

GATE: Bridging the Gap between Terminology and Linguistics

Diana Maynard

University of Sheffield, UK

One of the current problems many terminologists face is the lack of suitable tools to process their data. In particular, since needs differ widely, general and robust tools are required that can be adapted to different language processing tasks and even languages. GATE is an open source architecture for language engineering, developed by the University of Sheffield and used worldwide by thousands of scientists, companies, teachers and students. One of its key outcomes is that over a decade of collecting reusable code and building a community has led to a mature ecosystem for solving language processing problems quickly, efficiently and cheaply. It contains a specialist Integrated Development Environment for language engineering R&D, enabling users to visualise, create and edit text, annotations, ontologies and parse trees, to create new linguistic processing resources and to construct applications from pre-defined components, and to perform evaluation and benchmarking on the results. GATE contains a set of ready-made linguistic processing resources such as tokenisers, POS taggers, parsers and so on which are theory-neutral and therefore widely applicable to a number of tasks in various languages. It has been used not only in many research projects but also in industrial settings by commercial users such as AstraZeneca, the Press Association, Garlik, Fizzback, Innovantage, Roche and Intelius, amongst many others. GATE and its components are a key tool in today's world of information and data overload, enabling users to perform tasks such as document management, business intelligence, information retrieval, question answering, and knowledge indexing, modelling and conceptualisation.