

단백질 구조예측 지원을 위한 조정 에이전트의 설계

김현식, 예형석, 진훈, 김인철
경기대학교 전자계산학과
e-mail : advance7@kuic.kyonggi.ac.kr

Design of a Coordinator Agent for Supporting Protein Structure Prediction

Hyun-Sik Kim, Hyung-Seok Ye, Hoon Jin, In-Cheol Kim
Dept. of Computer Science, Kyonggi University

요 약

인터넷 상의 생물학 자원들을 이용하는 에이전트 시스템들에는 몇 가지 특징들이 있는데, 그 중 하나가 바로 사용자 에이전트와 공개 데이터베이스들을 대신하는 에이전트들 간의 자원 이용과정에서 발생하는 정보를 효율적으로 중재하고 조정하는 역할을 하는 기능이 존재해야 한다는 것이다. 공개 데이터베이스 자원들을 이용하는 경우 서로 다른 데이터 양식과 용어 체계, 제공하는 정보의 양과 정보를 제공하는데 걸리는 시간 등의 차이점이 존재하며, 또한 하나의 UI 에이전트로부터 복수 개의 질의를 수행하는 경우나 복수 개의 UI 에이전트들로부터 복수 개의 질의가 이루어지는 경우, 다양한 형태와 제공 정보로 인해 발생하는 자원들의 특성이 존재하게 된다. 이를 해결하기 위해서는 자원들의 다양성에 대한 고려, 메시지 트랜잭션의 관리와 더불어 다양한 플랫폼들에서 운용이 가능한 메시지 형식을 띠고 있어야 한다. 본 논문에서는 현재 개발중인 단백질 구조예측 연구를 지원하는 APSS 시스템 내에서의 사용자 에이전트와 여러 분석 및 자원 에이전트들 간의 효율적인 메시지 중재 및 조정기능을 갖는 CODY 에이전트를 설계하였고 이에 관하여 기술하고자 한다.

1. 서론

최근 들어 이미 개발되어 있는 330 여 종류 이상의 단백질 관련 정보를 제공하는 데이터베이스와 다수의 분석용 소프트웨어를 통합적으로 이용함으로써 좀더 빠르고 정확하면서도 다량의 결과를 도출하기 위한 단백질 관련 연구들이 진행되고 있다. 하지만, 이러한 연구에 있어 자원을 활용하기 위한 데이터베이스 간의 데이터 연결이나 통합, 양식, 용어체계 등의 차이로 인해 시간이 지날수록 점점 늘어나는 데이터간의 연결작업 또는 연동작업이 필요하다. 이러한 문제를 해결하고자 90 년대 말부터 에이전트 기술을 이용한 단백질 구조 예측 연구들이 이루어지고 있다.

인터넷 상의 자원들을 이용하는 에이전트 시스템들에는 요구되어지는 특징들이 몇 가지 있는데, 그 중 하나가 바로 사용자 에이전트와 공개 데이터베이스들을 대신하는 자원 에이전트들 간의 자원 이용과정에

서 발생하는 정보를 중재하고 조정하는 역할을 하는 기능이 존재해야 한다는 것이다. 이러한 기능은 단백질 자원을 다루는데 있어서, 데이터베이스 간의 양식이나 제공하는 정보, 용어 체계 등이 다르고, 또한 여러 자원들을 이용하는 구조예측 과정에 있어서 발생하는 문제를 해결하는데 중요한 역할을 한다.

본 논문에서는 현재 개발중인 단백질 구조예측 연구를 지원하는 APSS 시스템 내에서의 사용자 에이전트와 여러 분석 및 자원 에이전트들 간의 자원 이용과정에서 발생할 수 있는 용어 체계, 제공 정보의 양, 정보 제공에 걸리는 시간, 다수의 UI 에이전트로부터의 다수의 질의가 이루어지는 경우와 같은 문제를 해결하고자 효율적인 메시지 중재 및 조정기능을 갖는 CODY 에이전트를 설계하였고 이에 관하여 기술하고자 한다.

2. 관련 연구

2.1 에이전트 기반의 생물정보시스템

에이전트 및 다중 에이전트 시스템에 관한 연구는 소프트웨어 시스템 개발에 도움이 되는 직접적인 연구 성과를 거두는 단계이며, 에이전트 기법은 대규모 분산 소프트웨어 시스템 개발을 위한 차세대 개발 방법으로서 주목 받고 있다. 현재까지 알려진 대표적인 에이전트 기반의 생물정보 시스템들은 유전체 분석 및 단백질 구조 예측을 목적으로 고안된 시스템으로 동적이며 분산된 생물정보학 데이터 분석을 위해 MAS 개념을 도입한 GeneWeaver, 의학 분야에서 정보 검색을 위해 온톨로지 기반의 에이전트 기술을 이용한 시스템인 MeLiSA, 유전체 데이터 분석연구 작업을 수행하는 생물학자들을 위한 이동형 에이전트 시스템인 BioAgent, 예측 성능을 높이기 위해 에이전트 행위를 주로 다루며, 팀 단위의 협업 작업을 지원하는 X-MACoPS 등이 있으며, 이러한 시스템에서의 자원 이용과정에서 발생하는 정보에 대한 조정 기능을 보면 GeneWeaver의 경우 control module 이, MELiSA는 Query Generation 과 Query Evaluation 이 이러한 역할을 하게 된다.

2.2 APSS

APSS(Agent-based Proteomics Support System)는 단백질 구조예측을 지원하기 위한 시스템으로서, 다양하고 이질적인 에이전트들 간의 서비스를 지원할 수 있는 온라인 형태의 전세계적인, 협력작업을 지원하는 에이전트 시티 네트워크를 기반으로 하는 개방형 시스템으로, 인터넷상에 존재하는 다양한 단백질 관련 정보를 이용하여 이를 분석하고 활용하고자 하는 시스템이다[1].

APSS 시스템을 구성하는 에이전트들을 역할 및 수행 기능에 따라 분류해 보면 크게 자원 에이전트, 분석 에이전트, 관리 및 조정 에이전트, 제어 에이전트, 사용자 인터페이스 에이전트 그룹으로 나눌 수 있다.

- 자원 에이전트들 : PDB, GenPept, MEDLINE, 3D-PSSM, SWISSPROT, PFAM
- 분석 에이전트들 : BLAST, CLUSTALW, SWISS-MODELLER
- 관리 및 조정 에이전트들 : CODY, DM
- 제어 에이전트들 : AMS, DF
- 사용자 인터페이스 에이전트 : UI

시스템을 구성하는 에이전트들 중 CODY 에이전트는 본 시스템에서 가장 중요한 역할을 수행하는 에이전트로서 복수 개의 UI 에이전트로부터 받은 모든 질의에 대한 수행업무를 총괄하는 역할을 담당한다. 에이전트로서의 셋업 과정을 수행 후 메시지들을 수신 받게 되는데 이때 수신된 메시지들에 관한 분석과정을 거친 후 해당 질의에 대한 수행 가능한 서비스들이 존재하는 지를 DF 에이전트를 통해 확인한다. 그 후 AMS 에이전트로부터 해당 서비스를 수행하는 에

이전트들에 관한 정보를 구하게 되며, 해당 에이전트 리스트가 구해지면 수행해야 할 작업들을 미리 정의된 내용에 따라 할당하고 작업수행을 지시한다. 분석 및 자원 에이전트들로부터 받은 작업 수행 결과는 DM 에이전트를 통해 데이터베이스에 저장되고, 해당 UI 에이전트에게 전달된다. 현재까지는 개발 중이므로 인해 DM 에이전트를 통하지 않고 바로 DB 에 저장되는 형태로 이루어져 있다. 이와 같은 작업을 수행하기 위해 CODY 에이전트는 추출된 에이전트 리스트에서 우선순위 에이전트를 선택할 수 있도록 하기 위해 에이전트 평가값을 갖는다.

3. 조정 에이전트의 설계

PDB, SWISS-PROT, PIR 등과 같이 공개된 데이터베이스 자원들은 서로 다른 데이터 양식과 용어 체계, 제공하는 정보의 양과 정보를 제공하는데 걸리는 시간 등의 차이점이 존재한다. 또한 하나의 UI 에이전트로부터 복수 개의 질의를 수행하는 경우나 복수 개의 UI 에이전트들로부터 복수 개의 질의가 이루어지는 경우, 일관된 방식으로 구축되지 않고 다양한 기관으로부터 다양한 형태와 제공 정보를 가지고 개발된 자원들의 특성이 존재한다. 그러므로 이와 같은 문제들을 해결하기 위해서는 다음과 같은 메시지 중재 및 관리 방법이 필요하다.

1. 하나의 예측작업을 완성하기 위한 메시지 트랜잭션의 관리
2. 자원들의 다양성을 고려하는 메시지 컨텐츠 구조 개발
3. 다양한 플랫폼들에서 운용이 가능한 메시지 형식 고려

다음 절인 3.1 절에서는 예측작업을 위한 메시지 트랜잭션의 관리 부분을, 3.2 절에서는 자원들의 다양성을 고려하는 메시지 컨텐츠 구조 개발에 관한 부분을 기술하고자 한다. APSS 의 경우 다양한 플랫폼에서의 접속을 허용하기 위해 FIPA 에서 규정하고 있는 ACL 메시지 형식을 사용한다.

3.1 사용자 요구 정보 관리

CODY 에이전트는 정보를 관리하는 사용자 요구 정보 관리 테이블(URD, User's Requested Data management table)을 가지는데, 이러한 URD 는 여러 UI 에이전트와 자원 및 분석 에이전트들 간의 정보를 종합적으로 관리하기 위해 이용될 뿐만 아니라, 특정한 하나의 작업 처리 과정을 명세함으로써 단 단위의 작업수행 결과에 대해 총괄적으로 관리하기 위해 이용될 수 있다. URD 는 각 작업을 구분하는 Jkey, 정보를 요청한 에이전트의 AID 와 CID, Time, JID, Sub_JID 그리고 CODY 에이전트에서 생성한 AID 와 CID, 마지막으로 결과에 대한 RID 로 이루어져 있다. 각 항목에 대한 설명은 다음과 같다.

- Jkey : 각 작업에 대한 분류기
- AID : CODY 에이전트에게 정보를 요청하는

5. 결론

지금까지 본 논문에서 개발하고 있는 APSS 시스템에서의 다양한 자원 에이전트의 이용을 위한 CODY 에이전트의 기능과 메시지 콘텐츠 구조 및 응용에 대하여 기술하였다. 이를 통해서 APSS 는 생물정보 데이터베이스들 간의 이질성을 고려하면서도 재사용성을 높일 수 있게 되었다. 또한 다양하게 존재하는 단백질 자원들을 관리하고 사용자 측면의 편리성을 제공할 수 있으며, APSS 가 지향하는 시스템 측면에서의 개방성 및 분산성과 확장성을 높일 수 있게 되었다. 향후 계획으로는 자원 및 분석 에이전트의 기능 확장과 온톨로지를 이용한 자원 에이전트들로부터의 효과적인 정보 분석을 통해 시스템의 완성도를 높이고자 한다.

참고 문헌

- [1] 김현식 외, "단백질 구조예측을 위한 에이전트 시뮬레이션 기반의 다중 에이전트 시스템", 춘계 한국정보과학회 Proceeding. Vol.30, No.1, pp461-463, 2003
- [2] Andreas D. Baxevanis. "The Molecular Biology Database Collection: 2002 update", Nucleic Acids Research, Vol.30 No.1, 2002
- [3] Bryson, K. and Luck, M. and Joy, M. and Jones, D. "Agent Interaction for Bioinformatics Data Management", Applied Artificial Intelligence, Vol. 15, No.10, pp.917-947, 2001
- [4] Synthia Gibas, Per Jambeck., Developing Bioinformatics Computer Skills, O'Reilly, 2001
- [5] Willmott, S.N., Dale, J., Burg, B., Charlton, C. and O'brien, P., "Agentcities: A Worldwide Open Agent Network", Agentlink News, Vol.8, pp.13-15, Nov. 2001.