RESEARCH CENTER
Lille - Nord Europe

FIELD

Activity Report 2015

Section Application Domains

Edition: 2016-03-21

ATEAMS Project-Team (section vide)

DREAMPAL Project-Team (section vide)

DOLPHIN Project-Team

4. Application Domains

4.1. Transportation and logistics

- **Scheduling problems under uncertainty:** The flow-shop scheduling problem is one of the most well-known problems from scheduling. However, most of the works in the literature use a deterministic single-objective formulation. In general, the minimized objective is the total completion time (makespan). Many other criteria may be used to schedule tasks on different machines: maximum tardiness, total tardiness, mean job flowtime, number of delayed jobs, maximum job flowtime, etc. In the DOLPHIN project, a bi-criteria model, which consists in minimizing the makespan and the total tardiness, is studied. A bi-objective flow-shop problem with uncertainty on the duration, minimizing in addition the maximum tardiness, is also studied. It allows us to develop and test multi-objective (and not only bi-objective) optimization methods under uncertanty.
- **Routing problems under uncertainty:** The vehicle routing problem (VRP) is a well-known problem and it has been studied since the end of the fifties. It has a lot of practical applications in many industrial areas (ex. transportation, logistics, etc). Existing studies of the VRP are almost all concerned with the minimization of the total distance only. The model studied in the DOLPHIN project introduces a second objective, whose purpose is to balance the length of the tours. This new criterion is expressed as the minimization of the difference between the length of the longest tour and the length of the shortest tour. Uncertainty on the demands has also been introduced in the model.

4.2. Bioinformatics and Health care

Bioinformatic research is a great challenge for our society and numerous research entities of different specialities (biology, medical or information technology) are collaborating on specific themes.

4.2.1. Genomic and post-genomic studies

Previous studies of the DOLPHIN project mainly deal with genomic and postgenomic applications. These have been realized in collaboration with academic and industrial partners (IBL: Biology Institute of Lille; IPL: Pasteur Institute of Lille; IT-Omics firm).

First, genomic studies aim at analyzing genetic factors which may explain multi-factorial diseases such as diabetes, obesity or cardiovascular diseases. The scientific goal was to formulate hypotheses describing associations that may have any influence on diseases under study.

Secondly, in the context of post-genomic, a very large amount of data are obtained thanks to advanced technologies and have to be analyzed. Hence, one of the goals of the project was to develop analysis methods in order to discover knowledge in data coming from biological experiments.

These problems can be modeled as classical data mining tasks (Association rules, feature selection). As the combinatoric of such problems is very high and the quality criteria not unique, we proposed to model these problems as multi-objective combinatorial optimization problems. Evolutionary approaches have been adopted in order to cope with large scale problems.

Nowadays the technology is still going fast and the amount of data increases rapidly. Within the collaboration with Genes Diffusion, specialized in genetics and animal reproduction for bovine, swine, equine and rabbit species, we study combinations of Single Nucleotide Polymorphisms (SNP) that can explain some phenotypic characteristics. Therefore feature selection for regression is addressed using metaheuristics.

4.2.2. Optimization for health care

The collaboration with the Alicante company, a major actor in the hospital decision making, deals with knowledge extraction by optimization methods for improving the process of inclusion in clinical trials. Indeed, conducting a clinical trial, allowing for example to measure the effectiveness of a treatment, involves selecting a set of patients likely to participate to this test. Currently existing selection processes are far from optimal, and many potential patients are not considered. The objective of this collaboration consists in helping the practitioner to quickly determine if a patient is interesting for a clinical trial or not. Exploring different data sources (from a hospital information system, patient data...), a set of decision rules have to be generated. For this, approaches from multi-objective combinatorial optimization are implemented, requiring extensive work to model the problem, to define criteria optimization and to design specific optimization methods.

4.2.3. Molecular sampling and docking on large hybrid clusters

A Phd thesis is started in September 2015 in this context in collaboration with UMONS and University of Strasbourg. Flexible molecular docking is a very complex combinatorial opitmization problem especially when two components (ligand and protein) involved in the mechanism are together flexible. To deal in a reasonable time with such highly combinatorial process approximate optimization methods and massively parallel computing are absolutely The focus of the Ph.D thesis is on the flexibility-aware modeling and the design and implementation of near-approached optimization methods for solving the docking problem on large hybrid clusters including GPU accelerators and MIC coprocessors.

INOCS Team

4. Application Domains

4.1. Introduction

It is hard to find an aspect of our modern-day economy whose design, management and control do not critically depend on the solution of one or more CS decision problems. Even if they are pervasive, many of them are still not "satisfactorily" solved and constitute a strong challenge to research teams nowadays. The innovative research goals of INOCS have, without doubt, a strategic importance in the application field. CS problems appear in a broad range of application fields such as the next one cited hereafter.

1. *the energy sector* where decisions of distinct nature such as production and distribution are jointly determined;
2. *supply chain management* where location and routing decisions have to be defined jointly even if they refer to different time horizons;
3. *revenue management* where the determination of prices for services or products requires to take explicitly into account the strategic consumers' behaviour.

MEPHYSTO Team

4. Application Domains

4.1. Mechanics of heterogeneous media

The mechanics of heterogeneous materials aims at characterizing the macroscopic properties of heterogeneous materials using the properties of their constituents.

The homogenization theory is a natural tool for this task. In particular, for linear problems (linear conductivity or linear elasticity), the macroscopic properties are encoded into a single (conductivity or elasticity) homogenized tensor. The numerical approximation of this homogenized tensor is a typical objective of quantitative homogenization.

For nonlinear problems, such as rubber elasticity, the macroscopic properties are no longer characterized by a single tensor, but rather by a nonlinear energy density. Our aim is to relate qualitatively and quantitatively the (precise but unpractical) statistical physics picture to explicit macroscopic constitutive laws that can be used for practical purposes. This endeavor is relevant both in science and technology. The rigorous derivation of rubber elasticity from polymer-physics was indeed emphasized by John Ball as an important open problem of nonlinear elasticity in his survey [44] on the field. Its solution could shed light on some aspects of polymer-physics. The associated ab initio derivation of constitutive laws (as an alternative to phenomenological laws) would also be of interest to computational mechanics and rubber industry.

For this application domain, we work in close collaboration with physicists (François Lequeux, ESPCI) and researchers from mechanics and computational mechanics (Patrick Le Tallec, Ecole polytechnique).

4.2. Numerical simulation in heterogeneous media

Solving numerically PDEs in highly heterogeneous media is a problem encountered in many situations, such as the transport of pollutants or the design of oil extraction strategies in geological undergrounds. When such problems are discretized by standard numerical methods the number of degrees of freedom may become prohibitive in practice, whence the need for other strategies.

Numerical solution methods inspired by asymptotic analysis are among the very few feasible alternatives, and started fifteen years ago with the contributions of Hou and Wu [59] , Arbogast [40] etc. We refer to [54], [70],[3] for a recent state of the art. Numerical homogenization methods usually amount to looking for the solution of the problem (1) in the form $u_\varepsilon(x) \simeq u_0(x) + \varepsilon \nabla u_0(x) \cdot \Phi(x, \frac{x}{\varepsilon})$, where Φ is a proxy for the corrector field computed locally at point x (in particular, one does not use explicitly that the problem is periodic so that the method can be used for more general coefficients) and u_0 is a function which does not oscillate at scale.

Relying on our quantitative insight in stochastic homogenization, a first task consists in addressing the three following prototypical academic examples: periodic, quasi-periodic, and stationary ergodic coefficients with short range dependence. The more ambitious challenge is to address more complex coefficients (of interest to practioners), and design adaptive and efficient algorithms for diffusion in heterogeneous media.

4.3. Laser physics

Our contribution to the analysis of models in laser physics is motivated by the LabEx CEMPI (Centre Européen pour les Mathématiques, la Physique et leurs Interactions, a large eight-year research and training project approved by the French government in February 2012 as a "Laboratoire d'Excellence" and an initiative of mathematicians and physicists of the Université Lille 1). For this application domain, we work in close collaboration with physicists, which ensures our direct impact on these scientific issues. We focus on two applications: optical fibers and cold atoms.

In collaboration with physicists from the PhLAM laboratory in Lille, we aim at developing new techniques for the numerical integration of a family of 1D Schrödinger-like equations modelling the propagation of laser pulses in optical fibers. The questions arising are challenging since physicists would like to have fairly fast and cheap methods for their problems, with correct qualitative and quantitative behaviors. Another point is that they are interested in methods and codes that are able to handle different physical situations, hence different terms in the NLS equation. To meet these requirements, we will have to use numerical time-integration techniques such as splitting methods or exponential Runge-Kutta methods, space discretization techniques such as finite differences and fast Fourier transforms, and absorbent boundary conditions. Our goal, together with the physicists is to be able to reproduce numerically the results of the experiments they make in actual optical fibers, and then to be able to tune parameters numerically to get more insight into the appearance of rogue waves beyond the dispersive blowup phenomenon.

Recall that the Schrödinger equation also describes Bose-Einstein condensates. A second experimental team at PhLAM projects to study questions related to Anderson localization in such condensates. In fact, they will realize the "kicked rotor" (see [51]), which provides a paradigm for Anderson localization, in a Bose-Einstein condensate. We plan to collaborate with them on the theoretical underpinnings of their findings, which pose many challenging questions.

MODAL Project-Team

4. Application Domains

4.1. Application domains

Potential application areas of statistical modeling for mixed data are (credit scoring, marketing, environment, medical, economic, hydrology,...), but MODAL favors applications related to biology, phylogeny, genetics and medicine. Members of the team are already experienced in these directions with complementary skills.

NON-A Project-Team (section vide)

RAPSODI Team

4. Application Domains

4.1. Porous media flows

Porous media flows are of great interest in many contexts, like, e.g., oil engineering, water resource management, nuclear waste repository management, or carbon dioxyde sequestration. We refer to [47], [46] for an extensive discussion on porous media flow models.

From a mathematical point of view, the transport of complex fluids in porous media often leads to possibly degenerate parabolic conservation laws. The porous rocks can be highly heterogeneous and anisotropic. Moreover, the grids on which one intends to solve numerically the problems are prescribed by the geological data, and might be non-conformal with cells of various shapes. Therefore, the schemes used for simulating such complex flows must be particularly robust.

4.2. Corrosion

The concept for long term storage of high-level radioactive waste in France under study is based on an underground repository. The waste shall be confined in a glass matrix and then placed into cylindrical steel canisters. These containers shall be placed into micro-tunnels in the highly impermeable Callovo-Oxfordian claystone layer at a depth of several hundred meters. At the request of the French nuclear waste management agency ANDRA, investigations are conducted to optimize and finalize this repository concept with the aim to ensure its long-term safety and its reversibility. The long-term safety assessment of the geological repository has to take into account the degradation of the carbon steel used for the waste overpacks and the cell disposal liners, which are in contact with the claystone formation. This degradation is mainly caused by generalized corrosion processes which form a passive layer on the metal surface consisting of a dense oxide inner layer and a porous hydroxide outer layer in contact with the groundwater in the pore space of the claystones. The processes take place under anaerobic conditions, since the groundwater is anoxic.

As a tool to investigate the corrosion processes at the surface of the carbon steel canisters, the Diffusion Poisson Coupled Model (DPCM) for corrosion has been developed by Bataillon *et al.* [1]. The numerical approximation of this corrosion model and some associated models by accurate and efficient methods is challenging. Theoretical study of the models (existence of solutions, long time behavior) is also worth of interest.

4.3. Complex fluid flow simulations

The team is interested in some numerical methods for the simulation of systems of PDEs describing complex flows, like for instance, mixture flows or granular gases.

Let us first focus on fluid mixture flows. The fluid is described by its density, its velocity and its pressure. These quantities obey mass and momentum conservation. On the one hand, when we deal with the 2D variable density incompressible Navier-Stokes equations, we aim to study the ability of the numerical scheme to reproduce some instabilities phenomena such as the Rayleigh-Taylor instability. On the other hand, diffuse interface models have gained renewed interest for the last few years in fluid mechanics applications. From a physical viewpoint, they allow to describe some phase transition phenomena. If the Fick's law relates the divergence of the velocity field to derivatives of the density, one obtains the so called Kazhikhov-Smagulov model [61]. Here, the density of the mixture is naturally highly non homogeneous, and the constitutive law accounts for diffusion effects between the constituents of the mixture. The first phenomena that we want to reproduce are the powder-snow avalanches. We investigate the influence of the characteristics parameters (Froude, Schmidt and Reynolds numbers) on the progression of the front. Other similar hydrodynamic models arise in combustion theory or transport of pollutants.

Kinetic theory of molecular gases models a gas as a system of elastically colliding spheres, conserving mechanical energy during impact. Once initialized, it takes a molecular gas not more than few collisions per particle to relax to its equilibrium state, characterized by a Maxwellian velocity distribution and a certain homogeneous density (in the absence of external forces). A granular gas is a system of dissipatively colliding, macroscopic particles (grains). This slight change in the microscopic dynamics (converting energy into heat) cause drastic changes in the behavior of the gas: granular gases are open systems, which exhibits self-organized spatio-temporal cluster formations, and has no equilibrium distribution. They can be used to model silos, avalanches, pollen or planetary rings.

4.4. Stratigraphy

The knowledge of the geology is a prerequisite before simulating flows within the subsoil. Numerical simulations of the geological history thanks to stratigraphy numerical codes allow to complete the knowledge of the geology where experimental data are lacking. Stratigraphic models consist in a description of the erosion and sedimentation phenomena at geological scales.

The characteristic time scales for the sediments are much larger than the characteristic time scales for the water in the river. However, the (time-averaged) water flux plays a crucial role in the evolution of the stratigraphy. Therefore, defining appropriate models that take the coupling between the rivers and the sediments into account is fundamental and challenging. Once the models are at hand, efficient numerical methods must be developed.

4.5. Low frequency electromagnetism

Numerical simulation is nowadays an essential tool in order to design electromagnetic systems, by estimating the electromagnetic fields generated in a wide variety of devices. An important challenge for many applications is to quantify the intensity of the electric field induced in a conductor by a current generated in its neighborhood. In the low-frequency regime, we can for example quote the study of the impact on the human body of a high-tension line or, for higher frequencies, the one of a smartphone. But the ability to simulate accurately some electromagnetic fields is also very useful for non destructive control, in the context of the maintenance of nuclear power stations for example. The development of efficient numerical tools, among which the so-called“*a posteriori* error estimators", is consequently necessary to reach a high precision of calculations in order to provide estimations as reliable as possible.

SEQUEL Project-Team

4. Application Domains

4.1. Recommendation systems in a broad sense

Recommendation systems have been a major field of applications of our research for a few years now. Recommendation systems should be understood in a broad sense, as systems that aim at providing personalized responses/items to users, based on their characteristics, and the environment in which the interaction happens.

In that broad sense, we have collaborated with companies on computational advertizing and recommendation systems. These collaborations have involved research studies on the following issues:

- cold-start problem,
- time varying environment,
- ability to deal with large amounts of users and items,
- ability to design algorithms to respond within a reasonnable amount of time, usually below 1 millisecond.

We have also competed in challenges, winning some of them [0], and we have also organized a challenge [0], on those topics.

A company has been awarded an innovation award in 2015, thanks to the research work done in collaboration with SEQUEL (*cf.* sec. 1).

In these works, we develop an original [0] point of view on such systems. While traditional (before say 2010) recommendation systems were seen as solving a supervised learning task, or a ranking task, we have developed the idea that recommender systems are really a problem of sequential decision making under uncertainty.

We also started a new work aiming to introduce deep learning in recommender systems. An engineer (Florian Strub) was recruited to work on this topic and presented some results at the NIPS'2015 workshop on "Machine Learning for (e-)Commerce". Moreover we released some code to handle sparse data with the Torch7 framework and GPUs https://github.com/fstrub95/nnsparse.

4.2. Spoken dialog systems

A Spoken Dialogue System (SDS) is a system enabling human people to interact with machines through speech. In contrast with command-and-control systems or question-answering systems that react to a single utterances, SDS build a real interaction over time and try to achieve complex tasks (like hotel booking, appointment scheduling etc.) by gathering pieces of information through several turns of dialogue. To do so, besides the required speech and language processing modules (*e.g.* speech recognition and synthesis, language understanding and generation), there is a need for a dialogue management module that decides what to say in any situation so as to achieve the goal in the most natural and efficient way, recovering from speech processing errors in a seamless manner.

The dialogue management module is thus taking sequences of decisions to achieve a long-term goal in an unknown, noisy and hard to model environment (since it includes human users). For this reason, we work on machine learning techniques such as reinforcement and imitation learning to optimize this specific sequential decision making under uncertainty problem.

[0] SEQUEL ranked first and second at the "Pascal Exploration & Exploitation Challenge 2011"; SEQUEL ranked first at the "RecSys Challenge 2014: User Engagement as Evaluation".

[0] ICML 2012 new Challenges for Exploration & Exploitation 3.

[0] the originality fades away as years pass since this idea is exploited by other researchers.

In addition to bring novel and efficient solutions to this problem, we are interested in the new challenges brought to our research in machine learning by this type of application. Indeed, having the human in the learning loop typically requires dealing with non-stationarity, data-efficiency, safety as well as cooperation and imitation.

We collaborate with companies such as Orange Labs on this topic and several projects are ongoing (ANR MaRDi, CHIST-ERA IGLU). We will also be participating to a H2020 project on human robot-interaction starting in 2016 (BabyRobot). We organised a workshop at ICML this year: Machine Learning for Interactive Systems (MLIS). Olivier Pietquin was invited as a panelist at the NIPS Workshop on spoken language understanding and dialogue.

4.3. Adaptive/learning systems more generally

Reinforcement learning leads to the design of systems that adapts their behavior to their environment, hence adaptive systems. We have worked on various applications of this idea, beyond the two main applications domaines mentioned above (recommendation systems, and spoken dialog systems). Let us briefly mention: educative tutoring systems; adaptive heating system in buildings; players that adapt their strength to that of their human opponent; bioreactor.

4.4. Prediction in general

Since the goal of our research is to design systems that learn to act in an optimal way in their environment, prediction is a major issue. Hence, we are doing some research activities on this particular task, without always being in direct connection with learning a policy.

We have done some research in the area of prediction web-server load in a non stationary environment. We also have activities in the prediction in bug in software code.

BONSAI Project-Team

4. Application Domains

4.1. Life Sciences and health

Our research plays a pivotal role in all fields of life sciences and health where genomic data are involved. This includes more specifically the following topics: plant genomics (genome structure, evolution, microRNAs), cancer (leukemia, mosaic tumors), drug design (NRPSs), environment (metagenomics and metatranscriptomics), virology (evolution, RNA structures) ...

FUN Project-Team

4. Application Domains

4.1. Application Domains

The set of applications enabled through FUN and IoT is very large and can apply in every application area. We can thus not be exhaustive but among the most spread applications, we can name every area, event or animal monitoring, understanding and protection. To illustrate this, we may refer to the use cases addressed by our PREDNET project which goals is to equip rhinoceros with smart communicating devices to fight against poaching.

Other field of application is exploration of hostile and/or unknown environment by a fleet of self-organizing robots that cooperate with RFID and sensors to ensure a continue monitoring afterwards.

Also, IoT and FUN ca play a key role in logistics and traceability by relying on the use of sensors or RFID technologies as implemented in our TRACAVERRE project or our collaboration with the start up TRAXENS.

Finally, IoT and FUN leverage a lot of applications in Smart City concept , ranging from parking aid to a better energy consumption going through air quality monitoring, traffic fluidizing etc. (See our CityLab Inria and VITAL projects).

RMOD Project-Team

4. Application Domains

4.1. Programming Languages and Tools

Many of the results of RMoD are improving programming languages or development tools for such languages. As such the application domain of these results is as varied as the use of programming languages in general. Pharo, the language that RMoD develops, is used for a very broad range of applications. From pure research experiments to real world industrial use (the Pharo Consortium has around 20 company members) http://consortium.pharo.org Examples are web applications, server backends for mobile applications or even graphical tools and embedded applications.

4.2. Software Reengineering

Moose is a language-independent environment for reverse- and re-engineering complex software systems. Moose provides a set of services including a common meta-model, metrics evaluation and visualization. As such Moose is used for analysing software systems to support understanding and continous development as well as software quality analysis.

SPIRALS Project-Team

4. Application Domains

4.1. Introduction

Although our research is general enough to be applied to many application domains, we currently focus on applications and distributed services for the retail industry and for the digital home. These two application domains are supported by a strong expertise in mobile computing and in cloud computing that are the two main target environments on which our research prototypes are build, for which we are recognized, and for which we have already established strong collaborations with the industrial ecosystem.

4.2. Distributed software services for the retail industry

This application domain is developed in relation with the PICOM (*Pôle de compétivité Industries du Commerce*) cluster. We have established strong collaborations with local companies in the context of former funded projects, such as Cappucino and Macchiato, which focused on the development of a new generation of mobile computing platforms for e-commerce. We are also involved in the Datalyse and OCCIware funded projects that define cloud computing environments with applications for the retail industry. Finally, our activities in terms of crowd-sensing and data gathering on mobile devices with the APISENSE® platform share also applications for the retail industry.

4.3. Distributed software services for the digital home

We are developing new middleware solutions for the digital home, in particular through our long standing collaboration with Orange Labs. We are especially interested in developing energy management and saving solutions with the POWERAPI software library for distributed environments such the ones that equip digital homes. We are also working to bridge the gap between distributed services hosted on home gateways and distributed services hosted on the cloud to be able to smoothly transition between both environments. This work is especially conducted with the SALOON platform.

DEFROST Team

4. Application Domains

4.1. Surgery

Surgical procedures are often carried out using instruments made from stiff materials that interact with delicate biological tissues such as internal organs, blood vessel walls and small cavities. This is one of the source of danger for many surgical procedures. Soft-robotics open up new perspectives in minimally invasive approaches. Thanks to the highly deformability of their structure, similar to organic materials, and their motion, created by deformation in the same way as the muscles in living animals, they offer many advantage for surgical applications. Recent work anticipates that their compliant nature and their large number of degrees of freedom will provide key surgical positive outcomes:

- Improving the capacity of access with security to the fragile parts of the anatomy by applying less pressure to the anatomical walls
- Easy maneuvering through soft and confined spaces allowing new Minimally Invasive Surgery approaches.

These positive outcomes are expected given the properties of soft-robot. In a recent state-of-the art reports on soft robotics, surgery in the list of *killer applications* of soft-robotics. However, the lack of existing methodology for modeling and control remains an obstacle to be proved by a practical implementation. Given our background on surgical simulations: soft tissue and tool/tissues contact models we are particularly well positioned to address the challenge of using soft-robots in surgery.

4.2. Industry

Robotics in the manufacturing industry is already highly diffused and is one of the ways put forward to maintain the level of competitiveness of companies based in France and to avoid relocation in cheap labor countries. Yet, in France, it is considered that the level of robotization is insufficient compared to Germany for instance. One of the challenge is the high investment cost for buying robotic arms. In the recent years, it has led the development of « generic » and « flexible » (but rigid) robotic solution that can be produced in series. But their applicability to specific tasks is still challenging or too costly. The manufacturing of deformable robots could be very low compared to classical rigid robotics. Moreover, with the development of 3D printing, we can imagine the development of a complete opposite strategy: a « task-specific » design of robots. Given a task that need to be performed by a deformable robot: we would optimize the shape of its structure to create the set of desired motion (see in Challenge2: Exploring interactive and semi-automatic optimisation methods for design). An other remarkable property of soft-robots is their adaptability to fragile or tortuous environment. For some particular industry, this could also be an advantage compared to existing rigid solutions.

4.3. Personal and service robotics

The personal and service robotics are considered as an important challenge for industry in the coming years. The potential applications are numerous and particularly include the challenge of finding robotic solutions for active and healthy ageing at home. We plan to develop functional orthosis for which it is better not to have a rigid exoskeleton that are particularly not comfortable. These orthosis will be ideally personalised for each patient and built using rapid prototyping. Again the low manufacturing price and the robustness of deformable robots could be key advantages for this particular market. On this topic, the place of our team will be to provide algorithms for controlling the robots. We need to find some partners to build these wearable robots. Our team will also propose innovative technology for robotic games: we are currently working on a new technique of control for deformable puppets. If the project succeeds, a user will be able to build his/her own puppet with a 3D printer and control it with a Kinect. Finally, an other direction for the transfer of our research towards

society is art: soft-robotics seems a source of inspiration for artists. This year, we have been collaborating with the art school Le Fresnoy based at Tourcoing (near our Lab) and the result had a good impact for the visibility of our team. We may also collaborate in the close future with IRCAM in the context of the transversal project Inria-ART led by Arshia Cont and Laurent Grisoni.

LINKS Team

4. Application Domains

4.1. Linked Data Integration

There are many contexts in which integrating linked data is interesting. We advocate here one possible scenario, namely that of integrating business linked data to feed what is called Business Intelligence. The latter consists of a set of theories and methodologies that transform raw data into meaningful and useful information for business purposes (from Wikipedia). In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today's' enterprises and businessmen need to face the problem of information explosion, due to the Internet's ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of business intelligence within the enterprise.

4.2. Data Cleaning

The second example of application of our proposal concerns scientists who want to quickly inspect relevant literature and datasets. In such a case, local knowledge that comes from a local repository of publications belonging to a research institute (e.g. HAL) need to be integrated with other Web-based repositories, such as DBLP, Google Scholar, ResearchGate and even Wikipedia. Indeed, the local repository may be incomplete or contain semantic ambiguities, such as mistaken or missing conference venues, mistaken long names for the publication venues and journals, missing explanation of research keywords, and opaque keywords. We envision a publication management system that exploits both links between database elements, namely pointers to external resources and logical links. The latter can be complex relationships between local portions of data and remote resources, encoded as schema mappings. There are different tasks that such a scenario could entail such as (i) cleaning the errors with links to correct data e.g. via mappings from HAL to DBLP for the publications errors, and via mappings from HAL to Wikipedia for opaque keywords, (ii) thoroughly enrich the list of publications of a given research institute, and (iii) support complex queries on the corrected data combined with logical mappings.

4.3. Real Time Complex Event Processing

Complex event processing serves for monitoring nested word streams in real time. Complex event streams are gaining popularity with social networks such as with Facebook and Twitter, and thus should be supported by distributed databases on the Web. Since this is not yet the case, there remains much space for future industrial transfer related to Links' second axis on dynamic linked data.

MAGNET Team

4. Application Domains

4.1. Overview

The real-world problems we target include browsing, monitoring and mining in information networks. The discovered structures would also be beneficial to predicting links between users and texts which is at the core of recommender systems. More generally, all the learning tasks considered in the project such as node clustering, node and link classification and link prediction are likely to yield important improvements in these applications. Application domains cover natural language processing, social networks for cultural data and e-commerce, and biomedical informatics.

MINT Project-Team (section vide)

Mjolnir Team

4. Application Domains

4.1. Application Domains

Mjolnir works on fundamental aspects of Human-Computer Interaction that can be applied to diverse application domains. Our 2015 research concerned desktop, touch-based and mobile interfaces with notable applications to 3D animation, 2D illustration, clinical diagnosis and TV viewing experience.