# INDUSTRIAL PROJECTS 234313
## SPRING 2016

<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>Web application, DB, UI, HTML5</td>
<td>Refactoring business DB viewer tool</td>
<td>Amdocs</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Big Data, analytic, diagnostic, REST</td>
<td>Big Data - diagnostic package collector</td>
<td>Amdocs</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Optimization, algorithm, Real Time Analytics, REST, Cloud, energy optimization</td>
<td>Energy aware cloud workload placement</td>
<td>Amdocs</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>Full stack web development, SDN, Cloud, NFV</td>
<td>Lab manager</td>
<td>Amdocs</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>Social tool</td>
<td>Managing proposals for efficiency process</td>
<td>Amdocs</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Open Source, WebRTC, Unified communication, IoT</td>
<td>matrix.org open source bridges – communication hubs</td>
<td>Amdocs</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>Open Source, WebRTC, Unified communication, IoT</td>
<td>matrix.org open source bridges – management tools</td>
<td>Amdocs</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Open Source, WebRTC, Unified communication, IoT</td>
<td>matrix.org open source bridges – media services</td>
<td>Amdocs</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Open Source, WebRTC, Unified communication, IoT</td>
<td>matrix.org open source bridges - messaging services</td>
<td>Amdocs</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Automation, Monitoring, patching, NLP</td>
<td>My Instant Assistant (MIA)</td>
<td>Amdocs</td>
<td>16</td>
</tr>
<tr>
<td>11</td>
<td>Data Management, Dashboard, Telecom, User experience</td>
<td>Communication data consumption personal dashboard</td>
<td>Amdocs</td>
<td>19</td>
</tr>
<tr>
<td>12</td>
<td>Data Management, Quality, Telecom</td>
<td>Data Validation Rules Definition</td>
<td>Amdocs</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>Data Science, Analytics, Tableau, Testing, Performance Management</td>
<td>Test execution trend prediction</td>
<td>Amdocs</td>
<td>21</td>
</tr>
<tr>
<td>14</td>
<td>Web, UI</td>
<td>Self SMM Tool</td>
<td>Amdocs</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>Social Media, algorithms, User Interface, DB, Open APIs, Graph</td>
<td>Finding k-most influencing nodes in an interactions graph</td>
<td>Amdocs</td>
<td>23</td>
</tr>
<tr>
<td>No.</td>
<td>Field</td>
<td>Project Description</td>
<td>Company/Institution</td>
<td>Page</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------</td>
</tr>
<tr>
<td>16</td>
<td>Big data analytics and visualization</td>
<td>User Controlled Diversity of Received Content Recommendations</td>
<td>Outbrain</td>
<td>24</td>
</tr>
<tr>
<td>17</td>
<td>NLP, Chat bot programming</td>
<td>Chat bot able to provide personal content recommendations</td>
<td>Outbrain</td>
<td>25</td>
</tr>
<tr>
<td>18</td>
<td>Dynamic memory allocation, heap management, OS runtime</td>
<td>Modifying and optimizing dynamic memory allocation libraries and measuring the side effects</td>
<td>Intel</td>
<td>26</td>
</tr>
<tr>
<td>19</td>
<td>Computer security</td>
<td>Anomaly malware detection</td>
<td>Intel</td>
<td>27</td>
</tr>
<tr>
<td>20</td>
<td>Computer security</td>
<td>Integrated code review environment</td>
<td>Intel</td>
<td>27</td>
</tr>
<tr>
<td>21</td>
<td>Computer security</td>
<td>Finding vulnerabilities through fuzzing in SGX Enclave</td>
<td>Intel</td>
<td>28</td>
</tr>
<tr>
<td>22</td>
<td>Computer security</td>
<td>Genetic fuzzing</td>
<td>Intel</td>
<td>29</td>
</tr>
<tr>
<td>23</td>
<td>Computer security</td>
<td>Translator of x64 code to LLVM IR</td>
<td>Intel</td>
<td>30</td>
</tr>
<tr>
<td>24</td>
<td>Code Parsing, DB and graphics</td>
<td>Logic Visualizer</td>
<td>Rafael</td>
<td>30</td>
</tr>
<tr>
<td>25</td>
<td>Software Systems, Medical Devices, Monitoring Automation, Big Data</td>
<td>AUDIT – Advanced Usability Diagnostic Tool</td>
<td>Philips Medical Systems</td>
<td>31</td>
</tr>
<tr>
<td>26</td>
<td>Software Systems, Medical Devices</td>
<td>CT- REST Remote Embedded Scripting Tool</td>
<td>Philips Medical Systems</td>
<td>33</td>
</tr>
<tr>
<td>29</td>
<td>Software Systems, GUI, Visualization</td>
<td>TEvA – Tile Evolution Application</td>
<td>Philips Medical</td>
<td>38</td>
</tr>
<tr>
<td>Page</td>
<td>Domain</td>
<td>System</td>
<td>Company</td>
<td>Page</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>31</td>
<td>Machine Learning, Artificial Intelligence, Combinatorial Optimization</td>
<td>Statistical relational learning via factorization</td>
<td>LogicBlox</td>
<td>41</td>
</tr>
<tr>
<td>32</td>
<td>Machine Learning</td>
<td>Deep Learning with Factorization Machines</td>
<td>LogicBlox</td>
<td>44</td>
</tr>
<tr>
<td>33</td>
<td>Natural Language Processing</td>
<td>Artemis</td>
<td>HPE Enterprise</td>
<td>46</td>
</tr>
<tr>
<td>34</td>
<td>Big Data, Predictive Analytics</td>
<td>Predictive Analytics for Developers</td>
<td>HPE Enterprise</td>
<td>47</td>
</tr>
<tr>
<td>35</td>
<td>Operating Systems, Mobile</td>
<td>Android Threads Analysis Viewer</td>
<td>HPE Enterprise</td>
<td>48</td>
</tr>
<tr>
<td>36</td>
<td>Big Data processing</td>
<td>Portero</td>
<td>HPE Enterprise</td>
<td>49</td>
</tr>
<tr>
<td>37</td>
<td>Mobile applications</td>
<td>Windows Phone Agent for AppPulse Mobile</td>
<td>HPE Enterprise</td>
<td>50</td>
</tr>
<tr>
<td>38</td>
<td>Data Analytics</td>
<td>Gaia</td>
<td>HPE Enterprise</td>
<td>50</td>
</tr>
<tr>
<td>39</td>
<td>Big Data Analytic</td>
<td>Social Networks Data Collector</td>
<td>HPE Enterprise</td>
<td>51</td>
</tr>
<tr>
<td>40</td>
<td>Real time embedded systems</td>
<td>Bring Up Real time Lua + TLSF SW development environment for real time embedded mission critical avionic software applications</td>
<td>Elbit Systems</td>
<td>52</td>
</tr>
<tr>
<td>41</td>
<td>UI designer\Web &amp; DB application, Optimization</td>
<td>Verification dashboard</td>
<td>Elbit Systems</td>
<td>54</td>
</tr>
<tr>
<td>42</td>
<td>Programming</td>
<td>Stack-Analysis Tool</td>
<td>Elbit Systems</td>
<td>54</td>
</tr>
<tr>
<td>No.</td>
<td>Domain</td>
<td>Project Details</td>
<td>Company</td>
<td>Page</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>43</td>
<td>Compilers</td>
<td>Extend C/C++ Compiler – Static Code Analysis</td>
<td>Elbit Systems</td>
<td>55</td>
</tr>
<tr>
<td>44</td>
<td>Software Development Process, Programming, IDE</td>
<td>Eclipse plugin for Development-Process integration</td>
<td>Elbit Systems</td>
<td>56</td>
</tr>
<tr>
<td>45</td>
<td>HTML5, AngularJS, Mobile App</td>
<td>Advanced Hybrid mobile App framework</td>
<td>Elbit Systems</td>
<td>57</td>
</tr>
<tr>
<td>46</td>
<td>Networking</td>
<td>Real Time Video Streaming</td>
<td>Elbit Systems</td>
<td>58</td>
</tr>
<tr>
<td>47</td>
<td>Hierarchical clustering, Spark</td>
<td>Hierarchical Clustering Using Spark</td>
<td>IBM</td>
<td>59</td>
</tr>
<tr>
<td>49</td>
<td>Computer Vision, Machine Learning</td>
<td>Classification of medical images using semi-supervised learning</td>
<td>IBM</td>
<td>61</td>
</tr>
<tr>
<td>50</td>
<td>Web Applications, Cloud Systems, Platform-as-a-Service, Networking, Security</td>
<td>Cloud Binding for... Open-Source Web-Application Server Frameworks</td>
<td>IBM, Check Point</td>
<td>62</td>
</tr>
<tr>
<td>51</td>
<td>Logic, Artificial Intelligence</td>
<td>Software for encoding satisfiability problems into conjunctive normal form</td>
<td>IBM</td>
<td>63</td>
</tr>
<tr>
<td>52</td>
<td>3-tier application development, Databases, Web-based technology</td>
<td>Translyzer - A 3-tier analyzer &amp; visualizer for address-translation data</td>
<td>IBM</td>
<td>64</td>
</tr>
<tr>
<td>53</td>
<td>Cloud, PAAS, Docker</td>
<td>Video surveillance service using FFMPEG and Docker on IBM Bluemix cloud</td>
<td>IBM</td>
<td>65</td>
</tr>
<tr>
<td>54</td>
<td>Computer Security</td>
<td>Malware recognition with binary fingerprinting</td>
<td>Check Point</td>
<td>66</td>
</tr>
<tr>
<td>55</td>
<td>NLP (Natural Language Processing) &amp; Machine Learning</td>
<td>Semi-supervised learning NLP engine for food delivery and groceries ordering IoT system</td>
<td>Say2eat</td>
<td>68</td>
</tr>
</tbody>
</table>
PROJECT 1

Company: Amdocs

Project Domain: Web application, DB, UI, HTML5

Project: Refactoring business DB viewer tool

Supervisor: Zahi Gil

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 reveres engineering, run and learn, code review and more).
2. Design the target architecture
3. Build a work plan
4. Execute with mile stones and roadshow
5. Productization (SDK and Documentation)

Programming Languages and Development Platform: Java, Swing Spring, HTML 5, JQuery, SQL, ORACLE, Java web server knowledge and Eclipse.

Required Background
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)

---

**PROJECT 2**

**Company:** Amdocs

**Project Domain:** Big Data, analytic, diagnostic, REST

**Project:** Big Data - diagnostic package collector

**Supervisors:** Idan Raphael and Menachem Kaplan

**Project Description:** Big data analytics (BDA) is an Amdocs framework for providing data centric applications. It is based mainly on Hadoop and its standards ecosystems tools. The aim of the project is to provide a diagnostic package collector in order to create a utility that enable to collect diagnostic information from the Hadoop cluster and pack it in an archive, for two purposes:

1. Offline analysis of failures
2. Analysis of performance bottlenecks

The tool should receive a timeframe to work on, and a diagnostic level (verbose / trace / error) Then it should connect to the Hadoop cluster and other components (Elastic search cluster for example), and pull the diagnostic information from it using various protocols (mostly REST and java clients, but may be others, such as SSH / SFTP, etc.)
After pulling the data additional processing might be required (in error level, for example, error messages should be extracted from the logs)
Then the diagnostic information need to be stored in a hierarchical structure and packed in an archive
The tool should also be extendable with custom collectors, for future use, so, if for example data need to be collected from a new system, the user can inject new collector that the tool will use in order to collect data from the new system
Implementation should be done with Java or Groovy

**Programming Languages and Development Platforms:** Java and/or Groovy

---

**PROJECT 3**

**Company:** Amdocs

**Project Domain:** Optimization, algorithm, Real Time Analytics, REST, Cloud, energy optimization

**Project:** Energy aware cloud workload placement

**Supervisor:** Hemed Gur Ary

**Project Description:**

**Overview** - The solution will enable cloud energy optimization by recommending the right hosts to run workloads.

**Background** - Cooling datacenters is an expensive issue. It is estimated that in a traditional data center 43% of energy costs is spent on cooling. In order to lower cost organization usually work on two vectors: facility planning and hardware optimization. We propose that in cloud infrastructure there’s a new vector to tackle: application optimization.

The solution will be used as a policy provider to the cloud automation orchestrator. The orchestrator will turn to the solution for placement recommendation on physical host.

**Main Components:**

**Database** – Holds application data

**Data Collector** – collects sensor and metadata data and models the data to the database

**Real Time Analytics Engine** – analyze data in database and create recommendations

**API Provider** – provides REST API for external entities such as cloud orchestrator

**Proposed design**

Below is an initial abstract design of the solution:
Recommended Background:
Must: Python, REST API and Statistics
Preferred: Cloud/OpenStack and Real Time Analytics

PROJECT 4

Company: Amdocs

Project Domain: Full stack web development, Software Defined Network (SDN), Cloud, Network Functions Virtualization (NFV)

Project: Lab manager

Supervisor: Igor Shakhman

Project Description: Web based application with DB backend which manages lab operation. Including but not limited to: physical equipment inventory, continuous integration system, versions control and configuration manager. Lab Manager allows to create testing and development environments on demand. Lab Manager uses OpenStack Virtual Infrastructure Manager for virtual instances management and Open Contrail or Open Daylight Software Defined Network controller for network management.

Background: Amdocs Network solutions (ANS) NFV team activities are related to Network Functions Virtualization projects with following incubation of developed solutions. Many different POCs (Proof of Concept) run in lab which requires changes of lab’s layout to occur often. The changes need to be done fast.
Lab Manager main functions:
1. Manage physical equipment inventory
2. Deploy new Dev/QA setups
3. Run different test cases (TC) and TC suits on created setups
4. Snapshot setup state to repository
5. Reset setup.
6. Destroy setup.
7. Restore setup from snapshot
8. Solution environment blue-print designer
9. Dashboards presenting basic metrics

Main Components:
1. Web based application
2. Database

Systems to be integrated with and to manage:
1. Jenkins
2. Git
3. Ansible or Puppet
4. Nagios
5. OpenStack
6. Open Contrail/Open DayLight
7. Physical servers: HP Proliant XL series
8. Physical switches HP 5500 and 5900 series

Programming Languages and Development Platform:
Must: Web UI design, Python, REST API, Bash and MySQL
Preferred: XML, Git, Jenkins, Puppet or Ansible

PROJECT 5

Company: Amdocs
Project Domain: Social tool
Project: Managing proposals for efficiency process
Supervisors: Tomer Simon and Arik Rizer
Project Description: The collaborative platform will be Yammer, an enterprise social network, through which the proposals will be submitted and it will be the place for rating and conversations around them.

For the steering committee there is a need to develop a module that will be displayed as layer on top of Yammer, that would present information, allow to update details and give analytical insights. The system will display the information through the Yammer APIs and add additional parameters for management and control.

System Requirements:

- Connect through Yammer APIs for information retrieval
- Providing a unique number for every suggestion
- Display the time from the publication of the proposal - Tracing
  - The number of days from the posting
  - Allows user to determine the number of days until comment (SLA)
  - warning/ notification before the failure to comply the service level
  - special indicator when passing the specified time
- Build a dashboard, an aggregator that collects and displays all the information
- Possibility to communicate between members of the steering committee
- A final answer and the closure proposal
- Allowing generate reports in different segments

Background: Today, companies use various systems to capture ideas for improvement and innovation from their employees. Through them employees can publish a proposal for efficiency, and usually all communications around the process and idea will be by email. In order to advance the innovation in the organization there’s a need to build a managed collaborative arena where employees can share their suggestions and get feedback from their colleagues. The steering committee will have dashboard who will aggregate all the relevant information, and help to manage the process.

Recommended Background: Any programming language, Web development preferred for the dashboard/managerial module.
PROJECT 6

Company: Amdocs

Project Domain: Open Source, WebRTC, Unified communication, IoT

Project: matrix.org open source bridges – communication hubs

Supervisor: Yuval Lib

Project Description: Join and contribute to a cutting edge open source project community by developing bridges to external widely used leading communication hubs.

Background: Matrix is an open standard for interoperable, decentralized, real-time communication over IP. It can be used to power Instant Messaging, VoIP/WebRTC signaling, Internet of Things communication - or anywhere you need a standard HTTP API for publishing and subscribing to data while tracking the conversation history.

Matrix defines the standard, and provides open source reference implementations of Matrix-compatible Servers, Clients, Client SDKs and Application Services to allow anyone to create new communication solutions or extend the capabilities and reach of existing ones.

More details on Matrix can be found here: http://matrix.org/

Glimpse of the Matrix community sprit can be found in the Matrix blog: http://matrix.org/blog/2015/12/25/the-matrix-holiday-special/

Goal: Matrix goal is to act as a generic HTTP messaging and data synchronization system for the whole web and mobile: allowing people, services and devices to easily communicate with each other, empowering users to own and control their data and select the services and vendors they want to use.

Your Mission: In order to achieve that Matrix must seamlessly connect to other communication hubs. Our goal is to have bridges/connections/integrations between matrix and as many communication hubs as possible.

Your mission is to implement such a communication hubs bridge:

- Join the matrix community of topnotch open source programmers
- Design, develop and deploy a new bridge between Matrix and another management service, such as:
- Zapier - automate tasks between web apps
- Hubot - an automation tool that can sync with other chat services
- Nagios – open-source application that monitors systems, networks and infrastructure
- IFTTT - Create simple connections between the products you use every day (bridge the bridge)

(Of course you are free to suggest another communication hub – it is up to you to convince the community that it is necessary)

We will mentor you and guide you throughout the process. Your success is our success and we will make sure you will excel (and enjoy the journey)!

**Programming Languages and Development Platform:** Web programming, web services

---

**PROJECT 7**

**Company:** Amdocs

**Project Domain:** Open Source, WebRTC, Unified communication, IoT

**Project:** matrix.org open source bridges – management tools

**Supervisor:** Yuval Lib

**Project Description:** Join and contribute to a cutting edge open source project community by developing bridges to external widely used leading management tools

**Background:** Matrix is an open standard for interoperable, decentralized, real-time communication over IP. It can be used to power Instant Messaging, VoIP/WebRTC signaling, Internet of Things communication - or anywhere you need a standard HTTP API for publishing and subscribing to data while tracking the conversation history.

Matrix defines the standard, and provides open source reference implementations of Matrix-compatible Servers, Clients, Client SDKs and Application Services to allow anyone to create new communication solutions or extend the capabilities and reach of existing ones.

More details on Matrix can be found here: [http://matrix.org/](http://matrix.org/)

Glimpse of the Matrix community spirit can be found in the Matrix blog: [http://matrix.org/blog/2015/12/25/the-matrix-holiday-special/](http://matrix.org/blog/2015/12/25/the-matrix-holiday-special/)
Goal: Matrix goal is to act as a generic HTTP messaging and data synchronization system for the whole web and mobile: allowing people, services and devices to easily communicate with each other, empowering users to own and control their data and select the services and vendors they want to use.

Your Mission: In order to achieve that Matrix must seamlessly connect to other management tools. Our goal is to have bridges/connections/integrations between matrix and as many management tools as possible.

Your mission is to implement such a bridge:
- Join the matrix community of top-notch open source programmers
- Design, develop and deploy a new bridge between Matrix and another management service, such as:
  - Basecamp - web-based project-management tool
  - Trello - Trello keeps track of everything, from the big picture to the minute details
  (Of course you are free to suggest another management tool – it is up to you to convince the community that it is necessary)
  We will mentor you and guide you throughout the process. Your success is our success and we will make sure you will excel (and enjoy the journey)!

Recommended Background: Web programming, web services

---

PROJECT 8

Company: Amdocs

Project Domain: Open Source, WebRTC, Unified communication, IoT

Project: matrix.org open source bridges – media services

Supervisor: Yuval Lib

Project Description: Join and contribute to a cutting edge open source project community by developing bridges to external widely used leading media bridges.

Background: Matrix is an open standard for interoperable, decentralized, real-time communication over IP. It can be used to power Instant Messaging, VoIP/WebRTC signaling,
Internet of Things communication - or anywhere you need a standard HTTP API for publishing and subscribing to data while tracking the conversation history. Matrix defines the standard, and provides open source reference implementations of Matrix-compatible Servers, Clients, Client SDKs and Application Services to allow anyone to create new communication solutions or extend the capabilities and reach of existing ones.

More details on Matrix can be found here: http://matrix.org/

Glimpse of the Matrix community spirit can be found in the Matrix blog: http://matrix.org/blog/2015/12/25/the-matrix-holiday-special/

**Goal:** Matrix goal is to act as a generic HTTP messaging and data synchronization system for the whole web and mobile: allowing people, services and devices to easily communicate with each other, empowering users to own and control their data and select the services and vendors they want to use.

**Your Mission:** In order to achieve that Matrix must seamlessly connect to external media services. Our goal is to have bridges/connections/integrations between matrix and as many services as possible.

Your mission is to implement such a bridge:

- Join the matrix community of topnotch open source programmers
- Design, develop and deploy a new bridge between Matrix and another media services, such as:
  - Giphy - search, share, and discover the world’s best GIFs
  - Google Image Search - Comprehensive image search by Google.

(Of course you are free to suggest another management tool – it is up to you to convince the community that it is necessary)

We will mentor you and guide you throughout the process. Your success is our success and we will make sure you will excel (and enjoy the journey)!

**Recommended Background:** Web programming, web services
PROJECT 9

Company: Amdocs

Project Domain: Open Source, WebRTC, Unified communication, IoT

Project: matrix.org open source bridges - messaging services

Supervisor: Yuval Lib

Project Description: Join and contribute to a cutting edge open source project community by developing bridges to external widely used leading messaging services.

Background: Matrix is an open standard for interoperable, decentralized, real-time communication over IP. It can be used to power Instant Messaging, VoIP/WebRTC signaling, Internet of Things communication - or anywhere you need a standard HTTP API for publishing and subscribing to data while tracking the conversation history.

Matrix defines the standard, and provides open source reference implementations of Matrix-compatible Servers, Clients, Client SDKs and Application Services to allow anyone to create new communication solutions or extend the capabilities and reach of existing ones.

More details on Matrix can be found here: http://matrix.org/

Glimpse of the Matrix community spirit can be found in the Matrix blog: http://matrix.org/blog/2015/12/25/the-matrix-holiday-special/

Goal: Matrix goal is to act as a generic HTTP messaging and data synchronization system for the whole web and mobile: allowing people, services and devices to easily communicate with each other, empowering users to own and control their data and select the services and vendors they want to use.

Your Mission: In order to achieve that Matrix must seamlessly connect to other messaging services. Our goal is to have bridges/connections/integrations between matrix and as many messaging systems as possible.

Your mission is to implement such a bridge:
  • Join the matrix community of topnotch open source programmers
• Design, develop and deploy a new bridge between Matrix and another messaging service, such as:
  - HipChat - XMPP based group chat and video chat built for teams.
  - Whatsapp - cross-platform mobile messaging app
  - Twitter - online social messaging supporting 140-character messages called "tweets"
  - Facebook Messenger – social texting app
(Of course you are free to suggest another messaging system – it is up to you to convince the community that it is necessary)

We will mentor you and guide you throughout the process. Your success is our success and we will make sure you will excel (and enjoy the journey)!

**Recommended Background:** Web programming, web services

---

**PROJECT 10**

**Company:** Amdocs

**Project Domain:** Automation, Monitoring, parching, Natural language processing (NLP)

**Project:** My Instant Assistant (MIA)

**Supervisor:** Arkadiy Kats

**Project Description:**

**Main idea:** With the need of our customers for immediate support during and after the official working hours, we would like to offer the MIA system.

MIA system is 24X7, 365 availability, e-mail based system, which handles various tasks and provides online updates.

MIA will receive e-mail messages in human language, parse them, and answer back accordingly to the request in the incoming e-mail.

The reaction information will be taken from an up-to-date commands database, and will be sent back to the end-user via reply email.

Moreover MIA will also perform remote server/computer operations accordingly to the request.
MIA will act as an automation system that will spare from employees the need to remotely connect to their own PC for customer support needs after working hours, and will provide for the customers an instant reply all day long, whenever they ask for it.

Algorithm requirements - The three levels analysis

The algorithm will search the text (from the email body) for specified keywords, and will check the level of match:

- If there’s an exact-text (“Match”) to the DB repository command, MIA will have a direct decision, and will know how to answer the incoming mail and what process it should run.
- If there’s no exact-match to any command, MIA will look for possible user meanings. Then MIA will send the user those potential options for his/her decision.
- If no option is relevant for the user, MIA will ask guiding question to lead the user to what he/she wants to ask. The questions hierarchy is according to the decision-tree (decision-tree will be explained shortly) that configured.

The Decision-Tree: The Decision Tree is a tree which is stored in the DB. It categorizes the tasks (which MIA can do) into categories and sub-categories.

Example of a Decision Tree that can be used by MIA: In the example above we can see possible way to configure the decision-tree.
As we explained in the previous section (“The three levels analysis”), if MIA cannot analyze the email by finding the exact text on the DB, and also it cannot find relevant options by key-words, it will ask guiding questions according to the *decision-tree*.

<table>
<thead>
<tr>
<th>Examples for scenario:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User</strong>: Arkadiy</td>
</tr>
<tr>
<td>-&gt; unreadable text</td>
</tr>
<tr>
<td><strong>MIA</strong>: Did you mean:</td>
</tr>
<tr>
<td>2. Trigger a process.</td>
</tr>
<tr>
<td>3. My request is not listed above</td>
</tr>
</tbody>
</table>

**User**: Arkadiy

**MIA**: Did you mean:

1. CC related
2. Environment related
3. My request is not listed above

..... (and so forth according to Decision Tree)

As you can see the options that MIA sends to user is taken from above Decision-Tree

**Programming Languages and Development Platforms**: SQL is Must, Python and Oracle DB knowledge are advantages, nice to have NLP (Natural language processing) knowledge, Nice to have Oracle Apex knowledge
PROJECT 11

Company: Amdocs

Project Domain: Data Management, Dashboard, Telecom, User experience

Project: Communication data consumption personal dashboard

Supervisor: Gilli Shama

Project Description:

Project Overview: In this project you will define and develop a dashboard for a consumer of communication services, of his/her personal consumption.

Amdocs data management description: Amdocs is a global company, over 80 countries, with 20,000 employees, providing software services for the world communication service providers (Telecom). Amdocs exists for more than 30 years, starting in Israel. Amdocs is providing data management services to the telecom industry, including full maintenance of world largest data stores.

Detailed Project Description: You are probably one of the users of an Israeli communication provider. Think what, as a user you would like to see in a dashboard of your usage. Would you like to see trends of your bills, how the cost of your bill splits between different services you get. In this project you will define what a personal consumer of communication service would like to see as a web report of his communication services.

- You will design a plan from analysis to implementation.
- You will suggest creative new views
- You will interview consumers for their preferred information
- You will design a UI for your dashboard, on your selected platform, and collect feedback to this UI
- You will implement your design with Amdocs realistic data
- You will test your implementation on your pre-defined KPIs

Programming Languages and Development Platforms:

- Creativity, ability to interview people,
- UI and graphic design capabilities
- Knowledge in data structures, and SQL
- Knowledge in Tableau or other reporting tool is an advantage
- Knowledge in Hadoop is an advantage
**PROJECT 12**

**Company:** Amdocs

**Project Domain:** Data Management, Quality, Telecom

**Project:** Data validation rules definition

**Supervisor:** Gilli Shama

**Project Description:**

**Project Overview:** In this project you will define a set of rules that check for quality of data created by systems in the telecom world, you will interview Amdocs experts, define a set of logical rules, insert the rules into a DQM system, and run rules over realistic data.

**Amdocs data management description:** Amdocs is a global company, over 80 countries, with 20,000 employees, providing software services for the world communication service providers (Telecom). Amdocs exists for more than 30 years, starting in Israel. Amdocs is providing data management services to the telecom industry, including full maintenance of world largest data stores.

**Detailed Project Description:** Most data management projects in the world suffer from data quality issues, such as missing data, data not in correct format, contradicting data, and not aligned data. For example, the same City data value can be shown as “Tel-Aviv”, “Tel Aviv” and “Telaviv”, and then when you will ask for a histogram by city it will appear as 3 different cities. There are several tools that goes over the data and check for mismatches, and perform automatic data cleaning, per set of rules. There are also open shred rules for data cleaning.

Your project will include:

1. Creating data cleaning rules repository
   a. Collecting open-source rules for data cleaning
   b. Creating a set of Amdocs telecom specific data cleaning rules – interviewing Amdocs experts
2. Proof of concept development
   a. Populating a selected data cleaning tool with your rules
   b. Running the rules on a demi data set

**Programming Languages and Development Platforms:**

- Collaboration, and ability to interview in English (over the phone people from abroad)
- Knowledge in data structures, and SQL
- Knowledge in Hadoop is an advantage
PROJECT 13

Company: Amdocs

Project Domain: Data Science, Analytics, Tableau, Testing, Performance Management

Project: Test execution trend prediction

Supervisor: Gilli Shama

Project Description:

Work place: As part of this project students will need to come to Amdocs Haifa in order to work on the project.

Project Overview: In this project you will perform exploratory research of test execution data, build prediction models of upcoming trends in testing progress.

Amdocs Testing description: Amdocs is a global company, over 80 countries, with 20,000 employees, providing software services for the world communication service providers (Telecom). Amdocs exists for more than 30 years, starting in Israel. Amdocs Testing is the market leader of testing services outsourcing to the telecom industry. The leadership in revenue is driven from Amdocs leadership in testing innovations.

Detailed Project Description: Testing execution produces large ongoing work progress data. The execution progress data is mainly on defects (for example, defect status, opening time, detected tester, defect system and process) and test cases. Amdocs already collects defects and test cases full details into one aggregated data warehouse.

On this project you will enter the Amdocs DWH with Tableau. You will look for patterns in testing behavior, try to model trends for progress predictions. Once you will find formulas that model trends you will create a dashboard over Tableau that displays the prediction next to concurrent data progress.

Programming Languages and Development Platforms:

- Knowledge in data modeling
- Knowledge or experience in statistics and in Tableau is an advantage
**PROJECT 14**

**Company:** Amdocs  
**Project Domain:** Web, User interface (UI)  
**Project:** Self SMM tool  
**Supervisor:** Sol Yehezkel  

**Project Description:** The purpose of this project is to design and develop a web based questionnaire tool.

SMM (Service Maturity Model) is a Periodic centralized review of Account performance – comparing to a predefined guidelines. This audit is based (mainly) on answers received via questionnaire (currently in excel). The questionnaire contains ~400 questions for ~30 processes (a set of answers is predefined for each question). The grade for each item (question, process and overall grade) is 1-5. The account’s key member should choose the answer which represent the situation in his account. System main functions: auto grade calculation, parallel editing, reports execution, notification alerts and permission layer.  
The system should be also friendly, Self-guided and visible to account stakeholders and the auditors.  
The system should be able to support multiple accounts BUT a person should have the ability to view and edit limited parts in the questionnaire.  
As part of the project students will design the system according to the Business Requirements (BRs), the students will evaluate and choose the tool/framework for this project and will develop the system.  
Amdocs mentor will guide you in the development process; your success is our success.  

**Reports:**  
Reports should be available for the current status, % completion, PASS due date and so  
Also the ability to analyze and to create reports for several accounts or all accounts per region, process, line, year etc...  

**Programming Languages and Development Platforms:** Web portal design, UI knowledge, and plugin architecture.
PROJECT 15

Company: Amdocs

Project Domain: Social Media, algorithms, Graph Theory, machine learning, User Interface, DB, Open APIs

Project: Finding k-most influencing nodes in an interactions graph

Supervisor: Idan Shanny

Project Description: The goal is to find the k-most social-influencing users of a Telco company, based on the interactions between them. The output of the project can be used as valuable marketing data.

The project consists of two phases:

1. **Building the interaction graph:**
   - **Input:** raw interactions data (calls, messages) and communication interactions of messages application
   - **Output:** a graph, where the vertices are the users, and the edges weights are the ‘strengths’ of interactions between the 2 vertices.
   
   The graph will be built using the raw data extracted from the DB:
   - The number of incoming and outgoing interactions;
   - The interaction quality (duration, length);
   - The ‘quality’ of the interaction partners (their interactions with others – recursively, average monthly charges, cost and nature of owned wireless device, gender, etc.)

   Note: the students will verify how to interact (e.g. Open APIs) with Facebook messenger, matrix.org or WhatsApp, if possible, in order to monitor users interactions. Notice that we wish to verify only if it is possible to know only the very existence of the interaction between the users, and not the content of it.

2. **Analyzing the interaction graph:**
   - **Input:** the graph built in phase (1)
   - **Output:** the k-most social-influencing users

This task involves a research for finding a suitable algorithm and implementing it. The participants need to work independently, show research curiosity together with implementation ability and show innovative spirit
**Programming Languages and Development Platforms**: Programming (Java), algorithms in Graph Theory

---

**PROJECT 16**

**Company**: Outbrain Inc.

**Project Domain**: big data analytics and visualization

**Project**: **User controlled diversity of received content recommendations**

**Supervisor**: Ido Tamir

**Project Description**: Content recommendation engines provide users with recommendations on billions of page views per day, as those users consume content on media (a.k.a. “publisher”) sites, from CNN to YNET. The diversity of the recommendations displayed for users is known to have a profound impact on engagement metrics.

This project uses Outbrain’s external APIs\(^1\) to display a list of recommendations per user. Each returned recommendation is supplied along with descriptive metadata along several dimensions (e.g. the recommendation’s category). The goal of the project is to create an innovative user experience built around user-controlled diversity (namely, the user will control the amount of diversity of the recommendations by filtering less/more diverse recommendations based on their metadata). While this project is mainly focused around user experience, it will require connecting with Outbrain’s external APIs\(^1\). Therefore, some knowledge in javascript / client-side REST API consumption is needed, and will be acquired during the project.

- Programming languages and development platforms:
  - Javascript programming.
  - CSS web page styling.
  - Data visualization toolkit of the students’ choice. One such potential toolkit is [Google Charts](http://developers.sphere.com).

**Programming Languages and Development Platforms**: nothing specific.

---

\(^1\) [http://developers.sphere.com](http://developers.sphere.com)
**PROJECT 17**

**Company:** Outbrain Inc.

**Project Domain:** NLP, Chat bot programming

**Project:** Chat bot able to provide personal content recommendations

**Supervisor:** Ido Tamir

**Project Description:** Content recommendation engines provide users with recommendations on billions of page views per day, as those users consume content on media (a.k.a. “publisher”) sites, from CNN to YNET.

This project uses Outbrain’s external APIs\(^1\) to integrate with KIK messenger via a chatbot\(^2\). The bot is able to receive messages from the user and respond with content recommendations. A ‘conversation’ can be held between the user and the bot, leading to better-personalized recommendations for the user. For example, the bot might ask the user for his preferences regarding sites / categories he might be interested in, and present articles from those sites / categories.

This project will require developing a bot chat service, connecting to Outbrain’s external APIs (Sphere API\(^2\)) and integrating with a 3\(^{rd}\) party provider (KIK messaging\(^3\)). Java knowledge is required.

- Programming languages and development platforms:
  - Java programming
  - Basic understanding of NLP (Natural Language Processing)

**Programming Languages and Development Platforms:** nothing specific.

---

\(^1\) http://developers.sphere.com
\(^2\) https://engine.kik.com
PROJECT 18

Company: Intel

Project Domain: Dynamic memory allocation, heap management, OS runtime

Project: Modifying and optimizing dynamic memory allocation libraries and measuring the side effects

Supervisors: Ron Gabor and Tomer Stark

Project Description: Dynamic memory allocation libraries contain the dynamic heap management functions for languages such as C and C++. These functions include malloc, calloc, realloc and free. Multiple implementations for such libraries exist, including those in the OS runtime (e.g. glibc of Linux) as well as open source libraries (e.g. tcmalloc). These libraries are usually optimized for 16 bytes alignment of blocks. Recent technology progress may drive for changes in the alignment to 32 bytes or even 64 bytes, which would increase the memory overhead (potential fragmentation) and may impact performance for legacy workloads. The suggested project involves choosing one of the available libraries, modifying it to support 64 and 32 byte aligned blocks, optimizing the implementation, and measuring the memory overhead (due to fragmentation) and if possible also the performance and bandwidth overheads. The students are expected to be fluent in C and in dynamic heap management concepts, and be able to quickly study and understand code and algorithms, modify them and run them.

Programming Languages and Development Platforms: C, Linux (Windows is possible)

Courses:

- 234112 - Introduction to computer - C language
- 234114 - Introduction to Computer Science
- 234218 - Data Structures 1 - מבנים נתונים 1
- אלגוריתמים 1 - 234247
PROJECT 19

**Company:** Intel Corporation

**Project Domain:** Computer security

**Project:** Anomaly malware detection  
**Supervisors:** Sharon Martin and Avishai Redelman

**Project Description:** In this project, you will develop a malware detection system, based on supervised learning algorithm of system calls. During the supervised learning phase, malware samples will be collected and analyzed statically/dynamically for identifying the system calls used by each sample. The collected information will be used for training a classifier which distinguishes between malicious and legit programs.

**Programming Languages and Development Platforms:** Students decision

**Courses:** Course in machine learning (recommended not mandatory)

---

PROJECT 20

**Company:** Intel Corporation

**Project Domain:** Computer security

**Project:** Integrated code review environment  
**Supervisor:** Michael Atlas

**Project Description:** The LLVM Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, LLVM has little to do with traditional virtual machines, though it does provide helpful libraries that can be used to build them. One of its most usable advantages is its SSA capable internal representation. Translating x64 code to this representation allows writing code analyzers for finding well-known vulnerabilities.

In this project, you will develop a significant part of code review environment for taint analysis of the code under review. This project will be based on usage of LLVM framework and its internal code representation for data flow analysis.
**Programming Language and Development Platforms:**
C++, Linux, LLVM framework

**Courses:**
234118 - Computer Organization and Programming
236360 - Theory of Compilation

---

**PROJECT 21**

**Company:** Intel Corporation

**Project Domain:** Computer security

**Project:** Finding vulnerabilities through fuzzing in SGX Enclave

**Supervisors:** Sharon Martin and Avishai Redelman

**Project Description:** In this project, you will develop a generic tool for finding vulnerabilities in SGX enclave. SGX is the latest most advantage security protected technology from Intel, introduced in the last Intel processor launched in August 2015. You will use fuzzing methodologies for generating inputs to SGX enclave.

The input generation will be based on the EDL (Enclave Definition Language) file and known software attacking technics.

**Programming Language and Development Platforms:** Students decision
PROJECT 22

Company: Intel Corporation

Project Domain: Computer security

Project: Genetic fuzzing

Supervisor: Adi Kurtz

Project Description: Fuzz testing is a common method used by security researchers to identify application weaknesses. The very basic idea is to randomize many inputs to the application and check its behavior. Many tools and methodologies are available nowadays to smartly leverage this basic fuzzing approach.

The objective of this project is to build a framework for execution of genetic fuzzing methodology. The algorithm is based on creation the next generation of inputs based on the finest inputs from the previous generation. There are few challenges in this project (for some of them we can help in supplying ready code snippets), like how to detect the finest inputs, how to create next generation based on the previous one, and how to pack all the pieces into an efficient algorithm.

The project might be checked on commercial products, and there are lots of chances that it will find real life vulnerabilities.

Programming Language and Development Platforms:

Required prerequisites: Students decision

Courses: Algorithms 1, knowing of Genetic Algorithms is a plus
PROJECT 23

Company: Intel Corporation

Project Domain: Computer security

Project: Translator of x64 code to LLVM IR

Supervisor: Michael Atlas

Project Description: The LLVM Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, LLVM has little to do with traditional virtual machines, though it does provide helpful libraries that can be used to build them. One of its most usable advantages is its SSA capable internal representation. Translating x64 code to this representation allows writing code analyzers for finding well-known vulnerabilities.

In this project you will translate object files to LLVM IR for further analysis by other tools.

Programming Language and Development Platforms: C++, Linux, LLVM framework

Courses:
23411 - Computer Organization and Programming
236360 - Theory of Compilation

PROJECT 24

Company: Rafael

Project Domain: Code Parsing, DB and graphics

Project: Logic Visualizer

Supervisors: Ori Glick and Irena Kupershtok

Project Description: Here is how you can contribute to Israel's national security: We need a standalone Windows application to help our engineers visualize, analyze and document their software assurance tests. These tests provide the final seal of approval for some of the most advanced and confidential systems in Rafael.
The tests are written in a script language and stored in an SQL DB. The Logic Visualizer will make it come to life by turning syntax to flow charts and plain logical instructions into shiny UML graphics that make sense!

The principles are simple – extract the script from the DB, Parse and draw. But it's not that simple - it should work fast and flawless, look good, be robust and customizable. Our goal is to improve the tests' readability and make them accessible and understandable to all engineers involved in the project, whether they come from a software background or not.

**Programming Language and Development Platforms**: C#/C++, WPF (Visual Studio), SQL (MicroSoft SQL), The software should run on Windows (7 and up) 32bit.

**Courses**:
- 234112
- 234218
- 236363
- 236360 (An advantage)

---

**PROJECT 25**

**Company**: Philips Medical Systems

**Project Domain**: Software Systems, Medical Devices, Monitoring Automation, Big Data

**Project**: AUDIT – Advanced Usability Diagnostic Tool

**Supervisor**: Yan Tsitrin

**Project Description**:


The Philips Healthcare offers its customers various CT-scanners keeping them at the forefront of CT imaging (see [http://www.healthcare.philips.com/main/products/ct/products/scanners/](http://www.healthcare.philips.com/main/products/ct/products/scanners/) for details). A lot of SW is provided with these scanners to allow the eventual customers to use them in the most usable and safe way. Recently, several regulation standards (e.g. IEC 62366) have been proposed by the regulation authorities in order to guarantee that medical devices are indeed safe and the user errors are prevented whenever it is possible.
However, no formal metrics exist to evaluate and compare CT-Scanners of different manufacturer or several versions of the same CT-Scanner (say, before and after some improvement).

The objective of this project is to design an infrastructure layer of CT-Scanners Application providing its developers with ability to record events generated by the CT-operator in context of the Application usage. For this purpose, the student shall be provided with a relatively simple Platinum-DMS application and shall develop SW allowing record events generated by the Platinum-DMS operator. Data obtained this way shall be then analyzed, in order to reveal regarding any given UI component of the application:

- Whether this component is used by the operator in a reasonable and supposed manner;
- Does this component prevent the operator from reaching the desired functionality;
- For a given safety complaint, whether this UI component was involved; also statistics about <UI-component, Safety-complaint> could be provided.

When sufficient amount of data is collected this way, some Machine-Learning techniques (e.g. SVM) could be used in order to find common patterns in the UI related data (if time permits) http://www.cs.technion.ac.il/~cs234313/projects_sites/S12/18/site/objectives.html.

(see [Tube Lifetime Prediction Algorithm] for details)

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

Programming Language and Development Platforms:
- {C#}
- Advantage: Background in GUI development, Object-Oriented Design, Experience in the Embedded SW, WCF, WPF
Courses: 234122 - Introduction to Systems Programming

PROJECT 26

Company: Philips Medical Systems

Project Domain: Software Systems, Medical Devices

Project: CT- REST Remote Embedded Scripting Tool

Supervisor: Yan Tsitrin

Project Description: The interaction between Philips CT Scanner components and the HOST computer controlling the Scanner is done by message passing via the CAN communication bus (for more details see http://www.cs.technion.ac.il/~cs234313/projects_sites/W15/34/site/).

Detection Measurement System (DMS) is one of the major components of a CT-Scanner. It consists of detectors, which 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 DMS is controlled by embedded SW running on a processor residing on the DMS. Often it is necessary to update the embedded environment (e.g. to write to a DMS register) from the HOST computer. Such updates can be done one-by-one, when each write/read request is send via the CAN.

To improve this tedious and boring process we in Philips developed REST (Remote Embedded Scripting Tool), Figure 1.
**Figure 1:** REST: HOST – CT Scanner script base communication.

REST allows to create a script on the HOST side (1), inject the script (via the CAN) into the embedded processor (2). The script is a LUA (see http://www.lua.org/) based set of instructions executed by the embedded processor. The script is executed by the LUA interpreter (3) which is compiled and linked together with the embedded SW; and the execution results are sent to the HOST back (4). However, this tool is a command-line based and its user-experience should be improved.

**Figure 2:** REST: HOST – CT scanner script-based communication.

In this project the student shall be provided with the existing command-based REST tool and shall develop GUI for it. The GUI shall support editing, compiling and getting results which shall be presented in a convenient way. The development shall be done in .NET, C#, initially in the simulation environment, and then on the real Philips CT-Scanners of Brilliance and iCT families. It shall include all SW development stages: requirements definition, design, implementation, debugging, testing, verification.

**Programming Language and Development Platforms:**
- {C#, C++}
- **Advantage:** Background in Embedded Systems

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

>资本市场
PROJECT 27

**Company:** Philips Medical Systems

**Project Domain:** Software Systems, Medical Devices, Visualization

**Project:** **Platinum Plotter** – Platform Independent Utility for Monitoring Detection Measurement System Plotter

**Supervisor:** Yan Tsitrin

**Project Description:** Practical WPF Charts and Graphics, By Jack Xu

The Philips Healthcare offers its customers various CT-scanners keeping them at the forefront of CT imaging (see [http://www.healthcare.philips.com/main/products/ct/products/scanners/](http://www.healthcare.philips.com/main/products/ct/products/scanners/) for details). Detection Measurement System (DMS) is one of the major components of a CT-Scanner (Figure - [2]). It consists of detectors, which 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 DMS architectures (mechanical, electronic etc.) of various CT Scanners are different. In Philips different tools for monitoring of DMS of several kinds have been created. In order to cope with various tools in a common generic way we started to develop CT-DMS Platinum (Figure - [2]). Platinum DMS is intended for:

- Reading diagnostic data transferred from the DMS in various formats, via different channels;
- Managing and analyzing the collected data;
- Evaluating and visualization of the DMS components statuses
- For each component being monitored the Platinum provide its latest status and log of its values in a defined period of time within a DMS working session;
Plotter is one of the most important SW-component for the eventual users (Figure - [3]). However, at the moment only a primitive and partial implementation of the related functionality has been implemented.

The objective of this project is to define, design and implement the Platinum-Plotter performing the described above. The development shall be done in .NET, C#, using WPF technologies; initially in the simulation environment, and then on the real Philips CT-Scanners of Brilliance and iCT families.

**Programming Language and Development Platforms:**
- {C#};
- **Advantage:** Background in GUI development, Object-Oriented Design, Experience in the Embedded SW, WCF, WPF.

**Courses:** 234122 Introduction to Systems Programming

---

**PROJECT 28**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems, Medical Devices

**Project:** RCA Platinum DMS – Root Cause Analyzer for Platform Independent Utility for Monitoring Detection Measurement System

**Supervisor:** Yan Tshitin


Detection Measurement System (DMS) is one of the major components of a CT-Scanner (Figure - [2]). It consists of detectors, which 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 DMS architectures (mechanical, electronical etc.) of various CT Scanners are different. In Philips different tools for monitoring of DMS of several kinds have been created. In order to cope with various tools in a common generic way we started to develop CT-DMS Platinum (Figure - [2]).

Platinum DMS is intended for:
• Reading diagnostic data transferred from the DMS in various formats, via different channels;
• Managing and analyzing the collected data;
• Evaluating and visualization of the DMS components statuses
• For each component being monitored the Platinum provide its latest status and log of its values in a defined period of time within a DMS working session;

In groups of monitored items frequently appear some common behavioral patterns. *E.g. if a set of temperature sensors report similar error message, it is highly probable that the electronic board responsible for these sensors control is out of order.*

The objective of this project is to define an infrastructure for defining RCA (Root Cause Analysis) rules as independent knowledge-base (in XML based format) which would allow to formulate such rules with a convenient GUI, parse them using a generic mechanism and then use in problems analysis.

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

**Programming Language and Development Platforms:**
- {C#};
- **Advantage:** Background in GUI development, Object-Oriented Design, Experience in the Embedded SW, WCF, WPF.

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

םבםי雜 תחתנונצ מזרעסת
Company: Philips Medical Systems

Project Domain: Software Systems, GUI, Visualization

Project: 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). 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

Programming languages and development platforms:
- {C#};
- Advantage: Background in Statistics and SW Reliability, WPF

Courses: 234122 - Introduction to Systems Programming

---

**PROJECT 30**

**Company:** Philips Medical Systems

**Project Domain:** Software Systems, Medical Devices

**Project:** Web-CT- Platinum DMS – Web-Based CT Platform Independent Utility for Monitoring Detection Measurement System

**Supervisor:** Yan Tsitrin

**Project Description:** The Philips Healthcare offers its customers various CT-scanners keeping them at the forefront of CT imaging (see http://www.healthcare.philips.com/main/products/ct/products/scanners/ for details). Detection Measurement System (DMS) is one of the major components of a CT-Scanner.
consists of detectors, which 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 DMS architectures (mechanical, electronical etc.) of various CT Scanners are different. In Philips different tools for monitoring of DMS of several kinds have been created. In order to cope with various tools in a common generic way we started to develop CT-DMS Platinum (Figure - [1]). Platinum DMS is intended for: (a) Reading diagnostic data transferred from the DMS in various formats, via different channels; (b) Managing and analyzing the collected data; (c) Evaluating and visualization of the DMS components statuses; (d) For each component being monitored the Platinum provide its latest status and log of its values in a defined period of time within a DMS working session.

The objective of this project is to develop Web-Application with the same functionality – Web Platinum DMS.

The student shall be provided with the existing SW tool, Platinum-DMS (see figure below, [2]), and shall continue its development. The main project objectives are (1) Turn it to Client-Server Application (allowing to monitor Scanner’s Host remotely using MVC paradigm); (2) The visualization enhancement.

The development shall be done in .NET, C#, using WCF, MVC, ASP.NET 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: Background in GUI development, Object-Oriented Design, Experience in the Embedded SW, WCF, MVC, ASP

Courses: 234122 - Introduction to Systems Programming
PROJECT 31

**Company:** LogicBlox

**Project Domain:** Machine Learning, Artificial Intelligence, Combinatorial Optimization

**Project:** Statistical relational learning via factorization

**Supervisor:** Prof. Benny Kimelfeld

**Project Description:**

**Overview**
Statistical Relational Learning (SRL) is a branch of Machine Learning and Artificial Intelligence that focuses on models that combine relational structures, logic, and uncertainty. The challenge with SRL is the high computational complexity that SRL tasks (learning and inference) entail. The growing public appetite for Big Data analysis intensifies the need for scaling up SRL implementations. In this proposal, the student will implement a novel approximation approach that is based on tensor factorization. With this project, the student will deepen in, and gain experience with, the important concept of SRL, and may have a substantial impact on the research of facilitating practical SRL.

**Background**
Statistical Relational Learning (SRL) is a branch of Machine Learning and Artificial Intelligence that focuses on models that combine relational structures, logic, and uncertainty. SRL has attracted a lot of attention because of its generality, ease of development, and the applicability to multi-modal data. SRL has been the foundation for building knowledge bases (KB). Markov Logic Networks (MLN) and Probabilistic Soft Logic (PSL) are two state of the art approaches. For MLN based KBs, deepdive is a state of the art tool.

Despite its success in small scale data, there are some challenges in scaling it for real world problems. Generally, the techniques taken to computationally solve SRL tasks (training and inference) are of two approaches. In the first, the relational information is mapped down to a probabilistic graphical model. In the second approach, the relational information is modeled as multidimensional arrays, and efficient solutions are established through various kinds of factorizations. Factorization models give us an embedding of facts and relationships into a latent domain in the same way that Google’s word2vec deep net represents words in a k-dimensional latent space. Factorization models tend to scale better than graphical models. On the other hand, factorization models lack expressivity, since they do not provide any natural embedding for logical rules.
The goal of this project is to increase the expressivity and the complexity (model capacity) of the factorization models, by using more sophisticated methods, such as Factorization Machines along with objective functions borrowed by Probabilistic Soft Logic. One advantage of this approach is that it requires much simpler algorithms for training and inference than probabilistic graphical models, and it reduces the memory footprint. If the project is successful, then the result will provide a framework for easy-to-code solutions for SRL problems on real Big Data domains, and often on one’s personal laptop!

Probabilistic Soft Logic

For an introduction to PSL see video here. For applications of PSL, look at this video.
PSL uses the Lukasiewiewicz t-norm (Lnorm) to express the degree a rule is satisfied. For example, if we have predicates of the form \(R(x) \land L(y)\) and \(I(R(x))\) is the degree of satisfaction of \(R(x)\) in the range \([0, 1]\), then the degree of satisfaction of the rule is given by

\[
\max(I(R(x)) + I(L(y)) - 1, 0)
\]

If the predicate is \(R(x) \lor L(y)\), then the degree of its satisfaction is given by

\[
\min(I(R(x)) + I(L(y)), 1)
\]

The distance to satisfaction of a formula is given by

\[
\max(0, I(r_{body}) - I(r_{head}))
\]

Where \(r_{body}\) is the premise (or body) of the rule, and \(r_{head}\) is the conclusion (head) of the rule. A possible world is obtained by assigning a degree of satisfaction to each ground fact, and its probability is proportional to an exponent in the sum of satisfactions (where each rule has an associated weight coefficient).

Project Proposal
In this project the student will implement an SRL solution in a manner that is different from existing solutions, as we describe below. We conjecture that this implementation will be significantly more efficient than existing ones, with possibly more accurate results. The input will include rules and entity domains. In training, the goal will be to determine rule weights by fitting to examples. In inference, the goal will be to assign degrees of truth to facts by maximizing a goal function (that aggregates the degree of truth over all rule groundings).

The technique will be based on Logistic Tensor Factorization (LTF). In LTF, the relational data are modeled as a tensor, according to the following illustration.
\( x_{ijk} = 1 \) if the relationship \( \text{relation}_k(\text{entity}_i, \text{entity}_j) \) exists

The above method factors the relationship tensor \( R \) as \( X = AR \) \( A^T \). The loss functions are given below:

\[
\arg \min_{X,R} \text{loss}(X; A, R) + \lambda_A \|A\|^2 + \sum_{k} \lambda_{R_k} \|R_k\|^2 \\
\text{loss}(X; A, R) = \sum ||X_k - AR_kA^T||^2 \\
\text{loss}(X; A, R) = - \sum_{ijk} \log \sigma(\theta_{ijk}) + (1 - x_{ijk}) \log(1 - \sigma(\theta_{ijk}))
\]

\[
\sigma(\theta_{ijk}) = \frac{1}{1 + \exp(-a_i R_{kj}^T)}
\]

The problem of the above method is that it can only model binary predicates of the form \( R(x, y) \). We would like to be able to express first-order logical rules of the form:

\[
R_1(x, y, l) \land R_2(z, w) \rightarrow R_3(t)
\]

So, we need to face two challenges.
1. Extension to \( k \)-ary predicates with \( k > 2 \).
2. Support of logical rules.
For the first challenge, we propose to use Factorization Machines that can model higher order relationships. For the second problem, we propose to use the objective function that PSL is using.

Evaluation
The success of the project will be determined by benchmarks evaluated on standard SRL datasets, and will be compared against PSL.

PROJECT 32

Company: LogicBlox

Project Domain: Machine Learning

Project: Deep learning with factorization machines

Supervisor: Prof. Benny Kimelfeld

Project Description:

Introduction
Deep learning has been a highly successful approach in problems like image recognition, speech recognition and text mining. In all of these problems, data points are correlated either temporally or spatially. In the multibillion world of retail business, demand forecasting is a critical problem. A few percentage of accuracy improvement in prediction can save hundreds of millions of dollars in wastage. For many years, the retail community was limited by manually tuned linear-regression models that required careful feature engineering. Very recently, there has been a movement towards a new algorithmic approach - Factorization Machines (FM). The FM algorithm allows to automate a substantial portion of the tuning process. Both linear regression and FM have been preferred due of their natural interpretation. In this project, we intend to extend FM to accommodate a deep architecture while maintaining interpretability.

Background
The Factorization Machine (FM) algorithm has been successful in modeling multimodal information in recommender systems. It has also served very well as a regression solution that can model higher-order interaction terms between variables. For example, if the temperature and the product name are two features in a linear regression problem, then their combination (temperature,product) is another feature that specifies the effect of the combination. Such correlations are often extremely useful; for instance, in high temperatures coke demand goes up, while blanket demand goes down.

The second order FM is:
\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{l=0}^{d-1} \sum_{j=1}^{d-1} \langle y_{l,j}, x_l x_j \rangle \]

or

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{k=0}^{d-1} \sum_{l=0}^{k-1} \sum_{j=0}^{d-1} \langle y_{l,j}^{d-l}, x_l x_j \rangle \]

The third order FM would be:

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{k=0}^{d-1} \sum_{l=0}^{k-1} \sum_{j=0}^{d-1} \sum_{i=0}^{d-1} \langle y_{l,j}^{d-l}, x_l x_j x_i \rangle \]

where \( v \in \mathbb{R}^k \). Although FM has much more capacity than linear regression and it can model more complicated scenarios, it still has some limitations.

**Increasing Model capacity with nonlinearities**

It has been shown in the deep-learning community that introducing nonlinearities can significantly increase the model capacity. Nonlinearities also promote sparsity, which is important for generating interpretable data representations. This project will focus on demand forecasting where the dependent variable \( y \) is continuous. There, the Rectifier Linear Unit (ReLU) function is appropriate:

\[ r(x) = \max(0, x) \]

We propose the following modification to the original model:

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{l=0}^{d-1} \sum_{j=0}^{d-1} \langle r(y_{l,j}), x_l x_j \rangle \]

Now the FM model can be written as:

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{j=0}^{d-1} r(x_j) \]

This is a single-layer neural network. We can think of adding layers in two different ways. The first option is as follows.

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{l=0}^{d-1} \sum_{j=0}^{d-1} r(x_j) \]

The other option is to work on higher-order FM and introduce the nonlinearity on each order of interaction level:

\[ y = f(x) = w_0 + \sum_{l=0}^{d-1} w_l x_l + \sum_{l=0}^{d-1} \sum_{j=0}^{d-1} \langle r(y_{l,j}), x_l x_j \rangle \]

**Increasing Model Capacity with Recursive Data Partitioning**

When data is multimodal, a mixture model is typically more appropriate. In other words, increasing the capacity of your model by adding layers is not as effective as partitioning the data and training FMs or nonlinear FMs as described above. Random partitioning is the best you can
do in random forests (e.g., when you do not have any prior information on the data). FM and nonlinear FM both map every training point \((x_i, y_i)\) to a \(k\)-dimensional geometric representation. For example, in the nonlinear case we have:

\[
(x_i, y_i) \rightarrow ([r_0(x_i), \ldots, r_{d-1}(x_i)], y_i)
\]

Each dimension of the \(r\)-vector represents a mode. The sparsity or the variance of each dimension is an indication of complexity. Splitting the dataset on the dimension with the highest complexity gives two new datasets that can be retrained with a new FM. This process can be recursively applied to every partition of the data. The result is a binary tree where every node has an associated FM. It is very similar to a decision tree or a \(kd\)-tree. Tree based models have been highly successful in different domains. Moreover, such models often offer natural interpretations on the results.

**Evaluation**

All methods for increasing the model capacity of FM will be tested on a real production system for demand forecasting in big US retailers. Any improvement in the accuracy can measure the immediate impacts on savings for the retailers.

**PROJECT 33**

**Company:** HPE Software

**Project Domain:** Natural language processing

**Project:** Artemis

**Supervisors:** Ohad Assulin and Uri Kalish

**Project Description:** Software functional tests are programs that verify the behavior of a specific software/application. We would like to build a tool that let the test developer to ‘write’ it utilizing a natural language (e.g. enter ‘abc’ in the username field, enter ‘cde’ in the password field, click on the login button). We have done research that proves the feasibility of such a technology so we will guide the students step to by step. Starting with the language parser all the way to the reply engine. This tool will get open sourced.
Programming Languages and Development Platforms: JavaScript

PROJECT 34

Company: HPE Enterprise

Project Domain: Big Data, Predictive Analytics

Project: Predictive analytics for developers

Supervisor: Asaf Lahav

Project Description: Predictive analytics is one of the most promising types of analytics for businesses these days. One of the increasingly popular ways to tap into these possibilities, is through the Data Science method of Machine Learning. “Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data.” (Wikipedia, http://en.wikipedia.org/wiki/Machine_learning)

The challenge with Machine Learning is that it usually requires an experienced Data Scientist to engineer and optimize an effective predictive model.

Data Scientists, especially good ones, are rare. Becoming a good data scientist requires excellent knowledge, not only of relevant technologies and methods, but also deep understanding of Mathematics, Statistics and Graph Theory (a big plus).

A challenge lies in the fact that the expertise and understanding of the data usually requires a Business Analyst who understands how the data behaves, its quality, the business context and its’ meaning.

So basically, in order to run meaningful predictions you would need to gather both Business Analysts and Data scientists in order to research, investigate and develop the predictive models.

Then, in comes the developer... Once the analysis and modelling work is done, the solution still needs to be implemented in production systems in order to make it available to other business stakeholders and users. This usually requires a Developer to write some program code to finish the job. Herein lies the challenge faced by many organizations: Data Science is not only hard, but expert resources are usually even harder to come by.

There has to be an easier way.
As an alternative to hiring an army of Business Analysts and Data Scientists, organizations can now provide their Developers with access to the Prediction APIs from HPE Haven OnDemand. These
services hit right on the sweet spot of enabling coders with little to no Data Science expertise to develop new solutions that embrace Predictive Analytics simply through a few lines of code that make some API calls. That’s right, developers now have the tools to do the job of a Business Analyst and Data Scientist when building a solution that includes Prediction Modelling capabilities.

In this project, the student will work on the predictive analytics engine and will be invested in the effort of researching and automating the optimization of predictive models and algorithms.

**Programming Languages and Development Platforms:**
- Programming languages: Java, Scala, Python (and potentially R)
- Development Platforms: Apache Spark MLlib, Spring Framework, Tomcat
- Big Data Platforms: Spark, HADOOP, Vertica

**Courses:** Students who successfully finished one or more of the following courses are preferred: 236501 or 236756 or 236760 or 236779 or 236941 or 236757 or 46195 or 235602

---

**PROJECT 35**

**Company:** HPE Enterprise

**Project Domain:** Operating Systems, Mobile

**Project:** Android threads Analysis Viewer

**Supervisor:** Avi Kabizon

**Project Description:** Android applications need to be very efficient and run very fast. But on the other hand, Android developers tend to open a lot of threads of execution, sometimes unknowingly. There are almost no tools that can help a developer understand where the bottlenecks in a multi-threaded application in Android are.

The project purpose is to develop a new visualization tool that will help a developer to understand what is going on in the application threads.

**Programming Languages and Development Platforms:** There are 2 parts to this project:
* First, there is a cross platform tool for visualizing the threads. It will be written in Java.
* Second, the will be some android applications that will be used as samples. This will require the students to learn some Android programming, in Java, using Android Studio.
PROJECT 36

Company: HPE Enterprise

Project Domain: Big Data processing

Project: Portero

Supervisor: Ori Troya

Project Description: We research a method to create a Device Risk Profile that is created from finger print and the ongoing usage of a user to strongly identify the risk of a device and to give the customer the ability to decide. Furthermore, to bring this to the next level the risk profile will be share among all customers and can even be consumed by vendors around the world.

The Device Finger Print will be a combination of device sensors system parameters, user list of installed applications, installed applications, contact book and more that will give you the ability to de-anonymize a device and first impression of the device usage.

Pattern of Recent Use will aggregate the ongoing usage of a user in the device, the applications the user is using, different session, typing, gestures done, breaks between, operation that demand user interactions e.g login into different applications, taking images, other unique identifiers related to tracking methods, etc., will increase the probability of user action related to human. Therefore will improve the impression of the device authenticity.

The Pattern of recent use will be calculated by the combination of the above parameters with the ability for operation that demand human interaction to increase the probability and automatically effect other pending decisions related to the device. Meaning as much different applications will be installed on the device with significant traffic the probability for a human user action will increase.

Programming Languages and Development Platforms: Java, Linux, Android

Courses:
- 234247
- 236357
- 234319
- 236369
PROJECT 37

Company: HPE Enterprise

Project Domain: Mobile applications

Project: Windows phone agent for AppPulse mobile

Supervisor: Amichai Nitsan

Project Description: AppPulse Mobile is HPE’s product for monitoring mobile applications. The tools tracks the activity of mobile applications running on real users’ devices, enabling the application owner to understand what is the experience users are having while using the application.

AppPulse Mobile already supports Android and iOS, so we now need to develop a Windows Phone library. The students will build a library that will be embedded inside Windows Phone applications, enabling the report of events related to user experience. The students will also build a Windows Phone demo app to demonstrate the capabilities.

Programming Languages and Development Platforms: The project will be based on the Windows Phone development environment – building mobile apps in C# using Visual Studio.

PROJECT 38

Company: HPE Enterprise

Project Domain: Data Analytics

Project: Gaia

Supervisor: Shay Tsadok

Project Description: Distributed data analytics micro service architecture, focused on software development pipeline data. Real Continuous Delivery. Cloud based development: GitHub, AWS, Slack, CircleCi, Trello and more.
One of the goals of this project is to build a cloud based solution on Amazon using open source technology. We want to learn and be open to new ways of development and have a business goal of making our management system the best option for Amazon development.

**Programming Languages and Development Platforms:**
- Platforms: Docker, CoreOS, AWS, RabbitMQ, ElasticSearch and others.
- Languages: Python, Node.js, Java 8, Go

---

**PROJECT 39**

**Company:** HPE Enterprise

**Project Domain:** Big Data Analytic

**Project:** Social networks Data Collector

**Supervisor:** Avi Saidian

**Project Description:** Monitoring and analyzing social media information is becoming an essential task for any big company. In this project we will create a twitter connector to collect all published tweets on a certain account.

We will implement the following:
1. Connector – subscribe to twitter stream and collect tweets.
2. Configuration webapp - enable generic setting accounts to collect on.
3. Online Reporting – classify tweets and display summary in graphical dashboard using D3.

**Main focus:**
Concurrent Collection from several twitter accounts using non-blocking APIs.
Saving collected data to database (MongoDB or Cassandra).
Analyze tweets using online services as:
https://dev.havenondemand.com/apis --> Text Analysis

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

**Courses:** 236700 - Software design
PROJECT 40

Company: Elbit Systems

Project Domain: Real time embedded systems

Project: Bring Up Real time Lua + TLSF SW development environment for real time embedded mission critical avionic software applications

Supervisor: Dotan Halevi

Project Description: SW development for Real Time Embedded Avionic system is long and costly, due to Mission and Safety critical considerations (e.g. no dynamic memory allocations, strict static C code analysis etc.).

On the other hand, Dynamic languages (e.g. Java, Python) development is much (x10 or more) faster, but does require memory allocation, garbage collection management etc.

This project’s aim is to combine the two worlds of development, merging the requirement and advantages of both into one coherent solution.

The goal is to build an embedded application on a target (e.g. Arduino, or a VxWorks target) using C as a base layer, Lua as the application development language, and TLSF as the O(1) memory allocation layer.

The products are:

- A working Embedded application (E.g. a Digital Lock – קודן)
- A working host pc simulation, using exactly the same application code
- An embedded environment: loader, Source level debugger + Lua Debug terminal console
- Simple profiler for time and memory usage.

In short, Lua was chosen as the Dynamic language since Lua (+ TLSF – O(1) Malloc and free) fits naturally in this spot:

- Bare Lua engine code is about 15K lines of ANSI C code + 1K TLSF code
- Lua is Fast (x1.5 of C++ applications)
- Lua has many advanced Language features:
  - Functions as First class objects
  - Iterators and closures
o Modules and Packages
o Native bit manipulation operations
o Object Oriented
o Co Routines
o C Integration (C extending Lua and vice Versa)
o Platform independent code
o Controllable Garbage collection
o Data serialization
o ANSI C code base
o Changing Dynamic allocator to pull allocation easily
o IDE, local and Remote debugging

Some links: About the Lua language:
About: http://www.lua.org/about.html
History (why it is a small, fast and portable language): http://www.lua.org/history.html
An article about Java in Mission Critical software (stating the problems of dynamic languages in safety critical SW development):
https://books.google.co.il/books?id=0iWNu0nK070C&pg=PA211&lpg=PA211&dq=using+dynamic+languages+in+mission+critical+software&source=bl&ots=1Xjbuo8OgP&sig=TclLVNjwM5RH7oVfrjmKJFAhwk&hl=iw&sa=X&ved=0ahUKEwiB7-vTho3KAhXHzxQKHd5wDA04FBDoAQgZMAA#v=onepage&q=using%20dynamic%20languages%20in%20mission%20critical%20software&f=false

Programming Languages and Development Platforms:
- C
- Lua
- Windows PC (and/or Linux development station)
- Embedded target with Discrete inputs (keys) and outputs (LEDs)
  o E.g. Arduino, or preferably VxWorks target

Courses: מבנה, מערכות המחשב
PROJECT 41

Company: Elbit Systems

Project Domain: UI designer\Web & DB application, Optimization

Project: Verification dashboard

Supervisor: Alexander Vinnik

Project Description: The project purpose is development of a WEB Dashboard UI that connects to associated DB to have an ability to:
1. view all relevant information for specified system requirements
2. update verification status based on user decisions
3. display verification process status and results
4. export data to specified formats

Background details: During verification process system engineers required to combine data from different DB (QC, DOORs, Excels) for decision making process. This process can be optimized by using APIs for data mining from different DB and combining them to one place. Application shall display all required data to the user, receive user inputs and update DB accordingly. In addition application shall update dashboard that include progress status from different sources.

Programming Languages and Development Platforms:
- Must: WEB Development, DB applications, HTML 5
- Nice to have knowledge: Java, familiarity with Microsoft Office, HP Quality center and IBM DOORs APIs

PROJECT 42

Company: Elbit System

Project Domain: Programming

Project: Stack-Analysis Tool

Supervisors: Ori Arad and Carmel Ravid
Project Description:

If you always wanted to understand in-depth the relationship between the code you write and the compiled code, this is the project for you...

The project's goal is very simple: build a tool for stack analysis to ensure that the program does not exceed the memory size allocated to it (in our case – for C/C++ covered in the first semesters, but the principles you will learn throughout the project can be applied to any other language).

During the project, you will analyze the program and build its call graph. After that, you will extract the graph to find the "heaviest" path (i.e., the sequence of calls that uses the most memory on the stack) – and present it through a GUI.

This sounds simple (and it is also), but the project will beautifully show you how everything you learn in the lectures – connects to the real world (introduction to computer science, computer architecture, algorithms, and more...). (A project you should not miss, it will increase your understanding of programming – beneficial for interviews, by the way.)

Programming Languages and Development Platforms: Python or C# or Java (or any other – will be decided with the students)

Courses:
- Required: Intro to CS (234114), MATAM (234122), ATAM (234118)
- nice to have: Algorithms (234247)

PROJECT 43

Company: Elbit System
Project Domain: Compilers
Project: Extend C/C++ compiler – static code analysis
Supervisors: Ori Arad and Carmel Ravid

Project Description:

The goal of the project is to extend the capabilities of an existing compiler, adding additional warnings. In the industry, there are several standards for writing C/C++ in the Real-Time/Embedded domain. These standards are complex, often involving thousands of rules, but they make the code more readable, "correct" more, and reduce the risk of bugs. This project will expose you to these standards and their implementation, giving you a glimpse behind the scenes – for example, how the compiler handles pointers and more.
Programming Languages and Development Platforms: C Language (environment – in according with the chosen compiler – Probably GNU/GCC, Clang or similar)

Courses:
- **Nice to have:** Theory of Compilation (236360), Projects in Compilation
- **Required:** Students must have the ability to understand given open-source compiler, and be able to manipulate its code and add new features to it.

**PROJECT 44**

**Company:** Elbit System

**Project Domain:** Software Development Process, Programming, IDE

**Project:** Eclipse plugin for development-process integration

**Supervisors:** Ori Arad and Carmel Ravid

**Project Description:**

Eclipse plugin for Development-Process integration

In this project, students will extend an existing compiler with specific MISRA/JSF rules. This will allow the developer to enforce these rules at compile-time and identify any violations during the compilation process.

It is essential for students to choose a compiler and understand its architecture, as well as being able to modify its code and add new features to it. Experience in writing/extend existing compiler (or a high motivation and ability to develop this for the first time in this project) is highly recommended.

There is no doubt that students who complete this project will gain valuable experience – both in understanding the compiler architecture and in developing and maintaining existing code – which will be an advantage in future industry work.

**Company:** Elbit System

**Project Domain:** Software Development Process, Programming, IDE

**Project:** Eclipse plugin for development-process integration

**Supervisors:** Ori Arad and Carmel Ravid

**Project Description:**

Eclipse plugin for Development-Process integration

In this project, students will extend an existing compiler with specific MISRA/JSF rules. This will allow the developer to enforce these rules at compile-time and identify any violations during the compilation process.

It is essential for students to choose a compiler and understand its architecture, as well as being able to modify its code and add new features to it. Experience in writing/extend existing compiler (or a high motivation and ability to develop this for the first time in this project) is highly recommended.

There is no doubt that students who complete this project will gain valuable experience – both in understanding the compiler architecture and in developing and maintaining existing code – which will be an advantage in future industry work.

Writing code is only one step in the software development process. In various courses, you have learned about the different stages of software development: requirements specification, design, implementation, testing, etc. In an ideal world, we can achieve traceability (traceability) between all stages of the process, and see how the system requirements "flow" to the software design, detailed requirements, code, and at the same time be connected to the various tests.

The goal of the project is to enable the software engineer to see and visualize where the code he writes is linked: for example, from which system requirement or software requirement does the function he implements come. And which test (if it already exists) checks his code and so on.

In practice, the students will implement a plugin for eclipse (or a similar) that identifies which function/method the user is implementing (or selecting) – and using a test against the given model, presents the user with the relevant items. The plugin can be developed using, for example, eclipse plugin development tools, and should be designed to integrate with the existing Eclipse framework.
The project involves interns working on models for the development of software taught in various courses (Math, Software Engineering Methods, and others) and demonstrating their abilities in the real world, utilizing the knowledge gained during their studies.

**Programming Languages and Development Platforms:** Java, eclipse environment

**Courses:**
- Nice to have (but not required): Previous project included eclipse plugin development, Software Engineering Methods (236321)
- Required: Student must have the ability to develop plugins for eclipse environment—either by previous experience or by self-learning during the semester.

---

**PROJECT 45**

**Company:** Elbit Systems

**Project Domain:** HTML5, AngularJS, Mobile App

**Project:** Advanced hybrid mobile app framework

**Supervisors:** Alon Bialik and Guy Peled

**Project Description:** Client application built from client side layer running on the end-user’s device (Browser, Tablet, Mobile), communicating with other machines, and using a local DB that underly their data model.

In order to do so, we developed its own data model, which is designed to be simple, small in size, and UI oriented in terms of data visualization and data binding.

In order to perform the data conversion to/from data model of various backend systems, a JS based binary parser will be used. The parser need to be adapted per each Target type duo, and its inner mapping logic reflecting Target’s needs.

DB and Dynamic UI abilities are defined in another PC SW.

The goals of the project are to create a hybrid mobile application that enables end user to send and receive data to/form different target computers. Visualize the data using predefined UI Bindings.
Recommended background:
- HTML5 & JS programming knowledge
- Familiarity with the Angular JS is an advantage.
- Ability to work independently, following a given HLR (High Level Requirement), Open-minded, Innovative.

**Programming Languages and Development Platforms:** HTML5, Angular JS, Ionic

**Courses:**
- 234218 - Data Structures 1
- 236703 - Object Oriented Programming

---

**PROJECT 46**

**Company:** Elbit Systems, Aerospace division

**Project Domain:** Networking

**Project:** Real time video streaming

**Supervisors:** Eran Solomon and Ophir Levin

**Project Description:** The goal of this project is to propose, design and implement standard streaming of H.264 packets in real time over IP networks.
An important part of the pilot mission is the debriefing. For that our helmet records all helmet related mission data.

Instead of storing it on the helmet we can stream it to a video recorder on the A/C that stores all flight data.

Our intention is to stream video using the standard RTP Payload Format for H.264 Video protocol to a PC (that will simulate the video recorder) and there it will be captured by a streaming video recorder e.g. VLC.
Programming Languages and Development Platforms: ANSI C over platform of student’s choice

Courses:
- 234123 - Operating Systems
- 236334 - Introduction to Computer Networks
- 236341 - Internet Networking (advantage)

PROJECT 47

Company: IBM
Project Domain: Hierarchical clustering, Spark
Project: Hierarchical Clustering Using Spark
Supervisor: Boris Rozenberg

Project Description: Clustering is often an essential first step in data mining intended to define data categories. Hierarchical clustering [1], a widely used clustering technique, can offer a rich data representation by suggesting the potential group structures. However, many applications face the demand to perform this operation over large amount of data. A common approach to tackle this problem is distributed computation. However, parallelization of such an algorithm is challenging as it exhibits inherent data dependency during the hierarchical tree construction. We designed a parallel implementation of Hierarchical Clustering algorithm. What is requested in this project is to implement our algorithm using Spark Framework [2] and to evaluate its performance on datasets provided by IBM.

References:
Programming Languages and Development Platforms: Spark, Java

PROJECT 48

Company: IBM, Haifa Research Lab


Project: Cloud Application cyber security – closing a security hole

Supervisor: David Hadas

Project Description:
Background: Platform-as-a-Service (PaaS) systems such as the Open-Source Cloud Foundry project may deploy numerous concurrent web applications/services. Each web application uses a set of libraries and cloud services of specific versions. Over time vulnerabilities are discovered across the used libraries. Once a vulnerability is published, a race starts between offenders seeking to exploit the vulnerability and defenders seeking to close the security hole.

The Challenge: This is an Open-Source project. It is an opportunity to learn how a PaaS Cloud Platform – Cloud Foundry operates internally and to consider security aspects of the deployed applications. The project will build a scalable mechanism for interrogating the Cloud Platform, extracting data about running applications. The information will be matched against a database of known vulnerabilities. Once vulnerability is detected, measures will be taken to remediate the situation.
A potential contribution to Open-Source (Cloud Foundry).
There is a possibility for a continuation project in the next semester

**Programming Languages and Development Platforms:**
5th semester students with a good control over a relevant programming language (JS, Python, Ruby or JAVA). Background in cyber security is an advantage.

**Programming Languages:** JS, Python, Ruby or JAVA
See Cloud Foundry, an open-source Platform as a Service (PaaS) system ([www.cloudfoundry.org](http://www.cloudfoundry.org)).

---

**PROJECT 49**

**Company:** IBM Research - Haifa

**Project Domain:** Computer Vision, Machine Learning

**Project:** *Classification of medical images using semi-supervised learning*

**Supervisor:** Rami Ben-Ari, Guy Amit

**Project Description:** Computerized analysis of medical images often utilizes machine learning techniques, in order to derive complex relationships between image features and their semantic meanings. Supervised learning approaches require costly and laborious effort of image annotation by medical experts, which is often a bottleneck in medical imaging research. In this project, we will explore a semi-supervised approach for classification of medical images. This approach fuses a small amount of labeled data with a larger set of unlabeled data to yield a reliable classifier. The application domain of this project will be automatic lesion classification in X-ray mammograms and magnetic resonance images of the breast. The project will combine advanced computer vision techniques for content-based retrieval of unannotated images, with state-of-the-art machine learning algorithms. The students will have a chance to gain experience in the field of medical imaging analytics.

**Programming languages and development platforms:**
Matlab (image processing and statistics toolbox)

- Experience in Matlab programming
- Background in machine learning
Background in computer vision / image processing

PROJECT 50

Company: IBM, Haifa Research Lab

Project Domain: Web Applications, Cloud Systems, Platform-as-a-Service, Networking, Security

Project: Cloud binding for open-source web-application server frameworks

Supervisor: David Hadas

Project Description:
Background: Applications and services deployed in the cloud commonly use open-source Web Framework libraries as a front-end to handle the communication with clients across the web. The Web Framework library handles all common web tasks such as HTTP parsing, validation and routing as well as other common tasks required by web applications/services such as handling of cookies, authentication etc. Using the library, the web server developer offloads common tasks and concentrate on developing the application/service logic.

The Challenge: This is an Open-Source project. It is a unique opportunity to dive into the design of Web Framework libraries, learning about their use as part of current cloud applications and services and learning about the future of cloud platforms. The project will extend the Web Framework to include an interface for management by the Cloud Platform (see figure). The interface will be designed as a REST API and will enable Cloud Platforms and other management entities to exchange information with the Web Framework. Such binding will allow the Cloud Platform to extend the services it offers to Devops deploying their applications in the cloud.
It is preferred that code will be contributed back to the respective Open-Source communities.

It is possible for two teams to work on different Web Frameworks.

There is a possibility for a continuation project in the next semester.

**Programming Languages and Development Platform:**
- Passion for Open Source + good control over a relevant programming language.
- Programming languages and development platforms: The team must have good programming skills in one or two of the following languages: JS, Python, Ruby or JAVA.

**Web Frameworks examples:**

For additional context and background on the Cloud Platform, see Cloud Foundry, an Open-Source Platform-as-a-Service system ([www.cloudfoundry.org](http://www.cloudfoundry.org)).

---

**PROJECT 51**

**Company:** IBM

**Project Domain:** Logic, Artificial Intelligence

**Project:** **Software for encoding satisfiability problems into conjunctive normal form**

**Supervisors:** Yael Ben-Haim and Alexander Ivrii

**Project Description:** Boolean Satisfiability (SAT) is one of the most well-known NP-complete problems. As such, probably no algorithm can solve every SAT problem in polynomial time. Nevertheless, there has been an enormous advance in the development of highly efficient SAT solvers that routinely solve industrial problems with millions of variables and clauses. SAT solvers find many applications, including hardware and software model checking, automatic test pattern generation, planning, cryptography, and more.

SAT solvers operate on formulas in conjunctive normal form (CNF), and thus an essential part of the process (and a very active field of research) consists of translating the original satisfiability problem into a CNF instance.

The goal of the project is to write a software package for translating cardinality (and possibly pseudo-Boolean) constraints into CNF. There are many translation methods already available, so the students are expected to read papers, understand the translations, and to implement them.

We are especially looking for research-oriented students, who are also interested in improving
existing translations and collaborating on the development of new translations. From the practical side, the package must have a good design and a clean API to be used as a standalone component for various IBM applications.

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

---

**PROJECT 52**

**Company:** IBM

**Project Domain:** 3-tier application development, Databases, Web-based technology

**Project:** Translyzer - A 3-tier analyzer & visualizer for address-translation data

**Supervisor:** Shai Doron

**Project Description:**

**Background:** The IBM Power-Processor is at the core of IBM's high-end servers used by large corporations for a variety of applications. One present challenging application is Cloud, where modern servers required to have built-in hardware support for address-translation that is vital for their virtualization capabilities.

The Power-Processor is also used in some of the top world's supercomputers.

The processor is extremely complex due to the strong requirements for high performance and low power. Moreover, it has to be built bug-free as it fulfills premium services and at times critical mission devices.

As a result, IBM (like other leading CPU manufacturers) invests significant R&D resources in new technologies and practices (including AI techniques and advanced algorithms) for the verification of the Power-Processor.

The purpose of this project is to assist in this domain.
**Project definition:** Developing a 3-tier application (server, database, and client) for analysis and visualization of address-translation data from post-silicon verification of IBM's next generation processor.

**Project Scope**
A 3-tier application consisting:

1. **Server**
   1.1. Implementing JDBC API (Java Database Connectivity) to ingest address-translation data produced by post-silicon stimuli generator into database

2. **Database**
   2.1. DB2 Express
   2.2. Exercising database content with SQuirreL SQL Client (or other DB client of choice)

3. **Client**
   3.1. Developing Java client or web-based client (for example can be implemented via Node.js)
   3.2. Implementing relevant analysis metrics on the data and suggest possible statistical characteristics

**Programming Languages and Development Platform:** JDBC (Java Database Connectivity), DB2 Express, Web-based Java and/or JavaScript (like node.js)

**Courses:**
- 236703 - Object Oriented Programming
- 236363 - Database Management Systems

---

**PROJECT 53**

**Company:** IBM

**Project Domain:** Cloud, PAAS, Docker

**Project:** Video surveillance service using FFmpeg and Docker on IBM Bluemix cloud

**Supervisors:** Ophir Azulai and Yevgeni Burshtein

**Project Description:** Bluemix is an implementation of IBM’s Open Cloud Architecture based on Cloud Foundry, an open source Platform as a Service (PaaS). Bluemix delivers enterprise-level services that can easily integrate with your cloud applications without you needing to know how
to install or configure them. Bluemix provides access to a wide variety of services that can be incorporated into an application.

Docker containers wrap up a piece of software in a complete filesystem that contains everything it needs to run: code, runtime, system tools, and system libraries – anything you can install on a server. This guarantees that it will always run the same, regardless of the environment it is running in.

FFmpeg is an open-source software project that produces libraries and programs for handling multimedia data.

In this project we will develop a basic video surveillance service on BlueMix using Docker and FFmpeg. In the client side we will have a gateway which will get the video stream from the camera, divide it to short segments and upload it to the cloud. The server side will receive the video and detect motion in it using an existing algorithm that IBM will supply. When a motion is detected, an event will be sent to a test web application.

The motion detection algorithm will run in a Docker container and video segments will be sent to it using the Kafka message queue. Video decoding will be done using FFmpeg.

**Programming Languages and Development Platform:** Java, Node.js, Linux

**Courses:** Software Engineering Methods (234321) or Object Oriented Programming (236703)

---

**PROJECT 54**

**Company:** Check Point Software Technologies Ltd.

**Project Domain:** Computer Security

**Project:** Malware recognition with binary fingerprinting

**Supervisors:** Ben Herzog and Amir Mizrachi

**Project Description:** Every day, tens of thousands of new malware specimens are released into the wild. Facing this huge influx, malware researchers must answer difficult questions: Have we seen this before? Where? Are some of the incoming samples novel and worth further investigation? Is the next Conficker/Zeus hiding somewhere in this hay stack?
The sheer volume of new malware is too much for humans to handle. In this project, we will build an automated classifier to assist humans with this mission. The classifier should receive a new executable as input, and answer whether it is brand new - or looks similar to something the classifier has encountered before. In the latter case, it should report the similarities.

The scope of the project is limited to the use of static analysis methods (i.e. no execution of the file), with a heavy focus on artifact extraction and binary difference methods. Ultimately, the goal is for the resulting classifier to be precise and efficient. A part of the project will involve figuring out which design choices would best balance those two requirements.

**Programming Languages and Development Platforms:** Flexible; we recommend that the student(s) opt for a language with a well-curated standard/3rd party library that implements graph-theoretic algorithms, such as Python or Java.

**Courses:**

**Mandatory:**
- Introduction to Systems Programming (234122)
- Computer Organization and Programming (234118)
- Data Structures 1 (234218)
- Algorithms 1 (234247)

**Desirable:**
- Computer Security (236350)

---

**PROJECT 55**

**Company:** Say2eat

**Project Domain:** NLP (Natural Language Processing) & Machine Learning, REST API, Node.js

**Project:** Semi-supervised learning NLP engine for food delivery and groceries ordering for IoT system

**Supervisor:** Ilya Kritsmer

**Project Description:** The raise of Internet of Things era demands new ways of human-machine interaction to be developed. Smart homes, connected cars, wearable devices - all of them has no GUI thus need to understand what people say to them. To address the challenge, the NLP (Natural Language Processing) technology has started to be adopted widely in the industry.
There are still many problems to solve when the NLP, which was a pure theoretical field in computer science until recently, is taken to the real world. One of the most interesting – is to how make the NLP engine to understand a millions of ways people might say the same thing – for example, order their favorite pizza.

And here where the Machine Learning algorithms come to rescue. So called semi-supervised learning process, e.g. ability to learn from a large volume of untagged data based on (much) smaller set of tagged data, which takes it very close to the process of human learning, might do the work. The project participants will examine the latest Machine Learning engines recently open sourced by Google, Microsoft and Facebook and choose the most appropriate one to work with in order to accomplish the goal of the project.

Main Components
- Merchants DB – cloud engine connected to more than real 10,000 US based merchants databases via REST API.
- User Input DB – a set of English phrases which is used by customers to order food/groceries, which constantly grows.
- Verification Engine – the components constantly feeds the NLP Engine with the aforementioned DBs and separates the user input for recognized and unrecognized phrases.
- Learning Engine – learns the unrecognized phrases and configures the NLP Engine properly.

NLP Engine - the core engine which makes the hard work.

Programming Languages and Development Platforms: C#, Node.js and related frameworks, Azure Cloud.

Courses: Theory of Algorithms and/or Learning and Artificial Intelligence group of courses