on.
4. To help the project team perform some (initially) simple data analysis tasks on the collected data to identify interesting patterns and trends related to the temporal patent evolution across countries and categories.

This project will collaborate with the SNSF [GSKN project](#).

Required skills are: the student should be familiar with databases and have some experience with scripting languages.

For more details please contact Prof Fabio Crestani. The project will be supervised by Prof Crestani with the help of Manajit Chakraborty (PhD student).

## How Do Users Interact with a Search Engine on Mobile devices?

Recent years have witnessed a dramatic rise in usage of mobile devices among all types of users worldwide. Given the fact that many users have internet access on their mobile devices, therefore, it is critical to develop an Information Retrieval (IR) system that is specifically designed to be used on mobile devices. To this end, we are going to perform a thorough and comprehensive user study in the attempt to track and mine how users interact with their devices and express their information needs while they are on the go. Such study will enable us to anticipate users' information needs and understand more accurately the users' information needs when they submit a query on their mobile devices given their context and history of interaction with their personal device.

Mobile devices are equipped with a number of sensors such as GPS, light sensor, microphone, gyroscope, etc. Having the data from all such sensors *together* with users' interaction and usage patterns on their devices could provide us a wealth of information to analyse users search behaviour deeply.

In this project the applicant will have to design and develop a mobile app to track users' behaviour and analyse the data gathered through mobile phones. The student responsibility will be to:

1. Design and development of an Android app which volunteers will install on their mobile devices (or on devices that we will provide them with) to track all their activities (provided they agree before we start the experiments). We aim at tracking all sensor data readings on the mobile devices for a limited period of time. The applicant will be free to use any open source and/or publicly available application as long as he/she can integrate them in the system.
2. Help recruit some volunteers who will either give us the required access on their mobile devices or will agree to use mobile devices we will provide them with.
3. Monitor the gathering the data for short period of time.
4. Analyse the gathered data in order to answer some research questions about the relation between relevant document and users' mobility patterns.

This project will collaborate with the SNSF [ReIMobIR project](#). The work is expected to result in a research paper that will be submitted to one of the top tier conferences in IR. It is a very unique opportunity for the applicant to gain a very valuable experience in app development and data analysis that gives him/her a taste of a typical research experience in IR.

Required skills are: solid background in programming and problem solving. A background in data analytic and/or IR is a plus.

For more details please contact Prof Fabio Crestani. The project will be supervised by Prof Crestani with the help of Mohammad Nejadi (PhD student).

## Question Answering about Past Meetings

Question Answering is an Information Retrieval (IR) problem which is concerned with building systems that automatically answer questions posed to a system by humans in natural language. Most current Question Answering (Q&A) systems can handle factoid questions, i.e. questions that ask for simple facts. For example, “where is USI located?” is a factoid question.

Apart from the issues concerned with building the most suitable answer given a question (which are not the object of this project), a Q&A system may extract its answers by querying a structured database of knowledge or information that was build beforehand. Such systems are usually composed of a number of components that: 1. Process a given question to extract keywords from the question and determine the answer type. 2. A query formulation component that construct a query to be sent to an IR engine. 3. A search engine that retrieves documents related to the query. 4. An answer generating component which prepares a simple answer based on the retrieved documents.

In this project, we would like to extend the current factoid Q&A systems, to handle personal data such as a collection of recorded past meetings. In this case, the question answering system should be capable of answering questions about one’s past meetings and conversations. Example of such questions would be “When was the last time I met John?”, “Where did we meet?”, “What did we talked about?”, “Was John the only person I talked about that?”, “What was Tom’s opinion about that?”.

Required skills are: an interest in IR and databases technologies, some simple knowledge of machine learning and text processing. The required programming language for this project is Python.

For more details please contact Prof Fabio Crestani. The project will be supervised by Prof Crestani with the help of Seyed Ali Bahrainian (PhD student).

# **Hardware/software co-design of application-specific systems with Xilinx SDSoC.**

Field Programmable Systems-on-Chips (FP-SoCs) are heterogeneous architecture comprising processors, memories and reconfigurable hardware blocks. Their structure allows the generation of application-specific processing platforms, comprising flexible accelerators to speed-up the computation of intensive routines.

However, the design of such systems traditionally require complex design flows, and mandates a high level of expertise both in hardware and software design. Recently, Xilinx SDSoC [1] addressed this challenge by offering a unified, software-driven cockpit for the development of applications on FP-SoC platforms. Starting from standard C/C++ code, SDSoC allows to mark functions for either hardware or software implementation, and optimize the latter for different performance/area targets [2]. Transparently to the programmer, accelerators are generated, interfaced to the system, and software handlers are created to manage their execution and the required data transfers.

Goal of the project is to explore the achievable performance of SDSoC-based systems on a set of benchmark applications, evaluating diverse acceleration choices and architectural configurations. During the course of the project, the student will acquire proficiency with the generation of software-defined hardware platforms, and the trade-off ensuing from different optimization choices.

Requirements:- Interest in computer architectures and digital design- Experience with C is a plus.

Mentors:- Laura Pozzi- Giovanni Ansaloni - Lorenzo Ferretti

[1] SDSoC Development environment: [www.xilinx.com/products/design-tools/software-zone/sdsoc.html](http://www.xilinx.com/products/design-tools/software-zone/sdsoc.html).

[2] SDSoC Development environment demo: [www.xilinx.com/video/software/sdsoc-development-environment-demo.html](http://www.xilinx.com/video/software/sdsoc-development-environment-demo.html)

## **Fake news detection in social media**

In the past decade, social media have become one of the main sources of information for people around the world. Yet, using social media for news consumption is a double-edged sword: on the one hand, it offers low cost, easy access, and rapid dissemination. On the other hand, it comes with the danger of exposure to “fake news” with intentionally false information. The extensive spread of fake news has recently become a centerpiece of controversy in the United States and United Kingdom following the highly debated elections of President Donald Trump and the Brexit vote. It is alleged that the outcome of these votes resulted from the public opinion manipulation by a massive injection of fake news, possibly produced by influence agents or even sponsored by hostile foreign governments. Due to the heavy societal toll, some analysts have declared fake news among the most serious and unprecedented threats to the modern democracies.

The project will attempt applying a novel class of machine learning algorithms developed in our group to the problem of automatic detection of fake news. The project will encompass a variety of tasks, including development of applications and data collection from social networks, development of machine learning methods, and scaling them up in the AWS cloud. The student will have the chance to work on an important problem, learn valuable new skills, and be exposed to bleeding edge technology.

**Requirements:** excellent programming skills, in particular very good knowledge of python. Preferably experience with social networks (Facebook/Twitter) APIs

**Contacts:** Prof. Michael Bronstein ([michael.bronstein@usi.ch](mailto:michael.bronstein@usi.ch)), Dr. Davide Eynard ([davide.eynard@usi.ch](mailto:davide.eynard@usi.ch)), Federico Monti ([federico.monti@usi.ch](mailto:federico.monti@usi.ch))