<table>
<thead>
<tr>
<th>No.</th>
<th>Project Domain</th>
<th>Project Name</th>
<th>Company</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Search engines</td>
<td>Near-duplicate documents detector</td>
<td>Outbrain</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Mobile apps, voice activation</td>
<td>Android app for voice-activated playback of audio files</td>
<td>Outbrain</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Image Classification</td>
<td>Automatic storytelling</td>
<td>Shutterfly</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Ranking of images on mobile devices</td>
<td>Mobile Picteligence</td>
<td>Shutterfly</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Gmail API, Rule engines, Machine learning</td>
<td>Automatic email reply</td>
<td>Google</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>Big data algorithms</td>
<td>Parallelizing property testing algorithms</td>
<td>Google</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>VLSI HW verification</td>
<td>Build a formal verification kit to bridge between OCP and AHB bus protocols</td>
<td>Apple</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>VLSI HW verification</td>
<td>Fully automate the way that we handle BD memory loading</td>
<td>Apple</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>VLSI HW verification</td>
<td>Develop a GUI platform for DFT env generation</td>
<td>Apple</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>operating systems, computer security</td>
<td>Automatic Windows image anomaly detection</td>
<td>VMWare</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>Operating Systems, Application Management, Data Mining</td>
<td>“App Analyzer&quot;- Deliver the right apps users want!</td>
<td>VMWare</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>Operating systems, computer security, networking</td>
<td>“Quicksight” – Smart and fast automatic IT assessment for Windows networks</td>
<td>VMWare</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>Dev ops</td>
<td>Manage IT world: Using dockers with hyperic agent</td>
<td>VMWare</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>Operating systems, computer security, networking</td>
<td>Advanced desktop recovery with Intel® vPro and VMware Horizon Mirage</td>
<td>VMWare</td>
<td>11</td>
</tr>
<tr>
<td>15</td>
<td>Big Data &amp; monitor</td>
<td>Yoda resource management</td>
<td>Akamai</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>Big Data</td>
<td>Hive adaptor for data-catalog</td>
<td>Akamai</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>System monitoring</td>
<td>Log-Me</td>
<td>Akamai</td>
<td>13</td>
</tr>
<tr>
<td>18</td>
<td>Big Data</td>
<td>HBlaze</td>
<td>Akamai</td>
<td>14</td>
</tr>
<tr>
<td>19</td>
<td>Data Structures, Algorithms, Signal and Image Processing</td>
<td>Audiobook with Translation</td>
<td>Qualcomm</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td>Operating Systems, Data Structures, Software Design</td>
<td>Support for Chinese input</td>
<td>Qualcomm</td>
<td>15</td>
</tr>
<tr>
<td>21</td>
<td>Data Structures, Algorithms, Computer Networking, Databases</td>
<td>Teach for Free, Learn for Free – Social App of Knowledge Exchange</td>
<td>Qualcomm</td>
<td>16</td>
</tr>
<tr>
<td>22</td>
<td>Operating Systems, Software Design</td>
<td>Vector-based fonts to bitmap renderer</td>
<td>Qualcomm</td>
<td>17</td>
</tr>
<tr>
<td>23</td>
<td>Machine learning</td>
<td>Pattern detection and alerting system</td>
<td>Microsoft</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>Research Area</td>
<td>Tool</td>
<td>Organization</td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>24)</td>
<td>OO Software development GUI</td>
<td>Rafael software project</td>
<td>Rafael</td>
<td>19</td>
</tr>
<tr>
<td>25)</td>
<td>Software Systems</td>
<td>CANOLA (CS) - CAN Online Listening Application</td>
<td>Philips</td>
<td>20</td>
</tr>
<tr>
<td>26)</td>
<td>Software Systems</td>
<td>ROI2DB (Regions Of Interest to Database) Application</td>
<td>Philips</td>
<td>21</td>
</tr>
<tr>
<td>27)</td>
<td>Software Systems</td>
<td>TTeV – Tile Evolution Application</td>
<td>Philips</td>
<td>22</td>
</tr>
<tr>
<td>28)</td>
<td>Software Systems</td>
<td>Tiltant-CFGM – Tile Test Analyzer Configuration Manager</td>
<td>Philips</td>
<td>23</td>
</tr>
<tr>
<td>29)</td>
<td>Web application, DB, UI, HTML5</td>
<td>Refactoring Business DB Viewer Tool</td>
<td>Amdocs</td>
<td>24</td>
</tr>
<tr>
<td>30)</td>
<td>Information System, Development process management tool, GUI</td>
<td>Interfaces Management Tool</td>
<td>Amdocs</td>
<td>26</td>
</tr>
<tr>
<td>31)</td>
<td>Automation, Generation, Java, Web, Javadoc, Rest</td>
<td>Javadoc tool for REST Data mapping layer</td>
<td>Amdocs</td>
<td>27</td>
</tr>
<tr>
<td>32)</td>
<td>Communication networks, Mobile small cell, Analytics, Network planning and design</td>
<td>Mobile cell site process analysis and prediction</td>
<td>Amdocs</td>
<td>28</td>
</tr>
<tr>
<td>33)</td>
<td>ATDD, TDD, Testing, Web development, JS, HTML5</td>
<td>Modern Web UI Testing Suite (ATDD) with a Speck Wiki Reporter</td>
<td>Amdocs</td>
<td>28</td>
</tr>
<tr>
<td>34)</td>
<td>Automation, Search, Algorithm, Parsing</td>
<td>Code parsing engine for test-automation</td>
<td>Amdocs</td>
<td>29</td>
</tr>
<tr>
<td>35)</td>
<td>Mobile applications</td>
<td>Anti-patterns for Mobile hybrid applications</td>
<td>IBM</td>
<td>30</td>
</tr>
<tr>
<td>36)</td>
<td>Multi-threading, formal verification, GUI development</td>
<td>Eclipse plug-in for visualizing failures in multi-threaded C/C++ programs</td>
<td>IBM</td>
<td>31</td>
</tr>
<tr>
<td>37)</td>
<td>Information Retrieval/Text Mining</td>
<td>Wikipedia References Classification</td>
<td>IBM</td>
<td>31</td>
</tr>
<tr>
<td>38)</td>
<td>SW Tools for Hardware Verification</td>
<td>Static Analysis for Failure Localization in Hardware Validation</td>
<td>IBM</td>
<td>32</td>
</tr>
<tr>
<td>39)</td>
<td>Networks</td>
<td>Simulation of Flow-Policies Configuration in SDN</td>
<td>IBM</td>
<td>33</td>
</tr>
<tr>
<td>40)</td>
<td>Information retrieval, Data mining</td>
<td>Extracting IS-A relations from Wikipedia</td>
<td>IBM</td>
<td>34</td>
</tr>
<tr>
<td>41)</td>
<td>Social Networks, Social Analytics, Search</td>
<td>Real-Time Trend Detection in Social Networks</td>
<td>IBM</td>
<td>35</td>
</tr>
<tr>
<td>42)</td>
<td>Computer vision, image/video processing</td>
<td>Real-time moving object detection in aerial video</td>
<td>IBM</td>
<td>35</td>
</tr>
<tr>
<td>43)</td>
<td>Verification and testing, software engineering</td>
<td>Finding security vulnerabilities in applications using symbolic interpretation of C/C++ source code</td>
<td>IBM</td>
<td>36</td>
</tr>
<tr>
<td>44)</td>
<td>Machine learning</td>
<td>Deep learning algorithms implementation and validation</td>
<td>IBM</td>
<td>37</td>
</tr>
<tr>
<td>45)</td>
<td>Databases, Data Mining, Machine Learning</td>
<td>Feature Generation for Prediction over a Database</td>
<td>LogicBlox</td>
<td>38-39</td>
</tr>
<tr>
<td>46)</td>
<td>Databases, Artificial Intelligence</td>
<td>Translating Soft Logic into LogicBlox Optimization</td>
<td>LogicBlox</td>
<td>39-40</td>
</tr>
</tbody>
</table>
PROJECT 1

Company: Outbrain Inc.

Project Domain: Search engines

Project: Near-duplicate documents detector

Supervisor: Ronny Lempel

Project Description: The identification of near-duplicate documents is a basic problem in search engine technology. Search engines do not necessarily want to fully index the potentially hundreds of seemingly identical documents, which may differ from each other by very small amounts of tokens.

In this project, the students will build a service that ingests an RSS feed of URLs, crawls the documents, “fingerprint” each document, and then: (1) enters the fingerprint into an Apache Solr search server; and (2) searches Solr for previous documents with similar fingerprints, writing matches to a log file.

Programming Languages and Development Platforms: the project will be written in either Java or Scala, and will use the Apache Solr open source search server.

Required Background: Any third year student is eligible. The students will need to read about document fingerprinting as described in a research paper.

PROJECT 2

Company: Outbrain Inc.

Project Domain: Mobile apps, voice activation.

Project: Android app for voice-activated playback of audio files

Supervisor: Ronny Lempel

Project Description: Controlling the playback of a playlist on a mobile phone while driving can be dangerous. Voice activation of commands such as “skip, back, replay” can increase driver safety. This project will develop an Android app and a backend server such that:

- The app downloads a playlist spanning a certain time duration from a server;
- The app plays the playlist, while allowing voice-activated controls;
- The app sends back to the server, for logging, any voice-activated command that was given by the user during playback;
- The server builds playlists of certain time durations and sends them to the app upon request;
- The server listens to reports of voice-activated commands by apps, and logs them.

Programming Languages and Development Platforms: the project requires Android programming (the client) and web server development (Java/Scala).
PROJECT 3

Company: Shutterfly

Project Domain: Image classification

Project name: Automatic storytelling

Supervisor: Roman Sandler

Project Description: Today's image classification tools automatically label an image with a set of more or less relevant tags. They can find people, animals, trees, and buildings. They can also deal with more abstract concepts like cruise or hike. In this project we shall use that technology to annotate albums of images with a short description.

The first step of the project will be automatic tagging of thousands of images with one or more automatic classification tools. Then the performance of these tools will be evaluated and compared. The last stage will be summarization of the image tags for an album to describe the image set as a whole.

Programming Languages and Development Platform:
- Python and PHP development in Linux at AWS.
- Work with APIs of third party tools.

PROJECT 4

Company: Shutterfly

Project Domain: Ranking of images on mobile devices

Project Name: Mobile Pictelligence

Supervisor: Eran Cohen

Project Description: Statistical ranking of images has proven to provide great value for photo based companies. Most of the current known images ranking techniques are based on backend processing, which include computer vision analysis, face detection/recognition and image quality analysis.

In this project the goal is to develop lightweight ranking system on mobile devices, provide a new curated gallery view, which shows only the better photos.

Programming Languages and Development Platforms: Mobile platform can be either IOS/Android platforms

Courses:
- Experience in IOS/Android is required
- Course in image analysis is an advantage
PROJECT 5

Company: Google

Project Domain: Gmail API, Rule engines, Machine learning

Project Name: Automatic email reply

Supervisor: Yonatan Goldhirsh

Project Description: We get so much email these days, and while we have better methods of sorting and categorizing it, there is no solution for answering it for us. Email auto response hasn’t gone far beyond vacation auto-responders. In this project, students will implement an automatic email response system over Gmail API. Gmail API is a RESTful interface to a user’s email inbox, which is useful for programmatic message sending.

For each email, the system will select one of a set of prepared responses. The big question is how will it choose. The students will first implement a rule system for choosing the appropriate response. Then, the students will implement a system to automatically create these rules based on the users mailbox.

To automatically construct the rules, students will experiment with choosing the relevant features and clustering the existing responses.

Programming Languages and Development Platforms: Java or Python - at the students’ discretion.

Courses:
- Mandatory: Algorithms 1 (234247), Operating Systems (234123)
- Advantage: Artificial Intelligence (236501), Machine Learning (236756)

PROJECT 6

Company: Google

Project domain: Big data algorithms

Project Name: Parallelizing property testing algorithms

Supervisor: Yonatan Goldhirsh

Short Description: Property testing algorithms are approximate decision making algorithms that only read a constant number of bits. That is, these algorithm only look at an extremely small part of the input, and can still (approximately) decide some property of it. Because of their extremely low complexity, these algorithms have a very local nature and therefore seem embarrassingly parallelizable. Unfortunately, the literature on property testing algorithms is almost exclusively theoretical, and there have been very few attempts to implement them.

In this project the students will implement a property testing algorithm (or several) over Hadoop MapReduce, a modern parallel computing framework. The students will then perform experiments to judge several facets of the implementation:

- Does it parallelize well?
- Does a parallel implementation outperforms a sequential one?
- Does a parallel property testing algorithm outperforms other parallel algorithms?

Property testing algorithms depend heavily on a sampling mechanism. Time permitting; the students will also experiment with implementing a sampling mechanism within Hadoop MapReduce’s sharding algorithm.

**Programming Languages and Development Platforms:**

For the algorithm implementation - at the students’ discretion, preferably one of {C++, Java, Python}.

For modifying Hadoop MapReduce - Java.

**Courses:**

- Mandatory: Algorithms 1 (234247), Operating Systems (234123)
- Advantage: Big data technology (sometimes given as 236620), Probabilistic methods and algorithms (236374)

---

**PROJECT 7**

**Company:** Apple

**Project domain:** VLSI HW verification

**Project name:** Build a formal verification kit to bridge between OCP and AHB bus protocols

**Supervisor:** Yossi Levhari

**Short Description:** In a Soc (System-On-Chip) environment there are many crossing of standard busses, these busses come from different vendors (MIPS, ARM etc), and are usually general purpose. The project is to learn the two bus standards (AHB, and OCP), define a set of formal clauses that will ensure:

- a “bridging” design, between two of such protocols obey each protocol’s rules.
- The clause need to be generic enough (i.e., add parameters)
- Tests these clauses in a Formal Verification Tool (Such as Sugar, IFV, or Jasper)

The student will learn:

- Using a Formal Verification Tool
- The notion of standard protocols, and the nuances between them
- Debug of declarative programing

**Programming Languages and Development Platforms:** Linux, Verilog

**Courses:**

- Introduction to SW testing
- Introduction to Logic
- Computer Structure/ Architecture class
PROJECT 8

Company: Apple

Project Domain: VLSI HW verification

Project Name: Fully automate the way that we handle BD memory loading

Supervisor: Yossi Levhari

Project Description: In a SOC design there are several on-chip memories used, these memories may be of various topologies, and structures. The project is to develop a system to create automatic System Verilog (OOP language) code for the loading of memories with internal knowledge of their topology.

The stage would be:
- Define the structure to define the memory topology
- Create a system Verilog task (per memory kind), to enable loading them directly (called “Back Door” in professional language)

The student will learn:
- Defining a language that will enable to describe all memories in a clean manner (BNF)
- Creation of the OOP code from this system
- SOC addressing architecture
- The way that an RTL simulation environments use these loaders

Programming Languages and Development Platforms: Programming, System Verilog

Courses:
- Introduction to Logic
- Computer Structure/ Architecture class
- OOP class

PROJECT 9

Company: Apple

Project Domain: VLSI HW verification

Project Name: Develop a GUI platform for DFT env generation

Supervisor: Yossi Levhari

Project Description: Develop a GUI platform that will support the creation of tests to many platforms, otherwise created in a manual manner. The platform will initially read the specification of the DFT structure, the GUI will serve as a Canvas environment in which the user will input the DFT command, and will be flagged for the optional parameters, and after each command, the user will have the option to execute all the commands so far to create one of the following outputs:
- Jtag script (in the jsm language which is a special language to talk with debuggers)
• Bist script (will activate existing tools)  
• Scan script (will activate existing tools)  

The student will learn:  
• Creating a GUI  
• Parsing of complex structures, and creation of an efficient DB  
• Learning how the various debugging capabilities interact with the System On a Chip  

Programming Languages and Development Platforms:  
• Linux  
• TCL-TK or Java (for GUI building)  
• Java / C / C++ - for the actual output generation  

Courses:  
• SW engineering of some sort  
• Computer Structure/ Architecture class an advantage  
• Proficient with high level programing language a must  
• GUI an advantage  

PROJECT 10  

Company: VMware  

Project Domain: operating systems, computer security  

Project Name: Automatic Windows image anomaly detection  

Supervisor: Tal Zamir  

Project Description: Windows is exposed to all kinds of malware: classic viruses, sophisticated Trojans and rootkits and all kinds of annoying toolbars and bloat-ware. These pests can significantly reduce the productivity of both home and enterprise users. Therefore, it is in the interest of IT organizations in enterprises to detect and eliminate such disturbances.  

VMware Horizon Mirage is a new kind of desktop management suite. It centralizes the full file-system of every Windows endpoint in the enterprise into the data center, while users keep running their Windows operating systems natively on their physical desktops or laptops. Once the endpoints had been centralized, IT can use Mirage to automatically update, fix and recover endpoints – both local and remote – and with minimal network and user interruption.  

We would like to leverage Mirage’s image centralization technology to automatically detect anomalies in the users’ Windows images. These anomalies can be detected by mass analysis and comparison of the file-systems and registry entries of Windows machines and finding which machines are outliers which are not following the wide consensus.  

In this project, students will create a system which will integrate with Mirage’s central store to analyze Windows machines and find these anomalies automatically, providing the IT administrator with suggestions on which applications to fix or remove on which Windows endpoints.
Programming Languages and Development Platforms:
- The project will be mainly using Microsoft technologies.
- Programming languages: C/C++, C#
- Development platform: Visual Studio 2010

Courses:
- 234123 Operating Systems
- 236363 Database Management Systems
- 236350 Computer Security

PROJECT 11

Company: VMware

Project domain: Operating Systems, Application Management, Data Mining.

Project name: “App Analyzer” - Deliver the right apps users want!

Supervisor: Yoel Calderon

Project Description:
Organizations struggle to find and deliver the right applications and tools for their employees. Often the Desktop System Administrator, maintains and delivers applications to end users is not fully aware of the actual needs of end-users nor knows which tools users need to be more productive. App Analyzer addresses this problem by providing the following features:

- Scan the desktops in the organization and retrieve user related information such as: username, organizational hierarchy, site\geographic location group memberships together with application related information like: installed apps, app usage statistics, app versions.
- Analyze the gathered information and apply data mining algorithms to create clusters of applications with similar characteristics to guide the desktop administrator which applications packages and should be delivered to which users.
- Present application usage reports and insights to the desktop administrator.
- Facilitate application delivery process by integrate with VMware Horizon Mirage application virtualization management system to

Programming Languages and Development Platforms: C++, C#

Courses:
- Operating systems
- Database Management Systems

PROJECT 12

Company: VMware

Project Domain: operating systems, computer security, networking

Project name: “Quicksight” – Smart and fast automatic IT assessment for Windows networks
Supervisor: Tal Zamir

Project Description: Enterprise IT can potentially use a vast variety of advanced tools which can ease management of Windows devices and reduce IT costs. For example: IT can potentially leverage file de-duplication tools to reduce backup storage costs; IT can use advanced desktop management tools to automatically and robustly update and fix Windows devices.

However, it is usually hard for IT decision-makers to get a sense of the value such advanced systems will really provide, which can sometimes prevent IT from making the switch to these new systems.

In this project, students will create a system which can quickly and efficiently perform an assessment of a large Windows-based network (e.g. 1,000’s of machines), providing important automatic insights on the Windows devices in the network.

The system will provide the following insights (for example):

- Total size of duplicated files in all Windows devices.
- File-system statistics for all Windows devices.
- Software inventory: installed applications and their usage statistics.
- Automatic application suggestion / recommendation.
- Which applications are common to all Windows devices (applications which should be added to a common IT “golden image”).
- Operating system versions.
- Hardware models and device drivers used by all Windows devices.

This analysis should be done in a completely automated manner and complete execution as quickly as possible (e.g. less than one hour).

Programming Languages and Development Platforms:
  o The project will be mainly using Microsoft technologies.
  o Programming languages: C/C++, C#
  o Development platform: Visual Studio 2010

Courses:

234123  Operating Systems
236334  Introduction to Computer Networks

Optional:
236363  Database Management Systems
236370  Concurrent and Distributed Programming

PROJECT 13

Company: VMware

Project domain: dev ops

Project name: Manage IT world: Using dockers with hyperic agent

Supervisor: Itzick Kasovitch, Noa Gradovitch
**Project Description:** In the environment of data centers and cloud IT, containers in general and Docker in particular became the big next thing. Docker is an open platform for developers to build and run distributed applications. Hyperic is application monitoring and performance management for virtual, physical, and cloud infrastructures. Auto-discover resources of 75+ technologies, including vSphere, and collect availability, performance, utilization, and throughput metrics.

Hyperic agents are being installed on different machines in order to supply hyperic server the relevant data.

In this project you will have the opportunity to experience Docker.

The project has two phases:
1) Use Dockers to ship and run hyperic agent.
2) Write a web based application to control those Dockers

Check out those links:
https://www.docker.com/whatisdocker/
http://sourceforge.net/projects/hyperic-hq/

**Programming languages and development platforms:** Any programming language and development platform is welcome.

---

**PROJECT 14**

**Company:** VMware

**Project domain:** operating systems, computer security, networking

**Project name:** Advanced desktop recovery with Intel® vPro and VMware Horizon Mirage

**Supervisor:** Tal Zamir

**Project Description:** The vast majority of enterprises provide users with Windows-based laptops for their day-to-day work. When a user’s laptop fails due to a software problem, it usually takes several days to get the user back to work and may require re-imaging of the laptop to solve the problem, losing user data, applications and settings in the process. The problem is even worse if the user is travelling or away from the office with his laptop, making it harder for IT to fix the problem. Intel® vPro is a built-in feature of many Intel-based laptops, which allows IT to connect to the laptop remotely, even when the Windows (or other) operating system is malfunctioning. With the right software on top of vPro, this feature can potentially help tackle malfunctioning systems and also monitor and secure them, if necessary.

VMware Horizon Mirage is a new kind of desktop management suite which provides smart disaster recovery and image management tools for both physical and virtual machines. Specifically, it applies multiple network optimization techniques for WAN, which allow these tools to work efficiently even when users are away and over low-bandwidth / high-latency networks.
In this project, students will create a system which can quickly and efficiently fix Windows machines, even when these machines are remotely located. This system will be based on Intel® vPro and VMware Horizon Mirage.

The system will have the following capabilities:
- Automatic activation of vPro in supported laptops.
- Automatic detection of laptops which support vPro in a LAN.
- Launching a minimal recovery environment on a malfunctioning Windows machine, specifically over a low bandwidth network.
- Remote wipe of a stolen machine.
- Remote fix of a malfunctioning machine.
- Transferring KVM control to a remote IT admin.

**Programming Languages and Development Platforms:**
- The project will be mainly using Microsoft technologies.
- Programming languages: C/C++, C#  
- Frameworks: .NET, vPro  
- Development platform: Visual Studio 2010

**Courses:**
234123 Operating Systems  
236334 Introduction to Computer Networks

**PROJECT 15**

**Company:** Akamai  
**Project Domain:** Big Data & Monitor  
**Project:** Yoda resource management  
**Supervisor:** Amir Skovronik  
**Project Description:**

**Background:** Yoda is our home-brew query engine - very similar to Facebook Presto / Cloudera Impala. Currently it has no resource management, thus one query can: consume all available memory; run an endless loop taking 100% CPU, etc.

**Abstract:** The project is to add resource management both in terms of memory and CPU. One way to do it is through YARN. We'll define resource pools, and for each pool we'll spawn a Yoda instance limited by memory and CPU. The project includes researching how other query engines are doing that and trying to come up with a working prototype.

**What will the student learn:** In depth knowledge of Big Data query engines and develop a big data resource management solution.

**Programming Languages and Development Platforms:** Java
PROJECT 16

Company: Akamai

Project Domain: Big Data

Project: Hive adaptor for data-catalog

Supervisor: Asaf Mesika

Project Description:

Background: The CSI platform uses its home-brew solution named Data Catalog for managing the data it saves in the Big Data platform, be it HBase, HDFS, etc. This library has two parts: Catalog definition and APIs for reading/writing data defined in the Catalog. The catalog is composed of several schemas, which in turn are composed of several tables. Each table has a record structure (much like a RDBS table) and storage properties (HBase and its configuration, HDFS, etc.) Today we have the ability read/write from each table by using: Apache MapReduce (OutputFormat, InputFormat) our very own query engine named Yoda (similar to Impala/Presto).

Abstract: Our security research team required us to supply them with SQL language for querying the data in the Big Data platform, which is the reason we installed Hive for them. Today for we manually create Hive tables, and also resort to exporting much of our Data into CSV format and define manually Hive tables on top of them. The task is to enable querying in Hive any table defined in Data Catalog.

What will the student learn: Learn about HDFS, Hive and the HDFS MetaData repository

Programming Languages and Development Platforms: Java

PROJECT 17

Company: Akamai

Project Domain: System Monitoring

Project: Log-Me

Supervisor: Uri Shamay

Project Description:

Background: Analyzing cluster behavior with hundreds to thousands of machines is a time consuming task that just doesn’t scale, it’s close to impossible to ssh to each one of them and grep logs it’s sure to drive you nuts at some stage! This is where Log-Me become your best friend, Log-Me is a system enabling you to view any log file in your cluster in one centralized location with proper UI.

What would the students learn from the project: client/server architecture, networking, scalability

Description of the final product/feature that would be built: Log-Me product will provide a web management to see the whole logs from any machine/component in your cluster with the ability to query based on machine/component/time/pattern.

Programming Languages and Development Platforms: Java
PROJECT 18

Company: Akamai

Project Domain: Big-Data

Project: HBlaze

Supervisor: Uri Shamay

Project Description:

Background: DR (disaster recovery) is hard, in distribute system even harder. Think of how much time it took Gmail to get back to life from backup when they lost hundreds of thousands accounts (hint: very short time given such huge data lose). HBlaze will be part of a DR solution for the Hadoop eco-system allowing high speed replication over WAN (Wide Area Network) between multiple HDFS systems.

What would the students learn from the project: client/server architecture, networking, scalability, TCP over UDP

Description of the final product/feature that would be built: High performance replication solution for HDFS

Programming Languages and Development Platforms: Java (maybe JNI & C++)

PROJECT 19

Company: Qualcomm

Project Domain: Data Structures, Algorithms, Signal and Image Processing

Project Name: Audiobook with Translation

Supervisor: Andrey Markovytch

Project Description: Listening to Audiobooks is great way to substitute reading process in situations where you can’t open a real book. Whether it is while driving a car, riding a bicycle, running, working out in gym or even cleaning your apartment. It’s a great way to use your time wisely. It could be even wiser if you could not just listen to an audiobook but improve your skills in foreign language. However, the main problem with listening to audiobooks in foreign language is that whenever you encounter word that you are not familiar with, you can’t really do much about it. Of course you could stop the car, bicycle or a treadmill, open up your favorite translator app and look up the word hoping that you spelled it correctly. In practice, this doesn’t really work and instead you will just find the same book in your native language and enjoy it peacefully.

The purpose of the project is to come up with simple application that will allow listening to audiobooks while being able to hear the translation of any word or phrase in the book in the language of your choice. The main requirement is for UI to be as simple and intuitive as possible with the emphasis on usage during activities. No restrictions on platform, frameworks or API’s to use, however Android + Google voice API’s + Google Translate API’s would be a good starting point.
Requested Project Milestones:

1) Students will explore and find out how APIs for voice recognition and translation available on the internet can be utilized to work with arbitrary sections inside audio file with speech. APIs should support translation of at least 3 languages one to each other. Proof-of-concept should be provided at the end of this milestone.

2) Students will design and implement UI that will enable playing audiobooks and a smart way to hear the translation of any phrase or word in the book. The emphasis should be on user experience and convenience, distracting him as less as possible from driving/sport activities. Common auxiliary devices such as headphones with controls, smartwatch and other wearables can be part of the solution as well. The implemented UI should be presented at the end of this milestone.

3) Students will combine the above 2 and create a final product – application with convenient and simple UI that can play audiobooks and translate arbitrary word or phrase in audiobook from/to at least 3 languages.

Programming Languages and Development Platforms: Any mobile OS (preferably Android), Any (preferably Java)

PROJECT 20

Company: Qualcomm

Project Domain: Operating Systems, Data Structures, Software Design

Project Name: Support for Chinese input

Supervisor: Gilad Shelef

Project Description: As most of today daily experience is accompanied with the usage of various applications on mobile phones, there are different tools for creation, design and development of the User Interface (UI).

The UI exposed to the user, usually includes input fields (e.g. name, address etc.) and adding support for Chinese characters in those fields is a challenging task.

The Chinese language comprises of 7000+ characters (AKA logograms) and there are various common ways of receiving Chinese input, the most common of which is pinyin.

This method matches a combination of Latin characters with the corresponding Chinese logogram candidates, example:

("b" "a" "i") :: ("掰" "白" "百" "佰" "摸" "摸" "柏" "败" "拜")

In this project students will design and implement an input method, which will work with an external API in order to receive and process Chinese input, as part of the UI.

Requested Project Milestones:

1) Study the basics of common Chinese input methods, with focus on Pinyin
2) Design & Implement an input method to match existing needs, while handling with environment constraints
3) Perform integration with existing APIs and conduct on-target testing

Programming Languages and Development Platforms: C
Recommended Background: Any knowledge in Chinese is definitely not a must, but will be a huge plus 😊

PROJECT 21

Company: Qualcomm

Project domain: Data Structures, Algorithms, Computer Networking, Databases

Project name: Teach for Free, Learn for Free – Social App of Knowledge Exchange

Supervisor: David Arinzon

Short description:
This project is based on a simple idea of volunteering your free time for teaching what you know best and having someone else teaching you in return. For example, Yoni is a software engineer, he is good at math, has some spare time and wants to help someone. He also has some additional time and always wanted to learn guitar. Yossi, on the other hand, has played guitar since he could remember himself, but he is clueless in math and attestation exam is coming up fast. What they both are missing is a network where they could find each other.

The purpose of the project is to come up with well-designed mobile application that will be easy to use, have fair rating system and encourage people to participate in the network. The application should allow participants to post their area of expertise and allow others to contact them to get free lessons. After each given lesson participants will earn points and get rated by their students. Combination of those two will determine the amount of free lessons they can get from other participants.

The idea has been circulating around the world for some time under general title – Time Bank.

Some resources and existing projects:
http://en.wikipedia.org/wiki/Time-based_currency
http://timebank.org.uk
http://www.bankhazman.org.il/

Requested Project Milestones:
1) Students will propose, design and implement UI, including the exact formula for user ratings vs the amount of time they can get from other participants. The rating system should be based on quality of lessons and not on the market value of the area of expertise.
2) Students will propose design and implement the server database part and define client-server communication.
3) Students will combine the above 2 and come up with the required application.

Programming Languages and Development Platforms:
Application – any, Server - any
PROJECT 22

Company: Qualcomm

Project Domain: Operating Systems, Software Design

Project Name: Vector-based fonts to bitmap renderer

Supervisor: Gilad Shelef

Project Description:
As most of today daily experience is accompanied with the usage of various applications on mobile phones, there are different tools for creation, design and development of the User Interface (UI).

Such tools may allow the designer to choose a font of his choice for various text items, during the design process, and implement them as part of the UI.

In this project, students will create a method of converting vector-based fonts (such as ttf/otf) to a bitmap representation in order to render them on the mobile device during runtime, as part of the layout.

Requested Project Milestones:
1) Study the basics of different representations and file formats of fonts
2) Study existing solutions for conversion and rendering of vector-based fonts
3) Design & Implement a method to convert vector-based fonts to a bitmap representation, while handling with environment constrains
4) Perform integration with existing APIs and conduct on-target testing

Programming Languages and Development Platforms: C

Recommended Background: Computer Graphics

PROJECT 23

Company: Microsoft

Project Domain: Machine learning

Project Name: Pattern detection and alerting system

Supervisor: Tomer Brand

Project Description: The Protection Services team owns and operates a large scale cloud data platform that enables Microsoft One Protection team and the Microsoft security suite to block and remediate malware files on Windows users’ machines. Protecting you, our users, from malware and protecting the Windows brand.

The protection platform collects up to the minute suspected malware telemetry data and samples from millions of live Microsoft anti-malware clients including System Center Endpoint Protection and Microsoft Security Essentials, as well as other Microsoft security platforms and partners. The service leverages real-time, large scale scanning, processing, analysis and classification in order to generate data insights and intelligence to help Microsoft security researchers to detect, track, block and remedy malware attacks on all
Windows devices. The protection cloud service continuously works to deliver updated protection rules to our clients and leverage the cloud logic for connected Windows devices in order to quickly address malware threats as they emerge. Helping to keep Microsoft the #1 anti-malware solution for Windows devices worldwide.

PATTERN DETECTION AND ALERTING

Protection services telemetry system receives over 40 million reports per day. Each report represent a detection on a client machine based on an anti-malware signature (which is a set of heuristic rules to identify malware related behavior). This massive stream of information holds many interesting insights in it which we would like to expose. In particular we are interesting in developing a system which supports pattern learning and detections of deviation from the normal or search for other objects with similar pattern

use case #1 – Deviation from normal

<table>
<thead>
<tr>
<th>System settings</th>
<th>1. Defining a tracking object(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. Define a period to learn the tracked object pattern</td>
</tr>
<tr>
<td></td>
<td>3. Define high and low watermarks</td>
</tr>
<tr>
<td></td>
<td>4. [P2] Define a learning reset event – cases in which the system needs to re-learn the base pattern</td>
</tr>
<tr>
<td>Learning</td>
<td>5. The system shall learn the reporting pattern of all the defined objects per the settings defined in #1-#4</td>
</tr>
<tr>
<td>Tracking &amp; alerting</td>
<td>6. The system shall track the live, incoming, feed of telemetry data per each tracking object</td>
</tr>
<tr>
<td></td>
<td>7. The system shall look for deviation from the learned pattern. Once a deviation identified the system shall issue an alert</td>
</tr>
</tbody>
</table>

The below image illustrate the requirements.

- The **green** graph represent the seasonality pattern learned by the system for telemetry reports triggered by a given signature
- The following graphs represent the tracking and comparison to the learned pattern:
  - The **red** graph represent a potential deviation from the ‘normal’ expected pattern and the point where the high watermark was crossed for a long, consecutive, enough period to issue an alert
  - The **blue** represent a similar case to the above only that this time there is a match to the low watermark
use case #2 – similar pattern

System settings
1. Defining a tracking object
2. Define a period to learn the tracked object pattern
3. Define a period of time to search for other objects with a matching pattern

Similarity search
4. The system shall learn the reporting pattern of the defined object and search for similar patterns by other objects

Programming Languages and Development Platforms
- C# (we will support the learning curve for those who knows C / C++)
- SQL like language

Courses:
- 236363 Database Management Systems
- 234107 Numerical Analysis – advantage (not mandatory)
- [234122 Introduction to Systems Programming] OR [234218 Data Structures 1] OR [Some programming experience]

PROJECT 24

Company: Rafael

Project Domain: OO Software development GUI

Project Name: Rafael software project

Project Description: Some of Rafael software projects share an advanced object oriented architecture that utilizes existing components, utilities and development methods.
Today, a developer using this architecture does so by cloning the basic components and applying them to a specific project.

The purpose of this student project is to develop a user friendly wizard that simplifies development based on this architecture.

The project outcome should provide the user a convenient method of generating new components or modifying existing components is Visual Studio Environment.

**Programming Languages and Development Platforms:** C++

**Courses:**

234122 - Introduction to Systems Programming

---

**PROJECT 25**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems

**Project Name:** CANOLA (CS) - CAN Online Listening Application (Client-Server Model)

**Supervisor:** Yan Tsitrin

**Project Description:** The interaction between Philips CT scanners components is done via CAN communication bus (Figure 1).

![Diagram of CT scanner components and CAN communication bus](image)

**Figure 1:** [a] Philips CT scanner parts (Gantry – the scanner “body”, X-ray generator, DMS – detection system, Rotor, Couch) communicates via CAN (Control Area Network) bus. They send messages of different kinds: commands and requests to other components; answers to requests sent previously by other components; state reports etc. [b] The messages are monitored by CAN-Tracer, monitoring tool developed in Philips. [c] The result of the CT scanner is a huge amount of data that the DMS transfers to the IRS; the data is reconstructed there; [d] The resulting images are passed to the HOST computer and shown there.

In the framework of the Technion Industrial Affiliates Program we have developed a tool for CAN messages monitoring – CANOLA (CAN Online Listening Application), which is, in fact, the “eyes” of the CT developers; it is used for monitoring the CT CAN messages, managing database of the collected messages and their further analysis.
The student shall be provided with the existing SW tool, CANOLA, and shall continue its development. The main project objectives are: (1) Moving from Win-forms to WPF; (2) Turn Canola to Client-Server Application (allowing to monitor Scanner’s Host remotely; (3) The monitored messages visualization enhancement.

The development shall be done in .NET, C#, using WPF and WCF technologies; initially in the simulation environment, and then on the real Philips CT-Scanners of Brilliance and iCT families.

**Programming Languages and Development Platforms:** C#
Advantage: Experience in the Embedded SW, WPF, WCF.

**Courses:**
234122 - Introduction to Systems Programming

**PROJECT 26**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems

**Project name:** ROI2DB (Regions Of Interest to Database) Application

**Supervisor:** Yan Tsitrin

**Project Description:** A picture archiving and communication system (**PACS**) is a medical imaging technology which provides economical storage of and convenient access to images from multiple modalities (X-ray plain film (**PF**), computed tomography (**CT**) and magnetic resonance imaging (**MRI**)). The universal format for PACS image storage and transfer is **DICOM** (Digital Imaging and Communications in Medicine). Non-image data, such as scanned documents, may be incorporated (using formats like PDF) once encapsulated in DICOM. However, sometimes images with incorporated textual information are available in a non-readable image format (bmp, tif etc). In such cases we are interested in a tool that is able to obtain textual information from a non-readable image (with high precision) and store it into images database along with the image itself (see Figure 1).

![Diagram of ROI2DB Application](image-url)
Figure 2: [1] Given a set of non-readable images [2] We want allow the user to prepare (manually) layout of the image in order to specify ROI-s (regions-of-interest) that later shall be given to the OCR [3] The layout is stored to the Layouts-Database, it shall be used for analyzing of images with the same layout; [4] The OCR scans the ROI-s one-by-one and store the resulting data into database.

The objectives of this project is to implement a simple and efficient method for determining regions of interest (ROI) values from a non-readable viewer using optical character recognition (OCR). The development shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification. The development shall be done in .NET, C#.

Programming Languages and Development Platforms: C#
Advantage: Experience in the Embedded SW.

Courses: 234122 - Introduction to Systems Programming

PROJECT 27

Company: Philips Medical Systems

Project Domain: Software Systems

Project Name: TEvA – Tile Evolution Application

Supervisor: Yan Tsitrin

Project Description: The Tiles are main building blocks of the DMS (Detection Measurement System). Their primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface. The detector is composed of an array of small size X-Ray detectors pixels (16x16) that’s connected to a ceramic substrate by conductive glue. The detectors substrate is connected to a mixed signal Front End Electronic (FEE) VLSI chip. 4, 6 or 8 Tiles are mounted on a detector module (depending on the system). Each detector module provides the necessary power to its Tiles and a communication channel to/from the Tiles.

Figure 3:
[a] DMS – Detection system with Modules
[b] Module with Tiles
[c] Tile – Matrix Representation
The DMS modules with the Tiles on them serve the CT-scanners for years since they are installed, and we in Philips are interested in ensuring of their reliability. Reliability demonstration testing of units under use conditions is often impractical because test duration is expected to be in the order of several years while consuming hundreds of units in test samples. Therefore, an accelerated life test must be performed under stress conditions above and beyond the expected use conditions. The results of all tests are combined and extrapolated to use conditions, assuming some time to failure distribution and life-stress relationship, collectively known as reliability model. In the testing process the tests are applied to a Tile in groups (transactions) of 5 tests in each one. After each transaction a pixel in the tile is marked as defective (of one of the tests in the applied transaction fails) or survived otherwise.

The objective of this project is to develop TEvA (Tiles Evolution Application) providing visualization of a tile evolution (i.e. changing its matrix representation from the state where all its pixels are survived up to the moment when their major part is defective). The development shall be done in .NET, C# and shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

![Figure 2: A Tile pixels matrix evolution visualization](image)

**Programming Languages and Development Platforms:** C#

**Advantage:** Background in Statistics and SW Reliability.

**Courses:**
234122 - Introduction to Systems Programming

---

**PROJECT 28**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems

**Project Name:** Tiltant-CFGM – Tile Test Analyzer Configuration Manager

**Supervisor:** Yan Tsitrin

**Project Description:** The Tiles are main building blocks of the DMS (Detection Measurement System), one of the major components of a CT-Scanner. Their primary function in the CT system is to produce digital number (frequency) corresponding to the flux of impinging X-Ray radiation on their top surface. The detector is composed of an array of small size X-Ray detectors pixels (16x16) that’s connected to a ceramic substrate by conductive glue. 4, 6 or 8 Tiles are mounted on a detector module (depending on the system).
The DMS modules with the Tiles on them serve the CT-scanners for years since they are installed, and we in Philips are interested in ensuring of their reliability. Reliability demonstration testing of units is performed under stress conditions (like, temperature and humidity) above and beyond the expected use conditions.

While analyzing these experiments results numerous parameters should be set by the test operator, among them:
- Environmental parameters (temperature, humidity etc.),
- Description of tiles under test,
- Path/Fail criteria for different kinds of test and modules types,
- Test equipment,
- Pre-existing bad detectors
- etc

At the moment these parameters are mainly hard coded in the Tiltant-SW and the tool should be recompiled in order to change a parameter.

*The objective of this project is to develop Tiltant-CFGM (Tile Test Analyzing Tool Configuration Manager), a SW tool allowing the test operator to define experiments parameters in a convenient way. After the parameters are set they shall be stored in xml format and may be used in the current and further analyzing sessions.

The development shall be done in .NET, C# and shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

**Programming Languages and Development Platforms:** C#

Advantage: XML knowledge, background in GUI development.

**Courses:**
234122 - Introduction to Systems Programming

**PROJECT 29**

**Company:** Amdocs

**Project Domain:** Web application, DB, UI , HTML5

**Project Name:** Refactoring Business DB Viewer Tool

**Supervisor:** Zahi Gil. Project FP: Oleg Shestopal

**Project Description:** DB Viewer is an existing tool which can create a dynamic view on RDBMS. This tool is currently written in Java Swing technology on top of Spring platform. We would like to refactor this Amdocs Tool from Java Swing technology to Web Application (HTML 5, JQuery, etc.).
We would like to have full functionality of the existing tool capabilities (Please see below tool main functionality).

The steps are:

1. Get familiar with the current architecture and design of the tool (architecture reverse engineering, run and learn, code review and more).
2. Design the target architecture
3. Build a work plan
4. Execute with milestones and roadshow
5. Productization (SDK and Documentation)

Current Tool Screenshot:

Tool main functionality:

- Login/security mechanism
- Designer mode and user mode (permissions and role based)
- Define Data Entities
- Define query per data entity
- Define the data to display per data entity
- Connection to multiple data source
- Changes on the fly (during run time)
- Repository for the SQL queries
- Context to search and display the data entities
- Drill down per data entity (tree hierarchy)
- Conditional formatting configurable (icons, background, and more)

**Programming Languages and Development Platforms:** Java, Swing, Spring, HTML 5, JQuery, SQL, ORACLE, Java web server knowledge and Eclipse.
PROJECT 30

Company: Amdocs

Project Domain: Information System, Development process management tool, GUI

Project Name: Interfaces Management Tool

Supervisor: Benny Cohen

Current Situation: In Amdocs, there are number of applications (AKA: Products) that are developed by different groups. Usually when solution is applied, it comprised of few products.

In the solution suggested to Amdocs customers there are interactions between these products.

Each product can consume services exposed by other products.

In the development process a request between the products to create a new communication service are managed by emails or by XLS.

There is a need to create a management tool that will manage, track, and keep version history of these requests between a list of the products. Product names should be a text parameter, the tool will enable track due dates with alerts and contract technical definition.

Project Description:

There is a need to develop management GUI application:
- The application should define product as entity and keep metadata about this entity (Name, Description, Development Manager …)
- The application should maintain list of users entitled for each product
  - List of users should be imported from Microsoft outlook, Active Directory
- The application should define a request entity and keep metadata about this entity (Request Type, Description, Document Number, Required fields, Related Flow, Comments…)
- The application should have the ability to distinguish between requests versions
- The application should define a stage entity and keep metadata about this entity (stage is a step in Software Development Life Cycle: Scope, Design, Development, Testing …)
- The application should provide a flow to attach a request from service consumer (product a) to service provider (product b) in a certain stage, and manage a statuses in the request life cycle
- The application should provide email notifications to entitled users
- The application should keep a repository in which all requests will be stored
- The application should provide easy framework to provide reports on the repository (I.e.: All Requests from provider a, All requests by consumer a, All open requests from provider a. …)
- The application should be installed in a local user Desktop/Laptop and connect to central Repository OR the application can be a web application with a central repository.
- At the beginning the application should support ~2000 users, ~20 Products, ~10 Stages, ~total of 15,000 requests

Project Deliverables:

- GUI Tool that will satisfy the above requirements
Technical Expertise required:
- Basic understanding in SOA, basic understanding in interfaces types: Web Services, EJB, File based, Html ...
- Tool can be developed in Access, java, Web architecture + HTML5 ...

PROJECT 31

Company: Amdocs

Project Domain: Automation, Generation, Java, Web, Javadoc, Rest

Project Name: **Javadoc tool for REST Data mapping layer**

Supervisor: Michael Mashian

Project Description: Multi-Channel Self Service (MCSS) is a client application built from client-side layer running on the end-user’s device (Browser, Tablet, Mobile), and a server side layer which exposes REST services that expose operations of underlying backend systems while encapsulating the complexity of their data model. In order to do so, MCSS introduces its own data model, which is designed to be simple, small in size, and UI oriented in terms of what-you-need-is-what-you-get.

In order to perform the data conversion to/from data model of various backend systems, a set of Java based mapper objects has been created. A Mapper is created per each Source & Target type duo (where Source & Target are from different data models), and its inner mapping logic is responsible to map between the (first level) attributes of the Source object to the mapped attributes of the Target object. Mappers can be reused by other mappers in case the Source & Target attribute duo is mapped by another mapper. Different mappers for similar Source & Target type duo can be distinguished by context.

The goals of the project are to create a tool that documents the underlying mapping relation of a given set of mappers, as follows:
- The tool generated a Javadoc structured document
- The document is categorized by packages and mappers (classes)
- Each mapper document describes
  - The Source & Target it represents
  - The mapping relation between the set of Source & Target attributes (primitives)
  - In case of Objects, a link to another mapper will be embedded
- The document Chained mapper hierarchy should be reflected
  - Including link to mappers within the hierarchy

Programming Languages and Development Platforms: Java programming knowledge; Familiarity with the Javadoc tool; Ability to work independently, following a given HLD; Open-minded; Innovative.
PROJECT 32

Company: Amdocs

Project Domain: Communication networks, Mobile small cell, Analytics, Network planning and design

Project Name: Mobile cell site process analysis and prediction

Supervisor: Dani Livne

Project Description:
General background: Amdocs OSS division is offering a product for managing the operators’ complex procedure of cell site deployment. This includes site visits, radio parameters setup, different authority’s approval etc. In total there are around 200 processes involved for a single site rollout, where a customer can reach hundreds of sites per day.

Project Objective: Using a network probabilistic approach, and running over sample data, students should develop software that can “learn” the behavior of processes in mobile Networks planning and design environment in order to save the operator site deployment costs.

As part of the project a short evaluation process to be performed to choose the best algorithmic approach for graph dependency evaluation.

After assigning different probabilities to the nodes in the network, different queries might be supported, for example what is the effect in process duration of adding a new step in the “network deployment” project graph chain. The project user interface should present statistics of the engine results using tables and graphs.

Responsibilities: Develop the probabilistic engine. The engine should have a continuous learning process, thus updating its internal project graph for network deployment. The engine should expose standard interface to answer network related queries.

Requirements: Programming skills in Java, probability and probabilistic reasoning background is advantage, genetic algorithms background is an advantage. For UI development HTML5 programming skill is required.

PROJECT 33

Company: Amdocs

Project domain: ATDD, TDD, Testing, Web development, JS, HTML5

Project Name: Modern Web UI Testing Suite (ATDD) with a Speck Wiki Reporter

Supervisor: Chen Osipov

Project description: A main goal is to create an advanced ATDD (Acceptance Test Driven Development) framework for testing modern web applications based on one-page application principles and JavaScript stack (such as UXF in Amdocs).

We all know this; you deploy a new feature, do updates or change something in the config somewhere. How to check if everything is as it’s supposed to be? Unit testing, functional
testing, screenshot comparison or the usual manual random clicking through pages checking for odd things?

- **Functional testing / Unit tests**
  Project specific written test are the best that you can do. They make sure key features of your site will always work. As a result the CSS can still easily degenerate over time or content elements which aren’t covered through tests can change, move or vanish completely without you noticing.

- **Screenshot comparison**
  Great for ‘real regression tests’ when nothing ever changes on your site. Try to add 'body{margin-top:2px}' or a new text paragraph/image somewhere. The result will be a mess as simple pixel by pixel comparison doesn't care for shifted or changed elements. Simple screenshot comparison just doesn't cut it in real world projects where you want to be able to deploy new features without all your tests failing massively.

- **Random clicking**
  If you are working on a medium sized client project chances are pretty high you don't have the budget to write a lot of handcrafted tests. So you end up checking the homepage, some other major pages, random subpages and maybe the search and contact form (if you are really motivated). "Looks kind of ok, lets deploy it!" If the client doesn’t call and no one complains within 48h the main part of the site seems to be working ok. CSS, layout and content degeneration are your best friends though and you know this. Therefore you hate every single update.

- **Wiki Portal**
  On the wiki portal the client or the product manager can see in real time their specs and actual tests results against those specs executed in the last build.

The suite will be assembled from existing software solutions such as: PhantomJS, CasperJS, Wiki, Jankins, Jasminie, Grunt, etc….

**Programming Languages and Development Platforms:** JavaScript must, nice to have web development knowledge (HTML 5,CSS3)

### PROJECT 34

**Company:** Amdocs

**Project domain:** Automation, Search, Algorithm, Parsing

**Project name:** Code parsing engine for test-automation

**Supervisor:** Gilli Shama

**Project Description:** Amdocs and its more than 22,000 employees serve customers in over 80 countries, with revenue of $3.6 billion in fiscal 2014. Amdocs testing tis the world leading provider of testing in the telecommunication industry.
Amdocs Testing has an in-house automation console that records and enables to edit test automation activities, and run them in series. The automation console holds a repository of test automation activities for re-use. The goals of the project are to create a smart search engine in the automation repository. The user will record a full long business flow. The search engine will find a series of activities from the repository that will best to cover the recorded business flow in terms of testing coverage. The developers team role is to define what is the best testing coverage from all the found activities for the defined business flow, review several possible algorithms for this problem, define the best algorithm to aligned this testing coverage problem, and program the algorithm (no need in front end development), present verification to the coverage problem.

Business flow example and testing activities from the repository will be provided.

**Programming Languages and Development Platforms:** Strong algorithm capabilities; Java programming knowledge; Innovative.

---

**PROJECT 35**

**Company:** IBM research Haifa

**Project Domain:** Mobile applications

**Project Name:** Anti-patterns for Mobile hybrid applications

**Project field:** Mobile software engineering, mobile optimization and analysis, iOS and Android

**Supervisor:** Aharon Abadi

**Project Description:** Mobile hybrid applications are applications designed for mobile devices and written in the same languages as web applications e.g HTML, CSS and JavaScript. In addition the hybrid applications support cross-platform and may access device special capacities using libraries such as PhoneGap. In web applications all the resources are stored in the server and transformed to the device dynamically on demand. On the other hand in hybrid application all the resources are stored locally in the device on the device file system. Many of the guidelines and recommendations about applications written in web languages design to decrease the amount of data that is transferred over the network. Our hypothesis is that such as improvement is less relevant and will not affect hybrid applications. In this project, we suggest to evaluate the hypothesis by optimizing of hybrid applications using standard tools such as Google Closure Compiler, Google PageSpeed, MobileOk etc. In addition we suggest to further explore the applications to find the most important optimizations and guidelines to this domain.

**Programming Languages and Development Platforms:** Java, Eclipse

**Required Background:** knowledge in JavaScript, CSS and HTML
PROJECT 36

Company: IBM Haifa Research Lab (HRL)

Project Domain: Multi-threading, formal verification, GUI development

Project Name: Eclipse plug-in for visualizing failures in multi-threaded C/C++ programs

Supervisor: Tatyana Veksler and Dmitry Pidan

Project Description: Nowadays, when hardware industry evolves and computational power grows rapidly, parallel software becomes very common and essential. A number of possible interleavings (context switches) grows exponentially with the number of threads, and it is a real challenge for software developer to think of all possible behaviors of a parallel software in order to avoid deadlocks and race conditions.

In HRL we develop a tool called ExpliSAT that uses formal verification techniques to find bugs in C/C++ software. ExpliSAT builds a control flow graph (CFG) of a program and performs symbolic interpretation over the CFG, where the control flow is explored explicitly and the program data is hold symbolically. In contrast with testing, where a program is tested with one particular input, ExpliSAT explores all possible behaviors of a program for all program inputs, and checks assertions in the code and performs additional built-in checks, to find bugs in the code. When ExpliSAT finds a bug in a single-threaded software, it generates a counterexample, which is a C program that injects specific program inputs that reproduce the bug. A program developer compiles the generated counterexample file with his software, and debugs the resulting executable with his favorite debugger.

In case of multi-threaded software, ExpliSAT explores all possible thread interleavings, and thus is able to find bugs that are caused by race conditions. However, a multi-threaded counterexample cannot be represented like a single-threaded one, described above, because a multi-threaded counterexample represents a specific order of thread context switches, of which we have no control in a normal run.

The purpose of the current project is developing a graphical emulator of multi-threaded counterexample that can be plugged-in into Eclipse IDE. The emulator will get as its input a multi-threaded program and a textual counterexample file, describing a specific order of context switches that demonstrates the bug in the program. The emulator GUI will be similar to a debugger, it will traverse the program step-by-step along with the counterexample, while highlighting the relevant lines in the code. The emulator will run a simulation engine in the background, and will query it to get values of relevant program variables at each step.

Programming Languages and Development Platforms: C/C++, GUI development in Java/Python, multi-processing in Linux environment

PROJECT 37

Company: IBM Haifa Research Lab (HRL)

Project Domain: Information Retrieval/Text Mining

Project Name: Wikipedia References Classification

Supervisor: Haggai Roitman
**Project Description:** Wikipedia, the popular online encyclopedia, is a large source of human curated knowledge, contributed by thousands of authors. A Wikipedia article sometimes relies on external sources, whose identity is usually annotated by a `{{ref}}` template in the article’s content (http://en.wikipedia.org/wiki/Template:Ref). Such references are located at the bottom of each Wikipedia article. The main task in this project is to classify such external references with Wikipedia articles according to their source type, e.g., study, media, blog, etc.

**Programming Languages and Development Platforms:** Java/Python; Lucene/Weka/Matlab/R/open source NLP tools

**Courses:**
- מבוא למדעי נתונים 236756
- מ抽查 למדעי נתונים 236941

**PROJECT 38**

**Company:** IBM Research - Haifa

**Project Domain:** SW Tools for Hardware Verification

**Project Name:** Static Analysis for Failure Localization in Hardware Validation

**Supervisor:** Arkadiy Morgenshtein

**Project Description:** The latest Power8 processor contains more than 4 billion transistors. How do you make sure that such huge system works correctly? Here are some facts... Ratio of design and verification engineers today is 1:1. Designer spend more than 50% of his time on verification. Debug takes 40% of overall verification effort (more time than any other task).

In the debug process, one of the most complex and time-consuming tasks is the localization of an instruction that exposes a bug detected at system level. The task is particularly difficult due to the silicon’s limited observability and the long time between a failure’s occurrence and its detection.

Recent research [1] proposed a method for automatic architectural localization of post-silicon failures by leveraging the information derived from executing the test on an Instruction Set software Simulator (ISS). The proposed method identifies a set of instructions that could lead to the faulty final state. The usage of simulator allows near-100% accuracy in failure localization, but creates a dependency in simulator and is challenging when switching to new architectures.

In this project, we explore methods of static analysis that could provide comparable accuracy levels, while eliminating the need for dynamic simulation. The proposed method will apply the failure localization algorithms to the available test case data (program trace, architectural resources and initial and final state). We will develop techniques for static analysis of the test program and for determining confidence levels for failure localization findings. The new tool will allow high-quality analysis coupled with low usage complexity and high portability for various architectures.
Project stages

Input:
- Reports derived from test execution on processor being validated, including program trace with initial and final values of resources (registers and memories). The reports also point out the resources that were identified as carrying faulty data at end of test (e.g. by comparing several test executions that should have similar results)

The analysis tool implemented in this project should include:
- Implementing a parser for the test report, deriving identity of instructions and related input/output resources
- Building a dependency graph that represents the data flow dependency between the resources of instructions in the program trace
- Enabling traversal of the graph from given resources in the final state and marking of related dependant resources
- Developing heuristics for completion of missing info (such as addresses dynamically pointed by registers during run-time) based on available initial and final values
- Implementing a confidence metric for the final results, in case that some of the info cannot be found using the heuristics

Output:
Analytics report that provides:
- A set of suspected instructions in which a failure could occur
- Confidence level metrics based on rate of dynamic data missing during static analysis

The students will be provided with example data and with complementary information and explanations that will allow applying the work to verification flow of latest PowerPC processors.

Programming Languages and Development Platforms: C++ / Java

Courses: Data Structures, Algorithms

PROJECT 39

Company: IBM Research

Project Domain: Networks

Project Name: Simulation of Flow-Policies Configuration in SDN

Supervisor: Yaniv Ben-Itzhak

Project Description: Software-defined networking (SDN) is an approach to building computer networking equipment and software that separates and abstracts elements of these systems. SDN allow network administrators to manage network services more easily through abstraction of lower level functionality into virtual services. It allows network administrators to program the flows’ routes at each router using a central control without requiring physical access to the network’s hardware devices. However, current SDN solutions don’t guarantee consistency of the flow’s routes, while dynamically changing them. Generally, it can create many out-of-order packets, leading to a collapse of the flow’s
initial throughput, especially if the switch-to-controller RTT is larger than that seen by the flow’s packets.

In this project, you will simulate and evaluate a new technique to configure the flows` routes by exploiting the SDN capabilities. This technique preserves the flow`s routes consistency and therefore results-in higher flows` throughput as compared with traditional solutions.

The simulation is based on OMNeT++, which is well known simulation framework. OMNeT++ is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. OMNeT++ offers an Eclipse-based IDE, a graphical runtime environment, and a host of other tools. OMNeT++ is a free academic version of OMNEST, which is used by R&D staff, researchers and engineers worldwide to investigate various scenarios and design alternatives. OMNEST is used by: IBM, CISCO, QUALCOMM, Mellanox, Intel, HP, Alcatel-Lucent, BROADCOM, SUN, PHILIPS and more...

http://omnetpp.org/

**Programming Languages and Development Platforms:** Eclipse, C++

**Courses:**
- 236334 - Introduction to Computer Networks
- 234122 - Introduction to Systems Programming

---

**PROJECT 40**

**Company:** IBM Research Haifa

**Project Domain:** Information retrieval, Data mining.

**Project Name:** Extracting IS-A relations from Wikipedia

**Supervisor:** Shay Hummel

**Short Description:** When I was just a little child I needed to open encyclopedia to do my homework. If my teacher asked me to define concepts I needed to search in the local library for the right book with the information needed. Today, as technology evolves, we are exposed to infinite amount of data which can help us with our day to day tasks. One of this task is defining concepts. The student will create a system which defines concepts. Given a short phrase, the system will return an "is a" sentence of the concept using Wikipedia.

For example if the user enters the term:

- "United states" is a “federal republic consisting of 50 states and a federal district”.
- "IBM" is an “American multinational technology and consulting corporation, with headquarters in Armonk, New York, United States. “

These examples are easy as the first sentence of the Wikipedia page for United States and IBM is the definition sentence. However, this can get a bit complicated for unknown pages, pages with no definitions (and the info box can be utilized) and other IR problems.

**Programming Languages and Development Platforms:** Java, C++, Weka, Lucene, Indri, Lemur or any other text analytics tool.
Courses:
- 236756 Introduction to Machine learning and
- 236299 Introduction to Natural Language Processing and
- 096262 Information Retrieval

**PROJECT 41**

**Company:** IBM Research

**Project Domain:** Social Networks, Social Analytics, Search.

**Project Name:** Real-Time Trend Detection in Social Networks

**Supervisor:** Roy Levin

**Short Description:** Detecting trending events in online social media streams has become a very popular topic in recent years. However, since the same set of trends may not interest all users to the same degree, solutions have also begun to emerge for the problem of promptly recommending trending entities within some predefined personalized context. However, the problem of finding trends related to a search query which is NOT known in advance has not yet been studied. Such a setting would allow users to promptly get relevant trends as they issue a new search query over a dataset of documents. For example, recommending trending applications when a user searches for "children games". This presents a new challenge as, unlike existing methods, there is no predefined context for which relevant trends can be calculated in advance.

We have designed the outline of an algorithm for detecting such trends, in real time, over streaming social media data within the context of a search query. Using real data from an online social network, we wish to evaluate the algorithm based on its ability to predict an actual increase in the level of activity within entities that are found to be trending. In the experiments we wish to show an increased ability to predict an actual activity growth within the context of user searches and to examine their meaningfulness.

**Programming Languages and Development Platforms:** Java, Python or similar

Courses:
- Introduction to Systems Programming (234122)
- Algorithms 1 (234247)
- Data Structures 1 (234218)

**PROJECT 42**

**Company:** IBM Research – Haifa Lab

**Project Domain:** Computer vision, image/video processing

**Project Name:** Real-time moving object detection in aerial video

**Supervisor:** Dror Porat and Udi Barzelay

**Project Description:** Moving object detection in aerial video is a challenging computer vision task. The goal of the detection algorithm is to automatically identify motion in a video, e.g.,
in order to provide focus of attention or to serve as a precursor to subsequent vision tasks (such as object tracking). While traditional moving object detection algorithms typically assume a stationary camera, in the case of aerial video the video is captured by a freely moving camera, which poses a real challenge to reliably detect moving objects in the scene. In this project we aim to develop an efficient and robust moving object detection algorithm that is expected to work in real-time and to provide accurate results in various scenarios. The algorithm will be implemented using popular and useful C/C++ computer vision libraries, such as OpenCV (but no prior knowledge in using such libraries is necessary). The performance of the detection algorithm will be evaluated on real-world aerial videos acquired from a UAV (unmanned aerial vehicle).

NOTE: students must sign a special IBM-Technion agreement to participate

**Programming Languages and Development Platforms:** C++

**Courses:** Software Engineering Methods (234321) or Object Oriented Programming (236703), Digital Image Processing (236860)
Recommended: Computer vision (236873)

### PROJECT 43

**Company:** IBM Haifa Research Lab

**Project Domain:** verification and testing, software engineering

**Project Name:** Finding security vulnerabilities in applications using symbolic interpretation of C/C++ source code

**Supervisor:** Dmitry Pidan

**Project Description:** In today’s world of networks everywhere, security vulnerabilities become a major issue, and their early detection is gaining more and more attention. Frequently, a very harmful security vulnerability may result from a tiny error in the implementation of a program, and still remain undetected for a long time. One of the recent examples for such a vulnerability is a famous Heartbleed bug in the popular OpenSSL cryptographic software library (http://heartbleed.com).

Contrary to functional bugs, security vulnerabilities are much harder for detection using standard testing procedures for many reasons. One reason is that exploitation scenarios are very unusual and counter-intuitive for testcase developer. Another reason is the invisibility of the bug – when the vulnerability is exploited, usually no visible symptom (e.g. program crash) is available, however behind the scenes very bad things, like sensitive data leakage, happen.

To overcome the complexities described above, novel and sophisticated techniques for the program verification are required. IBM develops ExpliSAT – a tool for symbolic interpretation of C/C++ code. The purpose of the tool is, given an application under test, to explore the feasible execution paths of the application using symbolic data – a data that means “every possible value in the domain”, rather than a particular value as it is done in testing. Recent experimentations with the tool demonstrated its high potential in finding corner case bugs, and in particular security vulnerabilities like Heartbleed.
In the scope of the project, the students will take the ExpliSAT tool and explore with it the source code of security-critical application (e.g. OpenSSL), aiming at finding previously unknown vulnerabilities. Bugs that will be found will be reported to the application owners, and high-quality vulnerabilities (e.g. those that can lead to harmful results if exploited) will be published in the security community. From a technical perspective students will

- Learn how to work with the ExpliSAT tool and how to use it for performing qualified verification of the source code.
- Develop sophisticated verification environment for the application under test (using C/C++)

**Programming Languages and Development Platforms:** C, C++, Linux

**Courses:**
- 234218 – Data Structures 1
- 234247 – Algorithms 1

**PROJECT 44**

**Company:** IBM

**Project domain:** Machine Learning

**Project name:** Deep learning algorithms implementation and validation

**Supervisor:** Moran Gavish

**Project Description:** Deep learning is a set of machine learning algorithms that are inspired by a class of theories of brain development and typically focus on modeling high-level abstractions in data. Recently published papers present exciting results on numerous machine learning tasks in various application domains. The aim of this project is to build a prototype implementation of deep learning algorithms. We suggest following a tutorial given by Stanford University, which provides good basis for this task. The first phase would be creating a prototype implementation in Matlab or Python. The second phase would be of testing, validating and tuning the algorithms by creating system aimed at solving real world machine learning problems.

**Optional:** A follow up project aimed at creating efficient implementation for the developed algorithms in JAVA, may be considered upon successful completion of this project.

**Programming Languages and Development Platforms:** Matlab, Python or Java

**Courses:**
- שיטות סתברותיות ואלגוריתמים (236374)
- מבוא ליבנה מולכותית (236501)
- מבוא למערכות לומדות (236756)
- מטריצות לומדות (046195)
- מבוא לResolve עבירות (236941)
- студентים מצטיינים לשנה האחרות לה箢יכ עניקיימ ב-
- machine learning

Page 37 of 40
**PROJECT 45**

**Company:** LogicBlox  

**Project Domain:** Databases, Data Mining, Machine Learning  

**Project Name:** Feature Generation for Prediction over a Database  

**Supervisor:** Benny Kimelfeld  

**Project Description:** LogicBlox is a database product designed for enterprise software development, combining transactions and analytics. The underlying data model is a relational database, and the query language, LogiQL [1], is an extension of Datalog. As such, LogiQL features a simple and unified syntax for traditional relational manipulation as well as deeper analytics. Moreover, its declarative nature allows for substantial static analysis for optimizing evaluation schemes, parallelization, incremental maintenance, and sophisticated transactional management [2]. As part of its support for analytics, LogiQL has a collection of built-in collection of Machine Learning (ML) algorithms.

The goal of this project is to design and implement an automatic feature generator for ML algorithms, given a database in LogicBlox, by analyzing both the schema (LogiQL specification) and the data. The input consists of a collection of objects in the database, and training data, that is, pairs of the form \((o, y)\) where \(o\) is an object and \(y\) is the value that we wish to predict. The goal is to design a function that maps every object \(o\) into a vector \(x(o)\) of reals by applying database queries (involving joins and aggregates). The techniques include both logical reasoning and pattern mining.

The specific ML class in consideration is Factorization Machines (FM), a recent and successful model for predictors that operate over real-valued vectors (similarly to SVM) for either regression or classification. FM captures correlations among pairs of features, and can be learned and evaluated efficiently. A plethora of applications include recommender systems and retail predictions. To implement and evaluate the feature generator, LogicBlox will provide a specific retailer use case, including the database, LogiQL specification, and a goal function.


**Programming Languages and Development Platforms:**  
- Scala / Haskell  
- LogicBlox
Courses:
- 236363 Database Management Systems
- 236756 Introduction to Machine Learning

**PROJECT 46**

**Company:** LogicBlox

**Project Domain:** Databases, Artificial Intelligence

**Project Name:** Translating Soft Logic into LogicBlox Optimization

**Supervisor:** Benny Kimelfeld

**Project Description:** LogicBlox is a database product designed for enterprise software development, combining transactions and analytics. The underlying data model is a relational database, and the query language, LogiQL [1], is an extension of Datalog. As such, LogiQL features a simple and unified syntax for traditional relational manipulation as well as deeper analytics. Moreover, its declarative nature allows for substantial static analysis for optimizing evaluation schemes, parallelization, incremental maintenance, and sophisticated transactional management [2]. As part of its support for analytics, LogiQL has an intuitive interface for Mixed Integer Programming (MIP), by means rules over relational predicates with attributes marked as variables or objectives.

Probabilistic-logic concepts, such as Markov Logic Networks (MLN) [3] and Probabilistic Soft Logic (PSL) [4], are logical formalisms that allow for soft interpretation of rules. The canonical example is “R(x) ← R(y) and Friends(x, y),” where R describes a person property (e.g., “smokes” or “votes”); this rule states that whenever x and y are friends, the property R propagates from y to x, and such as rule should be taken as a hint on the unknown, and not as rigid truth. While an ordinary Datalog program specifies a unique extension of the database, probabilistic logic specifies a probability space over such extensions. Intuitively, the probability of a possible extension is determined by the extent to which the rules are satisfied (where weights of rules are taken into account). MLN and PSL have been used for solving a variety of computational problems, such as textual information extraction, entity resolution, predictions of drug and protein functionality, and robotics.

The goal of this project is to implement a translation of MLN and PSL into LogiQL. More specifically, the project involves parsing an input program in one of the formalisms for probabilistic logic, translation into LogiQL, and invocation the resulting LogiQL program towards Maximum A-Priori (MAP) inference. Moreover, the project will compare the evaluation of the implementation against existing implementations (e.g., Alchemy or Tuffy), in terms of quality and execution cost.

In Int Conf on Datalog in Academia and Industry, 2012.

**Programming Languages and Development Platforms:**
- Java / C++ / Python / Scala / Haskell
- LogicBlox

**Courses:**
- 236363 Database Management Systems
- 236330 Introduction to Optimization
- 236501 Introduction to Artificial Intelligence (recommended)