informatics / *mathematics*

Inria

RESEARCH CENTER
Lille - Nord Europe

FIELD

Activity Report 2017

Section Software

Edition: 2018-02-19

BONSAI Project-Team

6. New Software and Platforms

6.1. BCALM 2

KEYWORDS: Bioinformatics - NGS - Genomics - Metagenomics - De Bruijn graphs

SCIENTIFIC DESCRIPTION: BCALM 2 is a bioinformatics tool for constructing the compacted de Bruijn graph from sequencing data. It is a parallel algorithm that distributes the input based on a minimizer hashing technique, allowing for good balance of memory usage throughout its execution. It is able to compact very large datasets, such as spruce or pine genome raw reads in less than 2 days and 40 GB of memory on a single machine.

FUNCTIONAL DESCRIPTION: BCALM 2 is an open-source tool for dealing with DNA sequencing data. It constructs a compacted representation of the de Bruijn graph. Such a graph is useful for many types of analyses, i.e. de novo assembly, de novo variant detection, transcriptomics, etc. The software is written in C++ and makes extensive use of the GATB library.

- Participants: Antoine Limasset, Paul Medvedev and Rayan Chikhi
- Contact: Rayan Chikhi
- Publication: Compacting de Bruijn graphs from sequencing data quickly and in low memory
- URL: https://github.com/GATB/bcalm

6.2. NORINE

Nonribosomal peptides resource

KEYWORDS: Drug development - Knowledge database - Chemistry - Graph algorithmics - Genomics - Biology - Biotechnology - Bioinformatics - Computational biology

SCIENTIFIC DESCRIPTION: Since its creation in 2006, Norine remains the unique knowledgebase dedicated to non-ribosomal peptides (NRPs). These secondary metabolites, produced by bacteria and fungi, harbor diverse interesting biological activities (such as antibiotic, antitumor, siderophore or surfactant) directly related to the diversity of their structures. The Norine team goal is to collect the NRPs and provide tools to analyze them efficiently. We have developed a user-friendly interface and dedicated tools to provide a complete bioinformatics platform. The knowledgebase gathers abundant and valuable annotations on more than 1100 NRPs. To increase the quantity of described NRPs and improve the quality of associated annotations, we are now opening Norine to crowdsourcing. We believe that contributors from the scientific community are the best experts to annotate the NRPs they work on. We have developed MyNorine to facilitate the submission of new NRPs or modifications of stored ones.

FUNCTIONAL DESCRIPTION: Norine is a public computational resource with a web interface and REST access to a knowledge-base of nonribosomal peptides. It also contains dedicated tools : 2D graph viewer and editor, comparison of NRPs, MyNorine, a tool allowing anybody to easily submit new nonribosomal peptides, Smiles2monomers (s2m), a tool that deciphers the monomeric structure of polymers from their chemical structure.

- Participants: Areski Flissi, Juraj Michalik, Laurent Noé, Maude Pupin, Stéphane Janot, Valerie Leclère and Yoann Dufresne
- Partners: CNRS - Université Lille 1 - Institut Charles Viollette
- Contact: Maude Pupin
- Publications: Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing - Smiles2Monomers: a link between chemical and biological structures for polymers - Norine: a powerful resource for novel nonribosomal peptide discovery - NORINE: a database of nonribosomal peptides. - Bioinformatics Tools for the Discovery of New Nonribosomal Peptides
- URL: http://bioinfo.lille.inria.fr/NRP

6.3. Vidjil

High-Throughput Analysis of V(D)J Immune Repertoire

KEYWORDS: Cancer - Indexation - NGS - Bioinformatics - Drug development

SCIENTIFIC DESCRIPTION: Vidjil is made of three components: an algorithm, a visualization browser and a server that allow an analysis of lymphocyte populations containing V(D)J recombinations.

Vidjil high-throughput algorithm extracts V(D)J junctions and gathers them into clones. This analysis is based on a spaced seed heuristics and is fast and scalable, as, in the first phase, no alignment is performed with database germline sequences. Each sequence is put in a cluster depending on its V(D)J junction. Then a representative sequence of each cluster is computed in time linear in the size of the cluster. Finally, we perform a full alignment using dynamic programming of that representative sequence against the germline sequences.

Vidjil also contains a dynamic browser (with D3JS) for visualization and analysis of clones and their tracking along the time in a MRD setup or in an immunological study.

FUNCTIONAL DESCRIPTION: Vidjil is an open-source platform for the analysis of high-throughput sequencing data from lymphocytes. V(D)J recombinations in lymphocytes are essential for immunological diversity. They are also useful markers of pathologies, and in leukemia, are used to quantify the minimal residual disease during patient follow-up. High-throughput sequencing (NGS/HTS) now enables the deep sequencing of a lymphoid population with dedicated Rep-Seq methods and software.

- Participants: Florian Thonier, Marc Duez, Mathieu Giraud, Mikaël Salson, Ryan Herbert and Tatiana Rocher
- Partners: CNRS - Inria - Université de Lille - CHRU Lille
- Contact: Mathieu Giraud
- Publications: High-Throughput Immunogenetics for Clinical and Research Applications in Immunohematology: Potential and Challenges. - High-throughput sequencing in acute lymphoblastic leukemia: Follow-up of minimal residual disease and emergence of new clones - Diagnostic et suivi des leucémies aiguës lymphoblastiques (LAL) par séquençage haut-débit (HTS) - Multiclonal Diagnosis and MRD Follow-up in ALL with HTS Coupled with a Bioinformatic Analysis - A dataset of sequences with manually curated V(D)J designations - Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing - Multi-loci diagnosis of acute lymphoblastic leukaemia with high-throughput sequencing and bioinformatics analysis - Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing - The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL.
- URL: http://www.vidjil.org

6.4. MATAM

Mapping-Assisted Targeted-Assembly for Metagenomics

KEYWORDS: Metagenomics - Genome assembling - Graph algorithmics

SCIENTIFIC DESCRIPTION: MATAM relies on the construction of a read overlap graph. Overlaps are computed using SortMeRNA. The overlap graph is simplified into relevant components related to specific and conserved regions. Components are assembled into contigs using SGA and contigs are finally assembled into scaffolds. The process yields nearly full length marker sequences with a very low error rate compared to the state of the art approaches. Taxonomic assignation of the obtained scaffolds is performed using the RDP classifier and is represented using Krona.

FUNCTIONAL DESCRIPTION: MATAM provides targeted genes assembly from the short metagenomic reads issued from environmental samples sequencing. Its default application focuses on the gold standard for species identification, 16S / 18S ribosomal RNA SSU genes. The produced gene scaffolds are highly accurate and suitable for precise taxonomic assignation. The software also provides a RDP classification for the reconstructed scaffolds as well as an estimation of the relative population sizes.

- Participants: Hélène Touzet, Pierre Pericard, Yoann Dufresne, Samuel Blanquart and Loïc Couderc
- Contact: Hélène Touzet
- Publication: MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes
- URL: https://github.com/bonsai-team/matam

DEFROST Project-Team

6. New Software and Platforms

6.1. Simulation de neurochirurgie

Vascular neurosurgery simulation based on SOFA Framework
KEYWORDS: Simulation - Health - Computer-assisted surgery

- Participants: Christian Duriez, Eulalie Coevoet, Jérémie Dequidt and Laurent Thines
- Partners: Université de Lille - CHRU Lille
- Contact: Christian Duriez

6.2. SOFA

Simulation Open Framework Architecture
KEYWORDS: Real time - Multi-physics simulation - Medical applications
FUNCTIONAL DESCRIPTION: SOFA is an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation. It is mostly intended for the research community to help develop new algorithms, but can also be used as an efficient prototyping tool. Based on an advanced software architecture, it allows : the creation of complex and evolving simulations by combining new algorithms with algorithms already included in SOFA, the modification of most parameters of the simulation (deformable behavior, surface representation, solver, constraints, collision algorithm, etc.) by simply editing an XML file, the building of complex models from simpler ones using a scene-graph description, the efficient simulation of the dynamics of interacting objects using abstract equation solvers, the reuse and easy comparison of a variety of available methods.

- Participants: Christian Duriez, François Faure, Hervé Delingette and Stéphane Cotin
- Partner: IGG
- Contact: Stéphane Cotin
- URL: http://www.sofa-framework.org

6.3. SoftRobots

SoftRobots plugin for Sofa
KEYWORDS: Numerical simulations - Problem inverse - Soft robotics
FUNCTIONAL DESCRIPTION: This framework is based on a mechanical modeling of the robot elements combined with fast real-time direct/inverse FEM solvers. The keypoint of our approach is that the same modeling is used for interactive simulation of its behavior and interactive control of the fabricated robots.

- Participants: Christian Duriez, Olivier Goury, Jérémie Dequidt, Damien Marchal, Eulalie Coevoet, Erwan Douaille and Félix Vanneste
- Contact: Christian Duriez
- URL: https://project.inria.fr/softrobot/

<div align="center">

DOLPHIN Team

</div>

6. New Software and Platforms

6.1. Grid'5000

Grid'5000 experimental platform

FUNCTIONAL DESCRIPTION: The Grid'5000 experimental platform is a scientific instrument to support computer science research related to distributed systems, including parallel processing, high performance computing, cloud computing, operating systems, peer-to-peer systems and networks. It is distributed on 10 sites in France and Luxembourg, including Lyon. Grid'5000 is a unique platform as it offers to researchers many and varied hardware resources and a complete software stack to conduct complex experiments, ensure reproducibility and ease understanding of results.

- Participants: Christian Pérez, David Loup, Frédéric Desprez, Laurent Lefèvre, Laurent Pouilloux, Marc Pinhède and Simon Delamare
- Contact: Frédéric Desprez
- URL: https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home

6.2. ParadisEO

KEYWORD: Parallelisation

SCIENTIFIC DESCRIPTION: ParadisEO (PARallel and DIStributed Evolving Objects) is a C++ white-box object-oriented framework dedicated to the flexible design of metaheuristics. Based on EO, a template-based ANSI-C++ compliant evolutionary computation library, it is composed of four modules: * Paradiseo-EO provides tools for the development of population-based metaheuristic (Genetic algorithm, Genetic programming, Particle Swarm Optimization (PSO)...) * Paradiseo-MO provides tools for the development of single solution-based metaheuristics (Hill-Climbing, Tabu Search, Simulated annealing, Iterative Local Search (ILS), Incremental evaluation, partial neighborhood...) * Paradiseo-MOEO provides tools for the design of Multi-objective metaheuristics (MO fitness assignment shemes, MO diversity assignment shemes, Elitism, Performance metrics, Easy-to-use standard evolutionary algorithms...) * Paradiseo-PEO provides tools for the design of parallel and distributed metaheuristics (Parallel evaluation, Parallel evaluation function, Island model) Furthermore, ParadisEO also introduces tools for the design of distributed, hybrid and cooperative models: * High level hybrid metaheuristics: coevolutionary and relay model * Low level hybrid metaheuristics: coevolutionary and relay model

FUNCTIONAL DESCRIPTION: Paradiseo is a software framework for metaheuristics (optimisation algorithms aimed at solving difficult optimisation problems). It facilitates the use, development and comparison of classic, multi-objective, parallel or hybrid metaheuristics.

- Partners: CNRS - Université Lille 1
- Contact: El-Ghazali Talbi
- URL: http://paradiseo.gforge.inria.fr/

6.3. VRPsolve

KEYWORDS: C++ - Mobile Computing, Transportation - Optimization

SCIENTIFIC DESCRIPTION: VRPsolve is a software for solving vehicle routing problems dealing with last-mile delivery issues that arise as we approach the final customer. When modeling and solving combinatorial optimization problems, especially problems related to the transport of goods and people, the resulting models are generally subject to a specific development in order to be validated, as industrial needs are highly dependent of the application domain. However, a set of conventional objectives and constraints, such as vehicles capacities, incompatible parcels, time windows, are now commonly encountered. In addition to being efficient and effective, VRPsolve differentiates from other tools by allowing to quickly and conveniently integrate ad-hoc constraints and objectives into a generic software. Indeed, VRPsolve effectively deal with industrial last-mile delivery vehicle routing problems and is able to cope with multiple objectives and a large number of constraints by using advanced optimization algorithms which are usually not available with existing softwares. In addition, VRPsolve allows industrial collaborations to be addressed by solving real-world problems requiring geographic information systems (GIS).

- Participants: Arnaud Liefooghe, Clive Ferret-Canape and Sébastien Vérel
- Contact: Clive Ferret-Canape
- URL: http://gforge.inria.fr/projects/vrpsolve

6.4. Platform Grid'5000

The Dolphin project-team has been the Principal Investigator of the Grid5000@Lille project funded (budget: 750K€) within the framework of the CPER. This project consists in building in 2017 a new site of the Grid'5000 platform at Lille. This new site hosted by Inria Lille replaces the old one which was located in the supercomputing center (CRI) of the University of Lille. It consists in a GPU-enhanced computing cluster composed of over 1.000 CPU cores and 60.000 GPU cores corresponding to a 20FLOPS computational power. The Grid5000@Lille project allowed also the recruitment of two engineers for the system and network administration and the software development for two years. Another upgrade with more GPUs is planned for the beginning of 2018.

- Participants: F. Desprez, F. Huet, E. Jeannot, Y. Jegou, A. Lebre, L. Lefevre, F. Loui, D. Margery, N. Melab, J-M. Menaud, P. Neyron, L. Nussbaum, C. Perez, J-M. Pierson, O. Richard., S. Varette.
- Contact: Nouredine Melab
- URL: https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home

<div align="center">

FUN Project-Team

</div>

6. New Software and Platforms

6.1. AspireRFID ALE

FUNCTIONAL DESCRIPTION: AspireRFID middleware is a modular OW2 open source RFID middleware. It is compliant with EPC Global standards. This new module integrates the modifications of the new standard release, including new RP and LLRP definitions and fixing bugs. This module has been implemented in the framework of the MIAOU project.

- Participants: Ibrahim Amadou, Julien Vandaële, Nathalie Mitton and Rim Driss
- Contact: Nathalie Mitton

6.2. ETINODE-CONTIKI-PORT

FUNCTIONAL DESCRIPTION: Contiki is an open source embedded OS for Internet of Things (IoT). It is light and portable to different hardware architectures. It embeds communication stacks for IoT Il embarque aussi des piles de communication pour l'internet des objets. This driver allows the running of Contiki OS over Etinode-MSP430. The code dalso allows the use of radio chip and embedded sensors. This module has been implemented in the framework of the ETIPOPS project.

- Participants: Nathalie Mitton, Roudy Dagher and Salvatore Guzzo Bonifacio
- Contact: Salvatore Guzzo Bonifacio

6.3. ETINODE-DRIVERS

FUNCTIONAL DESCRIPTION: These drivers for Etinode-MSP430 control the different embedded sensors and hardware components available on an Etinode-MSP430 node such as gyroscope, accelerometer and barometric sensor. This module has been implemented in the framework of the ETIPOPS project.

- Participants: Nathalie Mitton, Roudy Dagher and Salvatore Guzzo Bonifacio
- Contact: Salvatore Guzzo Bonifacio

6.4. EVe-TCF

Embedded Verifier for Transitive Control Flow
KEYWORDS: Security - Embedded - Embedded systems - JavaCard - Control Flow - Code analysis
FUNCTIONAL DESCRIPTION: Verification of transitive control flow policies on JavaCard 2.x bytecode. Control flow policies expressed using a DSL language are embedded in JavaCard packages (CAP files) using EVe-TCF convert tool. Control flow policies are then statically verified on-device at loading-time thanks to an embedded verifier (designed for smart cards in EVe-TCF). EVe-TCF (Embedded Verifier for Transitive Control Flow) also contains an off-device (i.e. PC tool) to simulate on-device loading process of JavaCard 2.x platforms with GlobalPlatform 2.x installed.

- Participants: Arnaud Fontaine and Isablle Simplot Ryl
- Contact: Nathalie Mitton

6.5. GOLIATH

Generic Optimized LIghtweight communication stack for Ambient TecHnologies
KEYWORDS: WSN - WSN430

FUNCTIONAL DESCRIPTION: GOLIATH (Generic Optimized LIghtweight communication stack for Ambient TecHnologies) is a full protocol stack for wireless sensor networks. This module has been implemented in the framework of the ETIPOPS project.

- Participants: David Simplot Ryl, Fadila Khadar, Nathalie Mitton and Salvatore Guzzo Bonifacio
- Contact: Nathalie Mitton
- URL: https://gforge.inria.fr/projects/goliath/

6.6. IoT-LAB robots

KEYWORDS: Internet of things - Robotics
FUNCTIONAL DESCRIPTION: IoT-LAB robots is an embedded robot controler on a Turtlebot2 providing the IoT-LAB node mobility functionnality

- Partner: Université de Strasbourg
- Contact: Julien Vandaële
- URL: https://github.com/iot-lab/

6.7. T-SCAN

KEYWORDS: Rfid - RFID Middleware
FUNCTIONAL DESCRIPTION: T-Scan is an interface ensuring the translation from a SGTIN tag format to an ONS hostname format according to the EPCGlobal standards. It allows the sending of a DNS request to look up the EPC-IS aides to which the product belongs in order to access the data relative to that product. This module has been implemented in the framework of the TRACAVERRE project.

- Participants: Gabriele Sabatino and Nathalie Mitton
- Contact: Gabriele Sabatino

6.8. FIT IoT-Lab

Participants: Nathalie Mitton [correspondant], Julien Vandaele, Matthieu Berthome.

FIT IoT-LAB is a very large scale open testbed that features over 2700 wireless sensor nodes and more than 200 robots spread across six different sites in France. Nodes are either fixed or mobile and can be allocated in various topologies throughout all sites. A variety of wireless sensors are available, with different processor architectures (MSP430, STM32 and Cortex-A8) and different wireless chips (802.15.4 PHY at 800 MHz or 2.4 GHz). In addition, "open nodes" can receive custom wireless sensors for inclusion in IoT-LAB testbed. This platform is completely open and can be used by any one wishing to run experiment on wireless sensors and robots.

The Lille site displays 3 subsets of the platforms:

- Euratechnologies: this site features 256 WSN430 sensor nodes operating in the 2.4GHz band. 64 nodes are mobile, embedded on mobile trains.
- Haute Borne: this site features 256 M3 sensor nodes operating in the 2.4GHz band and 64 mobile robots (32 turtlebots and 32 wifibots) completely remotely programmable.
- Opennodes: this site will feature (opening beginning 2015) 64 hardware open slots to allow any one to plug his own hardware and benefits from the platform debugging and monitoring tools.

INOCS Team

6. New Software and Platforms

6.1. dapcstp

A dual-ascent-based branch-and-bound framework for the prize-collecting Steiner tree and related problems
KEYWORDS: Mathematical Optimization - Systems Biology
FUNCTIONAL DESCRIPTION: Variants of the Steiner tree problem appear in a broad range of diverse applications, ranging from infrastructure network design to the analysis of biological networks and pattern recognition. In this software, we provide a branch-and-bound (B&B) framework for solving the asymmetric prize-collecting Steiner tree problem (APCSTP). Several well-known network design problems can be transformed to the APCSTP, including the Steiner tree problem (STP), prize-collecting Steiner tree problem (PCSTP), maximum-weight connected subgraph problem (MWCS) and the node-weighted Steiner tree problem (NW-STP).

- Contact: Markus Sinnl
- URL: https://github.com/mluipersbeck/dapcstp

6.2. HappyChic-ApproPick

KEYWORDS: Operational research - Optimization - Java
FUNCTIONAL DESCRIPTION: This software is a prototype developed for the bilateral contract with the company HappyChic. This software is a solver for an integrated warehouse order picking problem with manual picking operations. More precisely, the following problems are solved: (1) the assignment of references to storage positions, based on the iterative solving of minimum cost flow problems, (2) the division of clients orders into several parcels, respecting weight and size constraints, using a dynamic programming algorithm based on the split algorithm, (3) the batching of parcels into trolleys to perform picking tours, using a dynamic programming algorithm based on the split algorithm. The objective function is to minimize the total walking distance. This software is designed to deal with the large-sized industrial instances of HappyChic (considering hundreds of clients, thousands of positions and product references) in a short computation time (few minutes).

- Contact: Maxime Ogier

6.3. rcmwcs

A Relax-and-Cut Algorithm for Maximum Weight Connected Subgraph Problems
KEYWORDS: Mathematical Optimization - Systems Biology
FUNCTIONAL DESCRIPTION: Finding maximum weight connected subgraphs within networks is a fundamental combinatorial optimization problem both from a theoretical and a practical standpoint. One of the most prominent applications of this problem appears in Systems Biology and it corresponds to the detection of active subnetworks within gene interaction networks. The software is a framework to solve the model by means of Relax-and-Cut, i.e., Lagrangian relaxation combined with constraint generation.

- Contact: Markus Sinnl
- URL: https://msinnl.github.io/pages/rcmwcs.html

6.4. MIBLPsolver

A Solver for Mixed-Integer Bilevel Linear Problems
KEYWORD: Mathematical Optimization

FUNCTIONAL DESCRIPTION: Bilevel optimization problems are very challenging optimization models arising in many important practical contexts, including pricing mechanisms in the energy sector, airline and telecommunication industry, transportation networks, optimal expansion of gas networks, critical infrastructure defense, and machine learning. In this software, we present a new general purpose branch-and-cut framework for the exact solution of mixed-integer bilevel linear programs (MIBLP), which constitute a very significant subfamily of bilevel optimization problems.

- Contact: Markus Sinnl
- URL: https://msinnl.github.io/pages/bilevel.html

6.5. PARROT

Planning Adapter Performing ReRouting and Optimization of Timing

KEYWORDS: Decision aid - Railway - Scheduling

FUNCTIONAL DESCRIPTION: This is a decision support system addressing the problem of the rescheduling railway schedules on the Belgian network when maintenance operations are planned in the short term (2-3 weeks in advance). The deliverable is a software tool that will take as input: (1) the schedules initially planned for the different trains, (2) the initial routes of the trains, (3) maintenance operations / changes of elements in the form of constraints (unavailable routes etc.). He then provides in output: (1) the new train schedule, (2) the new routing of the fleet. The modifications must respect the constraints corresponding to the operations of maintenance. For example, in some cases it is common to leave at least a few minutes interval between two trains using the same track in the station. This constraint must then be propagated if a maintenance operation delays the arrival of a train. New schedules and routings have to be created following a specific goal. Changes made to schedules and routings must minimize: (1) variations on the time spent at the station, (2) the number of partially canceled trains (additional correspondence (s) or stations that are no longer served), (2) the number of fully canceled trains (no stations served).

- Contact: Martine Labbe

LINKS Project-Team

6. New Software and Platforms

6.1. ShEx validator

Validation of Shape Expression schemas
KEYWORDS: Data management - RDF
FUNCTIONAL DESCRIPTION: Shape Expression schemas is a formalism for defining constraints on RDF graphs. This software allows to check whether a graph satisfies a Shape Expressions schema.

- Contact: Iovka Boneva
- URL: https://gforge.inria.fr/projects/shex-impl/

6.2. gMark

gMark: schema-driven graph and query generation
KEYWORDS: Semantic Web - Data base
FUNCTIONAL DESCRIPTION: gMark allow the generation of graph databases and an associated set of query from a schema of the graph.gMark is based on the following principles: - great flexibility in the schema definition - ability to generate big size graphs - ability to generate recursive queries - ability to generate queries with a desired selectivity

- Contact: Aurélien Lemay
- URL: https://github.com/graphMark/gmark

6.3. SmartHal

KEYWORD: Bibliography
FUNCTIONAL DESCRIPTION: SmartHal is a better tool for querying the HAL bibliography database, while is based on Haltool queries. The idea is that a Haltool query returns an XML document that can be queried further. In order to do so, SmartHal provides a new query language. Its queries are conjunctions of Haltool queries (for a list of laboratories or authors) with expressive Boolean queries by which answers of Haltool queries can be refined. These Boolean refinement queries are automatically translated to XQuery and executed by Saxon. A java application for extraction from the command line is available. On top of this, we have build a tool for producing the citation lists for the evaluation report of the LIFL, which can be easily adapter to other Labs.

- Contact: Joachim Niehren
- URL: http://smarthal.lille.inria.fr/

6.4. QuiXPath

KEYWORDS: XML - NoSQL - Data stream
SCIENTIFIC DESCRIPTION: The QuiXPath tools supports a very large fragment of XPath 3.0. The QuiXPath library provides a compiler from QuiXPath to FXP, which is a library for querying XML streams with a fragment of temporal logic.
FUNCTIONAL DESCRIPTION: QuiXPath is a streaming implementation of XPath 3.0. It can query large XML files without loading the entire file in main memory, while selecting nodes as early as possible.

- Contact: Joachim Niehren
- URL: https://project.inria.fr/quix-tool-suite/

6.5. X-FUN

KEYWORDS: Programming language - Compilers - Functional programming - Transformation - XML

FUNCTIONAL DESCRIPTION: X-FUN is a core language for implementing various XML, standards in a uniform manner. X-Fun is a higher-order functional programming language for transforming data trees based on node selection queries.

- Participants: Joachim Niehren and Pavel Labath
- Contact: Joachim Niehren

<div align="center">

MAGNET Project-Team

</div>

6. New Software and Platforms

6.1. CoRTex

Python library for noun phrase COreference Resolution in natural language TEXts

KEYWORD: Natural language processing

FUNCTIONAL DESCRIPTION: CoRTex is a LGPL-licensed Python library for Noun Phrase coreference resolution in natural language texts. This library contains implementations of various state-of-the-art coreference resolution algorithms, including those developed in our research. In addition, it provides a set of APIs and utilities for text pre-processing, reading the main annotation formats (ACE, CoNLL and MUC), and performing evaluation based on the main evaluation metrics (MUC, B-CUBED, and CEAF). As such, CoRTex provides benchmarks for researchers working on coreference resolution, but it is also of interest for developers who want to integrate a coreference resolution within a larger platform.

- Participant: Pascal Denis
- Contact: Pascal Denis
- URL: https://gforge.inria.fr/projects/cortex/

6.2. Mangoes

MAgnet liNGuistic wOrd vEctorS

KEYWORDS: Word embeddings - NLP

FUNCTIONAL DESCRIPTION: Process textual data and compute vocabularies and co-occurrence matrices. Input data should be raw text or annotated text. Compute word embeddings with different state-of-the art unsupervised methods. Propose statistical and intrinsic evaluation methods, as well as some visualization tools.

- Contact: Nathalie Vauquier
- URL: https://gitlab.inria.fr/magnet/mangoes

MEPHYSTO Project-Team (section vide)

MINT2 Team (section vide)

Mjolnir Team

6. New Software and Platforms

6.1. InspectorWidget

An opensource suite to track and analyze users behaviors in their applications

KEYWORD: Instrumentation

FUNCTIONAL DESCRIPTION: InspectorWidget is a set of opensource tools to track and analyze users' behaviors in interactive software. It works with closed applications that do not provide source code nor scripting capabilities, covers the whole pipeline of software analysis and does not require programming skills. To achieve this, InspectorWidget combines low-level event logging (e.g. mouse and keyboard events) and high-level screen features (e.g. interface widgets) captured though computer vision techniques, or through accessibility hooks when exposed by applications.

NEWS OF THE YEAR: InspectorWidget now supports the collection and annotation of User Interface accessibility features.

- Participants: Christian Frisson, Sylvain Malacria, Stéphane Huot and Gilles Bailly
- Contact: Sylvain Malacria
- Publication: InspectorWidget: a System to Analyze Users Behaviors in Their Applications
- URL: https://github.com/InspectorWidget/InspectorWidget

6.2. WhichFingers

WhichFingers: Identifying Fingers on Touch Surfaces and Keyboards using Vibration Sensors

KEYWORDS: Interaction - HCI

SCIENTIFIC DESCRIPTION: HCI researchers lack low-latency and robust systems to support the design and development of interaction techniques using finger identification. We developed a low-cost prototype using piezo-based vibration sensors attached to each finger. By combining the events from an input device with the information from the vibration sensors we demonstrate how to achieve low-latency and robust finger identification. Our prototype was evaluated in a controlled experiment, using two keyboards and a touchpad, showing single-touch recognition rates of 98.2% for the keyboard and 99.7% for the touchpad, and 94.7% for two simultaneous touches. These results were confirmed in an additional laboratory-style experiment with ecologically valid tasks. Last we present new interaction techniques made possible using this technology.

FUNCTIONAL DESCRIPTION: WhichFingers consists in a hardware and a software components.

The hardware component consists of five Minisense 100 vibration sensors attached to each finger. The sensors use flexible PVDF piezoelectric polymer film loaded by a mass to offer high sensitivity to detect contact vibrations. They produce a voltage as large as 90V depending on the intensity of the shock or vibration. The five sensors are plugged into a micro-controller and sends the raw values to the host computer at 1000 Hz.

The software component monitors low-level interaction touch and key events, and declares the vibration sensor that created the highest voltage as the finger that produced the input event.

- Participants: Géry Casiez and Sylvain Malacria
- Contact: Géry Casiez
- Publication: WhichFingers: Identifying Fingers on Touch Surfaces and Keyboards using Vibration Sensors

6.3. Lagmeters

Systems to measure end-to-end latency in interactive systems

KEYWORDS: Interaction - Latency

FUNCTIONAL DESCRIPTION: The first method works with most optical mice and allows accurate and real time latency measures up to 5 times per second. In addition, the technique allows easy insertion of probes at different places in the system – i.e. mouse events listeners – to investigate the sources of latency.

The second method relies on a vibration sensor attached to a finger and a photo-diode to detect the screen response. Both are connected to a micro-controller connected to a host computer using a low-latency USB communication protocol in order to combine software and hardware probes to help determine where the latency comes from. We provide source code and materials to replicate both the hardware and software.

- Participants: Géry Casiez, Nicolas Roussel, Stéphane Huot, Thomas Pietrzak, Sébastien Poulmane, Stéphane Conversy, Damien Marchal and Matthieu Falce
- Partners: Université Lille 1 - Inria
- Contact: Géry Casiez
- Publications: Characterizing Latency in Touch and Button-Equipped Interactive Systems - Looking through the Eye of the Mouse: A Simple Method for Measuring End-to-end Latency using an Optical Mouse
- URL: http://ns.inria.fr/mjolnir/lagmeter/

6.4. libParamTuner

Cross-platform library to ease the interactive tuning of parameters at run time and without the need to recompile code.

KEYWORD: Interaction

FUNCTIONAL DESCRIPTION: libParamTuner provides a lightweight syntax to bind some variables of an application to the parameters defined in an XML file. Each modification of the XML file updates in real time the associated parameters in the application. A graphical interface allows editing the XML file, using interactive controls dynamically created for each parameter.

- Participants: Géry Casiez, Marc Baloup and Veis Oudjail
- Partners: Université Lille 1 - Inria
- Contact: Géry Casiez
- Publication: libParamTuner : interactive tuning of parameters without code recompilation
- URL: https://github.com/casiez/libparamtuner

6.5. liblag

Library implementing latency compensation techniques for interactive systems

KEYWORDS: Interaction - Latency

FUNCTIONAL DESCRIPTION: The library comprises the management of a set of multitouch input devices, the implementation of latency compensation techniques from the state-of-the-art and new latency compensation techniques developed in the project, and a system to handle artificial latency.

The library is developed in C++ using the Qt framework to allow compiling the same code on a wide range of devices and platforms.

- Contact: Géry Casiez
- Publication: Dispositif à affichage prédictif
- URL: http://mjolnir.lille.inria.fr/turbotouch/

MODAL Project-Team

6. New Software and Platforms

6.1. MixtComp

Mixture Computation
KEYWORDS: Clustering - Statistics - Missing data
FUNCTIONAL DESCRIPTION: MixtComp (Mixture Computation) is a model-based clustering package for mixed data originating from the Modal team (Inria Lille). It has been engineered around the idea of easy and quick integration of all new univariate models, under the conditional independence assumption. New models will eventually be available from researches, carried out by the Modal team or by other teams. Currently, central architecture of MixtComp is built and functionality has been field-tested through industry partnerships. Three basic models (Gaussian, multinomial, Poisson) are implemented, as well as two advanced models (Ordinal and Rank). MixtComp has the ability to natively manage missing data (completely or by interval). MixtComp is used as an R package, but its internals are coded in C++ using state of the art libraries for faster computation.

- Participants: Christophe Biernacki, Étienne Goffinet, Matthieu Marbac-Lourdelle, Quentin Grimonprez, Serge Iovleff and Vincent Kubicki
- Contact: Christophe Biernacki
- URL: https://modal-research.lille.inria.fr/BigStat

6.2. BlockCluster

Block Clustering
KEYWORDS: Statistic analysis - Clustering package
SCIENTIFIC DESCRIPTION: Simultaneous clustering of rows and columns, usually designated by biclustering, co-clustering or block clustering, is an important technique in two way data analysis. It consists of estimating a mixture model which takes into account the block clustering problem on both the individual and variables sets. The blockcluster package provides a bridge between the C++ core library and the R statistical computing environment. This package allows to co-cluster binary, contingency, continuous and categorical data-sets. It also provides utility functions to visualize the results. This package may be useful for various applications in fields of Data mining, Information retrieval, Biology, computer vision and many more.
FUNCTIONAL DESCRIPTION: BlockCluster is an R package for co-clustering of binary, contingency and continuous data based on mixture models.

- Participants: Christophe Biernacki, Gilles Celeux, Parmeet Bhatia, Serge Iovleff, Vincent Brault and Vincent Kubicki
- Partner: Université de Technologie de Compiègne
- Contact: Serge Iovleff
- URL: http://cran.r-project.org/web/packages/blockcluster/index.html

6.3. CloHe

Clustering of Mixed data
KEYWORDS: Classification - Clustering - Missing data

FUNCTIONAL DESCRIPTION: Software of classification for mixed data with missing values with application to multispectral satellite image time-series

- Partners: CNRS - INRA
- Contact: Serge Iovleff
- URL: https://modal.lille.inria.fr/CloHe/

6.4. PACBayesianNMF

KEYWORDS: Statistics - Machine learning
FUNCTIONAL DESCRIPTION: Implementing NMF with a PAC-Bayesian approach relying upon block gradient descent

- Participants: Benjamin Guedj and Astha Gupta
- Contact: Benjamin Guedj
- URL: https://github.com/astha736/PACbayesianNMF

6.5. pycobra

KEYWORDS: Statistics - Data visualization - Machine learning
SCIENTIFIC DESCRIPTION: pycobra is a python library for ensemble learning, which serves as a toolkit for regression, classification, and visualisation. It is scikit-learn compatible and fits into the existing scikit-learn ecosystem.

pycobra offers a python implementation of the COBRA algorithm introduced by Biau et al. (2016) for regression.

Another algorithm implemented is the EWA (Exponentially Weighted Aggregate) aggregation technique (among several other references, you can check the paper by Dalalyan and Tsybakov (2007).

Apart from these two regression aggregation algorithms, pycobra implements a version of COBRA for classification. This procedure has been introduced by Mojirsheibani (1999).

pycobra also offers various visualisation and diagnostic methods built on top of matplotlib which lets the user analyse and compare different regression machines with COBRA. The Visualisation class also lets you use some of the tools (such as Voronoi Tesselations) on other visualisation problems, such as clustering.

- Participants: Bhargav Srinivasa Desikan and Benjamin Guedj
- Contact: Benjamin Guedj
- URL: https://github.com/bhargavvader/pycobra

6.6. STK++

Statistical ToolKit
KEYWORDS: Statistics - Linear algebra - Framework - Learning - Statistical learning
FUNCTIONAL DESCRIPTION: STK++ (Statistical ToolKit in C++) is a versatile, fast, reliable and elegant collection of C++ classes for statistics, clustering, linear algebra, arrays (with an API Eigen-like), regression, dimension reduction, etc. The library is interfaced with lapack for many linear algebra usual methods. Some functionalities provided by the library are available in the R environment using rtkpp and rtkore.

STK++ is suitable for projects ranging from small one-off projects to complete data mining application suites.

- Participant: Serge Iovleff
- Contact: Serge Iovleff
- URL: http://www.stkpp.org

6.7. rtkore

STK++ core library integration to R using Rcpp

KEYWORDS: C++ - Data mining - Clustering - Statistics - Regression

FUNCTIONAL DESCRIPTION: STK++ (http://www.stkpp.org) is a collection of C++ classes for statistics, clustering, linear algebra, arrays (with an Eigen-like API), regression, dimension reduction, etc. The integration of the library to R is using Rcpp. The rtkore package includes the header files from the STK++ core library. All files contain only templated classes or inlined functions. STK++ is licensed under the GNU LGPL version 2 or later. rtkore (the stkpp integration into R) is licensed under the GNU GPL version 2 or later. See file LICENSE.note for details.

- Participant: Serge Iovleff
- Contact: Serge Iovleff
- URL: https://cran.r-project.org/web/packages/rtkore/index.html

6.8. Platforms

6.8.1. MASSICCC Platform

MASSICCC is a demonstration platform giving access through a SaaS (service as a software) concept to data analysis libraries developed at Inria. It allows to obtain results either directly through a website specific display (specific and interactive visual outputs) or through an R data object download. It started in October 2015 for two years and is common to the Modal team (Inria Lille) and the Select team (Inria Saclay). In 2016, two packages have been integrated: Mixmod and MixtComp (see the specific section about MixtComp). In 2017, the BlockCluster package has been integrated and also a particular attention to provide meaningful graphical outputs (for Mixmod, MixtComp and BlockCluster) directly in the web platform itself has led to some specific developments.

<p align="center" style="color:red">NON-A Project-Team</p>

6. New Software and Platforms

6.1. Blimp

FUNCTIONAL DESCRIPTION: Scientific research and development on the control of autonomous airship have shown a significant growth in recent years. New applications appear in the areas such as freight carrier, advertising, monitoring, surveillance, transportation, military and scientific research. The control of autonomous airship is a very important problem for the aerial robots research.

The development of Blimp by Non-A is used for experimentation and demonstration of controlling algorithms. The blimp is required to provide some environment information and status of itself, such as surveillance video of surrounding environment, gesture of blimp, altitude of blimp. With these basic information, one could localize blimp with certain algorithm (visual SLAM for example) or implement one controller in order to improve the stability and maneuverability of blimp.

- Contact: Jean-Pierre Richard

6.2. SLIM

FUNCTIONAL DESCRIPTION: Multi-robots cooperation can be found as an application in many domains of science and technology: manufacturing, medical robotics, personal assistance, military/security and spatial robots. The market of robots is quickly developing and its capacity is continuously growing. Concerning cooperation of mobile multi-robots, 3 key issues have to be studied: Localization, path planning and robust control, for which Non-A team has worked and proposed new algorithms. Due to the ADT SLIM, we implement our algorithms (localization, path planning and robust control) and integrate them into ROS (Robotic Operating System) as a package, named SLIM.

- Contact: Jean-Pierre Richard

RAPSODI Project-Team (section vide)

<div style="text-align:center; color:red;">

RMOD Project-Team

</div>

6. New Software and Platforms

6.1. Moose

Moose: Software and Data Analysis Platform

KEYWORDS: Software engineering - Meta model - Software visualisation

FUNCTIONAL DESCRIPTION: Moose is an extensive platform for software and data analysis. It offers multiple services ranging from importing and parsing data, to modeling, to measuring, querying, mining, and to building interactive and visual analysis tools. The development of Moose has been evaluated to 200 man/year.

Mots-cles : MetaModeling, Program Visualization, Software metrics, Code Duplication, Software analyses, Parsers

- Participants: Anne Etien, Nicolas Anquetil, Olivier Auverlot and Stéphane Ducasse
- Partners: Université de Berne - Sensus - Synectique - Pleiad - USI - Vrije Universiteit Brussel
- Contact: Stéphane Ducasse
- URL: http://www.moosetechnology.org

6.2. Pharo

KEYWORDS: Live programmation objet - Reflective system - Web Application

FUNCTIONAL DESCRIPTION: Pharo is a pure object reflective and dynamic language inspired by Smalltalk. In addition, Pharo comes with a full advanced programming environment developed under the MIT License. It provides a platform for innovative development both in industry and research. By providing a stable and small core system, excellent developer tools, and maintained releases, Pharo's goal is to be a platform to build and deploy mission critical applications, while at the same time continue to evolve. Pharo 60 got 100 contributors world-wide. It is used by around 30 universities, 15 research groups and around 40 companies.

RELEASE FUNCTIONAL DESCRIPTION: Inspector/Playground/Spotter are new moldable development tools for inspecting, coding and searching objects. Slots model instance variables as first class entities and enable meta-programming on this level. ShoreLine reporter introduces a way to report system errors and collect statistics, that we will use for future improvements Dark theme.

- Participants: Christophe Demarey, Clement Bera, Damien Pollet, Esteban Lorenzano, Marcus Denker and Stéphane Ducasse
- Partners: Université de Berne - Cadence - Inceptive - Netstyle - Feenk - ObjectProfile - GemstoneSystems - Greyc Université de Caen - Basse-Normandie - BetaNine - Yesplan - RMod - Pleiad - Synectique - Sensus - Université de Bretagne Occidentale - École des Mines de Douai - Reveal
- Contact: Marcus Denker
- URL: http://www.pharo.org

6.3. Pillar

KEYWORDS: HTML - LaTeX - HTML5

FUNCTIONAL DESCRIPTION: Pillar is a markup syntax and associated tools to write and generate documentation and books. Pillar is currently used to write several books and other documentation. Two platforms have already been created on top of Pillar: PillarHub and Marina.

- Contact: Stéphane Ducasse
- URL: https://github.com/Pillar-markup/pillar

<p align="center" style="color:red">SEQUEL Project-Team</p>

6. New Software and Platforms

6.1. BAC

Bayesian Policy Gradient and Actor-Critic Algorithms

KEYWORDS: Machine learning - Incremental learning - Policy Learning

FUNCTIONAL DESCRIPTION: To address this issue, we proceed to supplement our Bayesian policy gradient framework with a new actor-critic learning model in which a Bayesian class of non-parametric critics, based on Gaussian process temporal difference learning, is used. Such critics model the action-value function as a Gaussian process, allowing Bayes' rule to be used in computing the posterior distribution over action-value functions, conditioned on the observed data. Appropriate choices of the policy parameterization and of the prior covariance (kernel) between action-values allow us to obtain closed-form expressions for the posterior distribution of the gradient of the expected return with respect to the policy parameters. We perform detailed experimental comparisons of the proposed Bayesian policy gradient and actor-critic algorithms with classic Monte-Carlo based policy gradient methods, as well as with each other, on a number of reinforcement learning problems.

- Contact: Michal Valko
- URL: https://team.inria.fr/sequel/Software/BAC/

6.2. GuessWhat?!

GuessWhat?! Visual object discovery through multi-modal dialogue

KEYWORDS: Deep learning - Dialogue System

FUNCTIONAL DESCRIPTION: This project train a AI to play the GuessWhat?! game. Thus, you can train an AI to ask questions, to answer questions about images. You can also perform basic visual reasoning. This project is a testbed for future interactive dialogue system.

- Partner: Universite de Montreal
- Contact: Florian Strub
- Publications: GuessWhat?! Visual object discovery through multi-modal dialogue - End-to-end optimization of goal-driven and visually grounded dialogue systems Harm de Vries

6.3. Squeak

Sequential sampling for kernel matrix approximation

KEYWORD: Machine learning

- Contact: Daniele Calandriello
- URL: http://researchers.lille.inria.fr/~valko/hp/serve.php?what=publications/squeak.py

6.4. OOR

Optimistic Optimization in R

KEYWORDS: Black-box optimization - Machine learning

- Contact: Mickael Binois
- URL: https://cran.r-project.org/web/packages/OOR/index.html

SPIRALS Project-Team

6. New Software and Platforms

6.1. APISENSE

KEYWORDS: Mobile sensing - Crowd-sensing - Mobile application - Crowd-sourcing - Android

FUNCTIONAL DESCRIPTION: APISENSE platform is a software solution to collect various contextual information from Android devices (client application) and automatically upload collected data to a server (deployed as a SaaS). APISENSE is based on a Cloud computing infrastructure to facilitate datasets collection from significant populations of mobile users for research purposes.

- Participants: Antoine Veuiller, Christophe Ribeiro, Julien Duribreux, Nicolas Haderer and Romain Rouvoy
- Partner: Université Lille 1
- Contact: Romain Rouvoy
- URL: http://apisense.io

6.2. Nopol

KEYWORD: Automatic software repair

FUNCTIONAL DESCRIPTION: Nopol is an automatic software repair tool for buggy conditional statements (i.e., if-then-else statements) in Java programs. Nopol takes a buggy program as well as a test suite as input and generates a patch with a conditional expression as output. The test suite is required to contain passing test cases to model the expected behavior of the program and at least one failing test case that reveals the bug to be repaired. The process of Nopol consists of three major phases. First, Nopol employs angelic fix localization to identify expected values of a condition during the test execution. Second, runtime trace collection is used to collect variables and their actual values, including primitive data types and objected-oriented features (e.g., nullness checks), to serve as building blocks for patch generation. Third, Nopol encodes these collected data into an instance of a Satisfiability Modulo Theory (SMT) problem, then a feasible solution to the SMT instance is translated back into a code patch.

- Contact: Martin Monperrus
- URL: https://github.com/SpoonLabs/nopol/

6.3. PowerAPI

KEYWORDS: Energy efficiency - Energy management

FUNCTIONAL DESCRIPTION: PowerAPI is a library for monitoring the energy consumption of software systems.

PowerAPI differs from existing energy process-level monitoring tool in its software orientation, with a fully customizable and modular solution that let the user to precisely define what he/she wants to monitor. PowerAPI is based on a modular and asynchronous event-driven architecture using the Akka library. PowerAPI offers an API which can be used to define requests about energy spent by a process, following its hardware resource utilization (in term of CPU, memory, disk, network, etc.).

- Participants: Adel Noureddine, Loïc Huertas, Maxime Colmant and Romain Rouvoy
- Contact: Romain Rouvoy
- URL: http://powerapi.org

6.4. Saloon

KEYWORDS: Feature Model - Software Product Line - Cloud computing - Model-driven engineering - Ontologies

FUNCTIONAL DESCRIPTION: Saloon is a framework for the selection and configuration of Cloud providers according to application requirements. The framework enables the specification of such requirements by defining ontologies. Each ontology provides a unified vision of provider offers in terms of frameworks, databases, languages, application servers and computational resources (i.e., memory, storage and CPU frequency). Furthermore, each provider is related to a Feature Model (FM) with attributes and cardinalities, which captures its capabilities. By combining the ontology and FMs, the framework is able to match application requirements with provider capabilities and select a suitable one. Specific scripts to the selected provider are generated in order to enable its configuration.

- Participants: Clément Quinton, Daniel Romero Acero, Laurence Duchien, Lionel Seinturier and Romain Rouvoy
- Partner: Université Lille 1
- Contact: Lionel Seinturier
- URL: https://gitlab.irisa.fr/drome00A/saloon

6.5. SPOON

KEYWORDS: Java - Code analysis

FUNCTIONAL DESCRIPTION: Spoon is an open-source library that enables you to transform (see below) and analyze Java source code (see example) . Spoon provides a complete and fine-grained Java metamodel where any program element (classes, methods, fields, statements, expressions. . .) can be accessed both for reading and modification. Spoon takes as input source code and produces transformed source code ready to be compiled.

- Participants: Gérard Paligot, Lionel Seinturier, Martin Monperrus and Nicolas Petitprez
- Contact: Martin Monperrus
- URL: http://spoon.gforge.inria.fr