

US

University of Sussex

MLCL Laboratory, CI-1 Room 203

MLCL Laboratory

Summary of Capabilities

Summary of MLCL Laboratory Offering

Our Capabilities

As a result of recent advances in the research field of computational linguistics, computers are rapidly becoming much more effective at processing text - therefore expanding the practical tasks that they can undertake.

We aim to stay on the leading edge of the field, figure out how to turn these new capabilities into practical solutions, and provide customers with the benefit of this expertise.

We provide consulting and software development services to companies that need to get a handle on very large sets of unstructured textual data. We work with a wide range of businesses, including specialist companies that themselves provide end-users with tools for handling such data.

We apply statistical approaches to the problems of encoding and interpreting text. We use specific technological approaches in the field of distributional semantics coupled with more widely understood machine learning methods. Our code platform consists of software developed in-house, combined with general algorithms for machine learning.

All of the proceeds from our work are used to fund research and to support the running of the University.

How we Work with Customers

Types of Project Undertaken

We only engage on projects where the project requirements are well aligned with our research interests and capabilities, and we believe can provide customers with what they are looking for in a cost-effective manner.

Scoping the Project

For a typical project, we identify in advance the required outcome and lay out a simple plan for achieving this goal. The cost is specified and agreed at this stage.

For larger projects, we break the project up into a series of phases. Costs for each phase are specified at this stage, but customers only commit to the next phase on completion of the previous phase. This allows companies to manage their overall cost exposure.

With larger projects, we also build flexibility into the plan. We are working in a rapidly evolving arena with companies whose business is usually likewise changing quickly; later phases of a project will therefore often be defined more precisely on the basis of the outcome of previous phases and any changes in the external environment.

Contracts are undertaken between the customer and the University of Sussex.

Project Team

A dedicated team is assigned to a project. We construct the team based on the specific requirements of a project. Team sizes vary, but are typically between 2 and 5 people.

Team members are all actively conducting research in the field and their time commitments are scoped to reflect this. One ancillary benefit is that it is not unusual for new research findings to be incorporated immediately into 'live' commercial projects.

Summary of MLCL Laboratory Offering

Typical Challenges Facing Clients

How can I search for specific entities ?

(Entity-Based Knowledge Management)

People often want to find documents that are referring to a particular entity (e.g. the company called Apple Inc.), but end up having to express that requirement as a hand-crafted query based on a suitable combination of words. These queries aim to eliminate spurious documents and include all relevant documents, but results are often poor.

Recent advances allow more sophisticated handling of this issue, which we are working to implement both internally and with customers.

Can we accurately interpret sentiment across a wide range of documents ? (Sentiment Analysis)

Sentiment analysis is the evaluation by computer of a document to determine the sentiment being expressed by the author. A typical implementation analyses whether a document's sentiment is broadly favourable, neutral, or unfavourable.

Applied rapidly to a large set of documents (e.g. web content), the approach can give an overview of what people are saying about a particular topic. This capability has been enthusiastically seized upon by many executives keen to understand what is being said online about their companies and their products.

Unfortunately, this first-generation technology is crude and liable to provide inaccurate assessments; we are developing ways to expand these capabilities, and as a result we have the ability to implement state-of-the-art sentiment analysis approaches.

Can I automate the process of categorising data ?

(Automated Classification)

Extracting knowledge from data today typically requires extensive categorisation by humans. In many circumstances, machine-learning techniques can be used to free up those people so they can undertake more valuable (and more interesting) work.

A related problem is that many installed computer systems automate less challenging instances of this type of decision making using hard-crafted heuristics; these heuristics can now be improved upon by using much more flexible machine-learning algorithms.

How do I find information meaningfully related to the information I have ? (Linking by Meaning)

Once people have information that they are interested in, they often want access to other information that is meaningfully related.

For example, they may want 'more of the same' or (alternatively) to avoid being troubled by duplicates. They may wish to be prompted to look at issues or entities that are meaningfully related but which they haven't yet considered, or to have the information organised by topic.

We have provided clients with working solutions addressing some of these challenges, using distributional semantics approaches, and are conducting research aimed at extending these techniques.