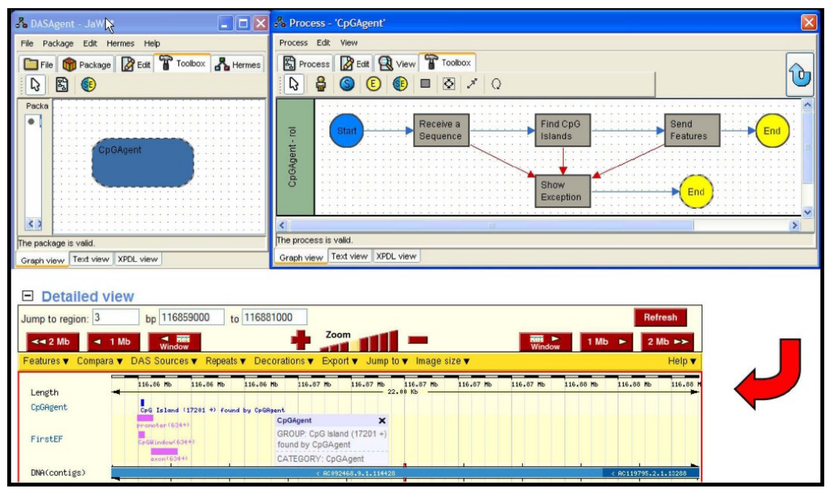
A DAS view of agent-based workflows

Ezio Bartocci[1],Emanuela Merelli[1],Steffen Möller[2]

[1]Dipartimento di Matematica e Informatica, Università di Camerino
[2]Institut für Neuro- und Bioinformatik, Universität zu Lübeck, Lübeck, Germany

Motivation

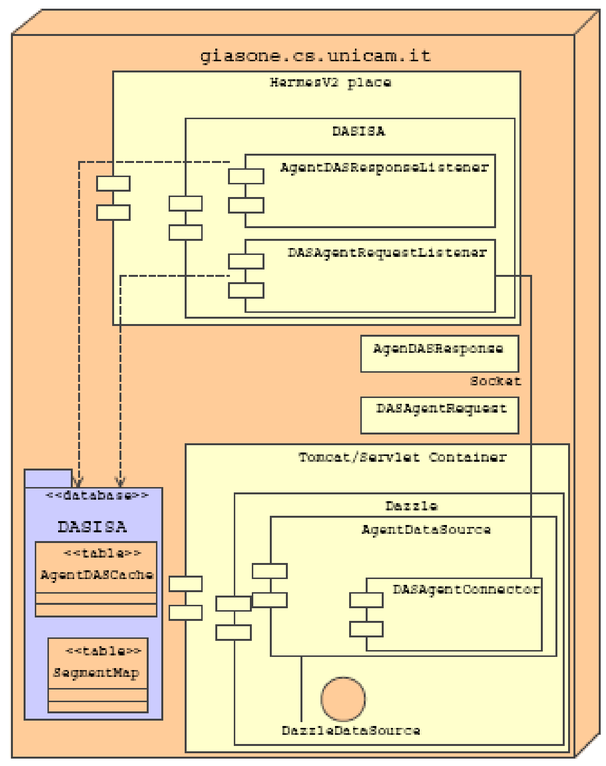
The continuous technological advancement in DNA sequencing led to an avalanche of data on biological sequences and their variants. Today, even several large mammalian genomes are fully sequenced and made available i.e. by the Open Source project Ensembl [1]. Sequence annotation could be a very complex task. New sequence features are often inferred after the interpretation, comparison or integration of several data sources and tools. The composition of several activities, described by a workflow, is becoming increasingly important and complex in bioinformatics. The avalanche of data, very often, is too huge to be transferred over the network (e.g. raw images of microarrays). Based on this issue, our approach uses mobile agents as executors of those workflow activities concerning the automation of sequence annotation. A mobile agent is a computational unit capable of migrating to different places from any location. An agent can behave in an opportunistic and reactive way. Agents do not require the user's presence and can be assigned a task to be exploited over distributed resources [2]. In the context of sequence annotation, the agent system should be interactively usable, both by a human and by other programs. This led to the selection of the BioDAS interface [3] for the exchange of biological information. The protocol allows the agent system to disguise itself as a regular DAS server. The data presentation and interactive interpretation is provided by independently developed DAS clients like the Ensembl contig view.



A DAS view of a CpG Agent workflow

Methods

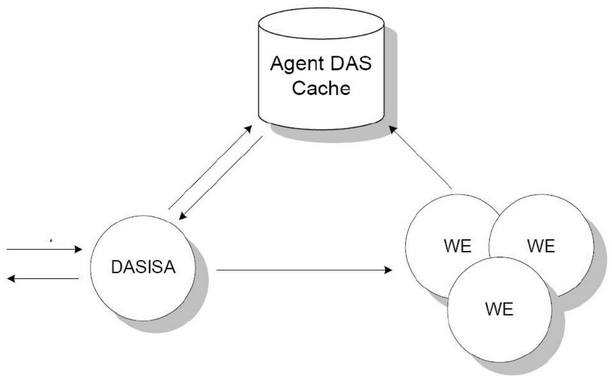
Our aim was to design an assistant agent -called DAS Interface Service Agent (DASISA)- capable to communicate with the external environment using the Distributed Annotation System [4] (DAS) protocol. DASISA to communicate with the DAS protocol, use the Dazzle (http://www.biojava.org/dazzle/) Java Servlet as a general purpose server for DAS protocol. Dazzle is a modular system providing access to several databases. For this work, we have considered an agents' activity results as a data source. Thus, we have developed an AgentDataSource component implementing DazzleDataSource interface. This component communicates through with DASISA exchanging two kind of XML messages: DASAgentRequest and AgentDASResponse. A DAS protocol request is translated by AgentDataSource in a DASAgentRequest document, a message comprehensible by DASISA. The DASAgentRequest document contains the name of the Agent to be created by DASISA and the specific sequence region (segment) id to be explored looking for new possible interesting features. An additional table called "SegmentMap" stores, for each possible segment id, the relative sequence, length and the start position in the inspected chromosome. After having retrieved the necessary segment informations, DASISA creates the agent and sends it a message with the sequence to analyse. Upon termination of the agent's job, its results are returned to DASISA which in turn converts these in an AgentDASResponse XML document. The agents' activity may require considerable computational effort. In order to be most responsive for repeated queries, a component called "AgentDASCache" has been implemented to collect the latest results. In the implementation we use BioAgent/Hermes [5] mobile agent-based middleware.



DAS Interface Service Agent Architecture

Results

DAS combined with agent technology provides all advantages of dynamic data generation and data integration to otherwise static DAS sources. The approach also addresses the integration of algorithms that may not be feasible to precompute for complete genomes. With the provision of schemas for caching, the individual interests of researchers in a particular family of genes or in a distinct disease-associated locus may be well-addressed.



Outline of DASISA annotations caching

References

1. Clamp, M., Andrews, D., Barker, D., et al. (2003) Ensembl 2002: accommodating comparative genomics., Nucleic Acids Res., 31(1): 38-42.
2. Hall, D., Miller, J., Arnold, J., Kochut, K., Sheth, A., and Weise, M. (1999) Using Workflow to Build an Information Management System for a Geographically Distributed Genome Sequencing Initiative., Genomics of Plants and Fungi, R.A. Prade and H.J.
3. Bohner, Editors. 4. Stein, L.D., Eddy, S., Dowell, R. Distributed Annotation System., (1999-2002) http://www.biodas.org/documents/spec.html
4. Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L. The distributed annotation system. BMC Bioinformatics. 2001;2:7. Epub 2001 Oct 10.
5. Corradini, F., Merelli, E. (2005A) Hermes: agent-based middleware for mobile computing., Mobile Computing, LNCS, 3465:

Aknowledgements:
This work was supported by the Italian FIRB-MIUR LITBIO: Laboratory for Interdisciplinary Technologies in Bioinformatics
http://www.litbio.org