

## SeqWeB: SOA 기반의 서열 주해 시스템

남성혁<sup>○</sup>, 정태성, 김태경, 유재수, 조완섭  
충북대학교 충북BIT연구중심대학육성사업단

{daks\_gada<sup>○</sup>, mispro, tkkim, yjs, wscho}@chungbuk.ac.kr

### SeqWeB: Sequence Annotation System based on SOA

Seong-Hyeuk Nam<sup>○</sup>, Tae-Sung Jung, Tae-Kyung Kim, Jae-Soo Yoo, Wan-Sup Cho  
CBITRC, Chungbuk National University

#### 요 약

서열을 분석하고, 기능을 예측하는 서열 주해는 생명 현상 규명을 위한 필수 과정이다. 서열 주해는 다수 응용 프로그램간 상호 연계를 통한 복잡한 처리 과정을 거쳐 이루어진다. 현재 사용자는 다양한 응용 프로그램들 중 적합한 응용 프로그램을 선택한 후, 운영환경에 맞도록 설치하고, 사용법을 익혀야 한다. 또한 각 프로그램들의 연계를 위해 입출력 데이터 형식을 변환해야 하는 불편함이 있다. 이를 위해 자동화된 솔루션들이 개발되고 있지만, 각 단계별 프로그램들이 강결합(tightly coupled)되어 있어 유연성(flexibility)이 떨어지고, 기능의 확장 및 변경에 어려움이 있다.

본 논문에서는 기존 시스템들의 한계를 극복하기 위하여 SOA (Service Oriented Architecture) 기반의 서열 주해 시스템인 SeqWeB을 제안한다. SeqWeB은 서열 주해에 필요한 7개의 응용 프로그램(Phred, cross\_match, RepeatMasker, ICAtools, Phrap, CAP3, Blast)들을 웹 서비스 기술을 통해 단위 서비스로 개발하고, BPM 기법을 이용하여 통합하였다. SeqWeB은 각 응용 프로그램간 상호 운용성을 높이기 위하여 XML 형식의 입/출력 데이터를 사용하며, SOA 기반의 시스템 통합으로 각 응용 프로그램들을 약결합(loosely coupled)하여 시스템의 확장 및 변경이 용이하다. 또한 웹을 기반으로 하는 다양한 조합의 서열 주해 솔루션 제공이 가능한 특징이 있다.

#### 1. 서 론

포스트 지능 시대가 도래하면서, 전세계적으로 생명현상을 규명하고, 관련된 생물학적 기능들을 이해하기 위한 수많은 연구들이 진행되고 있다. 생물학적 기능에 관한 정보는 염기 또는 아미노산 단위의 배열인 서열 정보로 표현되는데, 이 서열 정보를 분석하여 기능을 예측하는 것이 서열 주해 (Sequence Annotation)이다.

서열 주해는 생명현상 규명을 위한 필수 단계로써, 보통 5~7단계를 거치며, 단계별로 다양한 응용 프로그램들이 개발되어 사용되고 있다. 각 단계별 대표적인 프로그램으로 Phred[11], cross\_match[11], RepeatMasker[12], ICAtools[13], Phrap[11]과 Blast[14]가 있다. 이러한 응용 프로그램들은 세계 각지의 연구소나 대학교에서 개발되어 인터넷을 통해 배포되고 있다.

하지만 사용자가 여러 응용 프로그램들을 이용하여 서열 주해 작업을 하기 위해서는 다음과 같은 작업이 수반되어야 한다. 첫째, 각 단계별 동일 기능을 수행하는 다양한 응용 프로그램들 중 연구목적에 맞는 응용 프로그램을 선택해야 한다. 둘째, 사용자는 여러 단계로 이루어지는 서열 주해 과정의 각 단계에 필요한 응용 프로그램들을 선택하여 설치하고, 그 사용법을 익혀야 한다. 특히, 현재 대부분의 응용 프로그램들은 유닉스나 리눅스 기반에서 운영되므로 생물학자들은 해당 시스템에 대한 지식을 익혀야 한다. 셋째,

프로그램마다 데이터 입출력 형식 (Data format)이 다르므로 사용자는 각 단계별 응용 프로그램에 적합한 입출력 형식을 맞추기 위해 추가 작업을 해야 한다. 이러한 불편함을 해소하기 위해서 여러 연구가 진행되었으며, GeneMaster[15]와 Genotator[5] 등이 그 예이다. 그러나 기존의 연구들은 확장 및 변경이 어려울 뿐만 아니라, 사용자는 강결합된 응용 프로그램만을 사용해야 하며 인터넷을 통한 서비스 제공이 불가능 하다.

따라서 보다 효율적이고 서비스 지향적이며 자동화된 서열 주해 서비스를 위해서는 각 단계별로 다양한 응용 프로그램들을 인터넷 환경에서 사용 가능한 통합 서비스로의 개선이 필요하다. 그러나 다양한 형태로 개발된 많은 응용 프로그램들을 웹 기반으로 통합하는 것은 다음의 이유들로 어려움을 겪고 있다[4, 6].

첫째, 응용 프로그램들의 구동 환경이 이질적이다. 모든 응용 프로그램들이 서로 다른 개발환경에서 개발되었으며, 이질적인 운영체제와 데이터베이스 환경에서 구동된다. 둘째, 각 응용 프로그램들은 서로 다른 입출력 형식을 사용하고 있다. 다시 말해서 프로그램 사이의 데이터 교환을 위한 표준 입출력 형식이 없으며 메시지 교환 방식도 이질적이다. 그러나 서열 주해를 웹 기반의 통합된 서비스로 제공하기 위해서는 다양한 응용 프로그램들을 표준화된 방식으로 서비스화하고, 표준 데이터 교환을 위한 입출력 형식과 메시지 교환 방식을 사용해야 한다.

본 논문에서는 웹 서비스 기술을 통한 서비스 지향 구조 (Service Oriented Architecture: SOA) 기법을 이용하여 서열 주해 작업의 각 응용 프로그램들을 단위 웹 서비스로 개발하고, 비즈니스 프로세스 관리 (Business Process Management: BPM) 기법을 이용하여 서열 주해 작업 프로세스를 설계하였다. 그 결과, SeqWeB을 이용하여 인터넷 환경에서 한번의 데이터 입력만으로 서열 주해 결과를 얻을 수 있으며, 다양한 조합의 솔루션을 제공할 수 있는 기반을 마련하였다. 또한 실제 실험에서는 기존의 서열 주해 방법과 비교하여 더 신속하고 편리하게 신뢰할 수 있는 서열 주해 결과를 얻을 수 있음을 증명하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 연구를 조사하고, 3장에서는 본 논문에서 제시하는 서열 주해 시스템인 SeqWeB을 설명한다. 4장에서는 SeqWeB의 성능을 검증하고, 5장에서 본 논문의 결론을 맺는다.

## 2. 관련연구

본 장에서는 일반적인 서열 주해 과정과 사용되는 응용 프로그램의 통합 기법에 대해서 살펴보고, 최근에 이슈가 되고 있는 SOA, 웹 서비스 그리고 BPM 에 대해서 살펴본다.

### 2.1 서열 주해 과정

서열 주해 작업은 목적에 따라 보통 5~7단계의 전처리 과정을 거친다.

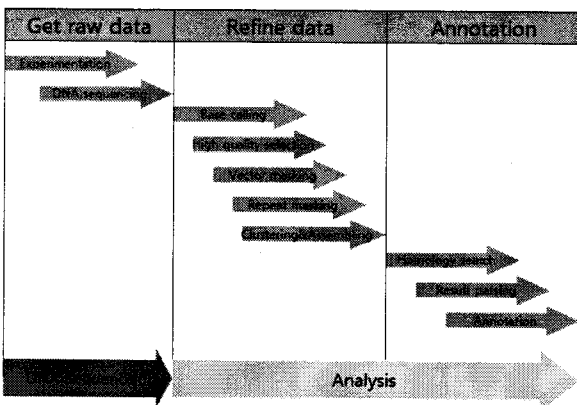


그림 1. 서열 주해 과정

그림 1은 서열 주해의 일반적인 과정을 보여준다. 생물학자가 실험을 통해 새로운 서열을 얻게 되면 Base calling 등을 비롯한 여러 단계의 정제 과정을 거쳐 새로운 서열의 기능을 예측한다. 이러한 서열 주해 과정은 5~7단계 별로 여러 가지 프로그램을 이용하여

처리할 수 있으며, 생물학자가 단계별 응용 프로그램들을 직접 사용하여 서열 주해를 수행하게 된다. 그러나 생물학자가 연구 목적에 적합한 응용 프로그램을 선택하기 어려우며, 선택한 응용 프로그램을 설치하고 그 사용법을 익혀야 하는 불편함이 있다. 또한 프로그램마다 다른 입출력 형식을 사용하므로 입출력 데이터에 대한 사용자의 추가 작업이 요구된다. 따라서 이러한 문제점을 해결하기 위하여 관련 연구 기관과 기업 등이 서열 주해 솔루션을 개발하고 있다.

### 2.2 서열 주해 솔루션

대표적인 서열 주해 솔루션으로서는 Genotator[5]와 GeneMaster가 있다. Genotator는 자동화된 서열 주해 작업과 브라우징을 위한 응용 프로그램으로 서열 주해의 2단계 전처리 과정과 Blast 처리 기능을 제공한다. Genotator는 비교적 단순한 서열 주해 기능을 제공하며 운영체제는 Unix OS만을 지원한다.

GeneMaster[15]는 국내에서 개발된 솔루션으로써 서열 분석 과정에 사용되는 응용 프로그램들 중 대표적인 프로그램들을 하나씩 선정하여 단순 통합/자동화한 프로그램이다. GeneMaster는 상용 프로그램이며, Genotator보다는 비교적 세분화된 서열 주해 과정을 제공한다.

그러나 Genotator와 GeneMaster는 모두 응용 프로그램 기반 솔루션으로 로컬 서버에 설치해야 하며 긴밀하게 연결된 프로그램들의 조합으로 구성되어 있어서 시스템 확장과 변경에 어려움이 있다. 즉, 사용 환경 변화에 따라 시스템 레벨의 수정 작업이 요구된다. 또한 서열 분석 과정의 각 단계마다 다양한 응용 프로그램이 존재함에도 불구하고 하나씩만을 선정하여 통합하였기 때문에 사용자는 Genotator와 GeneMaster에 사용된 프로그램만을 사용해야 하는 한계성이 있다.

### 2.3 서비스 지향 구조, 웹 서비스 그리고 BPM

기존 솔루션들이 가진 시스템 확장의 어려움과 사용의 한계성 등의 문제점을 극복하고 사용자의 요구에 유연하게 대응할 수 있는 솔루션이 요구된다. 이러한 솔루션은 웹 기반의 SOA, 웹 서비스, BPM 등과 같은 기술들을 이용하여 제공할 수 있다.

서비스 지향 구조 (Service Oriented Architecture: SOA)는 표준 인터페이스의 개념으로, 서비스를 기반으로 전체 어플리케이션을 구축하는 소프트웨어 아키텍처이다. SOA는 급격한 외부 환경의 변화에 어플리케이션을 민첩하게 조합할 수 있도록 지원하며, 재사용 (Reusability), 통합 (Integration), 민첩성 (Agility)을 지향한다. SOA 기반의 시스템에서는 하나의 서비스를 외부 시스템에서 호출하여 사용하는 것이

가능함으로써 시스템의 효율을 높이고 서비스들을 효과적으로 통합할 수 있다. 또한 사용 환경 변화에 따른 변경 요구사항이 발생했을 때 신속하게 대응할 수 있다. 이러한 SOA를 실현하기 위해서는 웹 서비스와 BPM과 같은 기술이 필요하다.

웹 서비스는 이질적 환경에서 사용되는 모든 응용 프로그램들이 직접 소통하고 실행될 수 있도록 동적 시스템 환경을 구현하여, 프로그램간 상호운용성을 제공하는 기술이다[16]. 바이오 관련 데이터베이스와 응용 프로그램들의 통합에 있어서 웹 서비스의 잠재력은 이미 많은 연구에서 검증되었다[1, 3, 6, 7, 9]. 웹 서비스 구현에는 SOAP, WSDL, UDDI 기술이 이용된다. SOAP(Simple Object Access Protocol)은 플랫폼, 개발언어 독립적인 기술로서 이기종 분산 환경에서 정보를 교환할 수 있는 XML 기반의 표준 프로토콜이다. WSDL(Web Service Definition Language)은 웹 서비스로 전환된 서비스의 사용 설명서를 기술한 XML 문서이다. UDDI(Universal Description, Discovery and Integration)는 웹 서비스 정보를 분류체계에 따라 저장할 수 있는 저장소로 서비스를 등록 검색할 수 있는 절차를 제공한다.

BPM은 사람과 어플리케이션의 상호 작용에 대한 지원과 함께 명확한 프로세스 관리 기능을 지원하는 서비스 및 도구이다[16]. SOA 환경에서 서비스들을 유연하게 관리하기 위한 기술로서, 프로세스를 관리한다는 것은 프로세스를 정의하고, 정의된 프로세스에 맞게 활동을 수행하며, 수행 결과를 측정하는 일련의 활동이다[8].

SOA나 BPM은 함께 사용하지 않아도 각각 강력한 효과를 나타내지만, 두 기법을 함께 사용함으로써 시스템은 훨씬 더 민첩하고, 유연하게 된다[16]. (Gartner Group. J.Sinur, D.McCoy, 2003).

본 논문에서는 웹 서비스 기법과 SOA, BPM을 함께 사용함으로써 변화에 민첩하게 대응할 수 있고, 확장성이 용이한 동적인 서열 주해 시스템을 제안하고자 한다.

### 3. SeqWeB

본 장에서는 서열 주해 과정을 일련의 프로세스로 정의하고 SeqWeB의 시스템 구조와 실제 SeqWeB을 이용한 서열 주해 각 단계를 설명한다.

#### 3.1 서열 주해 프로세스의 정의

본 연구는 서열 주해 작업의 각 프로세스에 초점을 맞추고, 각 단위 프로세스들을 하나의 통합된 프로세스로 생성하는 것이 목적이므로, 서열 주해 작업에 필요한 단위 프로세스들의 명확한 정의가 선행되어야 한다.

본 논문에서는 서열 주해 작업의 전 과정을 다음과 같이 총 6단계로 정의하였다.

- 제1단계: 염기 호출
- 제2단계: 벡터 서열 제거
- 제3단계: 반복 서열 제거
- 제4단계: 클러스터링
- 제5단계: 서열 조립
- 제6단계: 상동성 검색

제1단계는 생물학자가 얻어낸 크로마토그램을 읽어 들여 각각의 해당 염기를 결정해주는 염기 호출의 단계이다. 워싱턴 대학의 Phil Green과 Brent Ewing에 의해 개발된 Phred라는 응용 프로그램이 많이 사용되는데, Phred는 SCF (AVI model 373 및 377), ESD (MegaBACE) 등의 크로마토그램 파일을 자동으로 읽어 들여 각각의 서열 정보를 출력하는데 높은 성능을 보여준다[11].

제2단계는 1단계에서 읽혀진 염기 서열 중 cDNA 클로닝을 위해 사용한 벡터 서열이 전체 혹은 부분적으로 포함 될 수 있기 때문에 이를 제거해주는 단계이다. 이 단계를 거치지 않으면 4, 5단계의 결과가 다르게 나타나므로 필수 과정이라 할 수 있다. Vector masking은 cross\_match라는 응용 프로그램이 많이 사용된다[11].

제3단계는 반복 서열을 제거하는 과정이다. DNA 서열 내에는 생물 종마다 유전자의 발현과 무관한 것으로 보이는 서열이 단순 반복되는 것으로 밝혀졌다. 이러한 반복 서열은 매우 불규칙적으로 DNA 서열 내에 광범위하게 존재하기 때문에 제거해야 하며, RepeatMasker라는 응용 프로그램이 많이 사용된다[12].

제4단계는 확실하지 않은 대량의 DNA 서열 조각들을 클러스터링하는 단계이다. 이 단계를 거치면서 대표 서열을 얻게 되고, 그 대표 서열을 통해 서열 전체의 기능을 예측하게 된다. 이 단계에는 ICATools라는 관련 프로그램 패키지를 많이 사용한다[13].

제5단계에서는 하나의 서열로 조립(Assembling)하는 과정이다. 4단계의 전처리 과정을 거치게 되면 몇몇 조각의 대표 서열을 얻게 되는데 이러한 대표 서열들을 조립하여 하나의 서열로 만들어야 한다. 이 단계에서는 CAP3와 Phrap 등의 응용 프로그램들이 사용된다[11].

5단계의 전처리 과정을 모두 마친 하나의 서열을 이용하여 제6단계에서는 BLAST[14]를 이용한 상동성 검사를 통해 서열의 기능을 예측할 수 있게 된다.

#### 3.2 시스템 구조

본 논문에서 제안하는 SeqWeB의 구조는 그림2와 같다. SeqWeB의 구조는 네 계층으로 구성된다. Legacy application 계층은 서열 주해에 사용되는 기존의 응용 프로그램들로서, 서열 분석의 각 단계별 다양한 응용 프로그램들을 포함한다. Web service interface 계층은 다양한 프로그래밍 언어로 개발된 이질적 구동환경을 갖는 응용 프로그램들이 시스템과 통신하기 위한

인터페이스이다. Service Orchestration 계층은 사용자가 자신의 목적에 적합한 서열 분석 프로세스를 설계하는 부분이다. Presentation 계층은 사용자의 데이터 입출력 및 프로세스 설계를 위한 GUI이다.

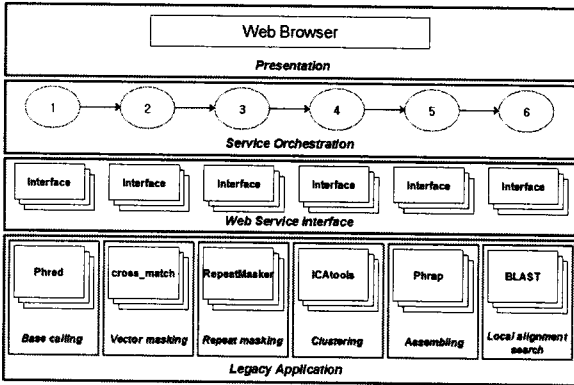


그림 2. SeqWeB의 구조

### 3.1.1 Legacy application

Legacy application은 서열 주해 작업을 실질적으로 수행하는 해당 응용 프로그램들로 구성된다. 본 논문에서는 표 1과 같이 서열 주해 작업의 프로세스를 총 6단계로 정의하고, SeqWeB의 프로토 타입을 위한 대표 프로그램들을 선정하였다.

단계	기능	대표 프로그램	개발 언어
1단계	염기 호출	Phred	C
2단계	벡터 서열 제거	cross_match	C
3단계	반복 서열 제거	RepeatMasker	C
4단계	클러스터링	ICAtools	C
5단계	서열 조립	Phrap, CAP3	C
6단계	상동성 검색	BLAST	C

표 1. 서열 주해 프로세스 및 대표 프로그램과 개발 언어

표 1의 대표 프로그램들은 모두 Linux OS 환경의 SeqWeB 서버에 설치되었다.

### 3.1.2 Web service interface

SeqWeB은 웹 서비스의 가장 큰 장점 중 하나인 응용 프로그램간 상호 운용성 보장을 극대화하기 위해 플랫폼 독립성을 보장하는 java로 구현되었다. 하지만 각각의 응용 프로그램들은 대부분 C 언어로 구현되었고, 처리 과정에서 Perl 등의 언어도 사용되기 때문에 java로 구현된 시스템이 각 응용 프로그램을 호출하고 결과값을 반환 받는데 어려움이 있다. 그래서 SeqWeB은 모든 응용 프로그램들과 Service

orchestration 계층 사이에 Interface를 위치시켜 시스템과 응용 프로그램 사이에 원활한 호출과 결과값 반환이 이루어지도록 설계되어 있다. 한 응용 프로그램과 하나의 Interface는 단위 웹 서비스를 위한 하나의 세트이다. 그림 1의 Legacy application에서 Phred가 아닌 제1단계를 수행하는 다른 응용 프로그램이 추가 될 경우 추가의 Interface를 생성하여 Service orchestration에 연결할 수 있게 된다.

본 논문의 SeqWeB 프로토 타입은 각각의 대표 프로그램과 해당 인터페이스를 생성하여 7개의 단위 웹 서비스를 제작/배포하였다.

### 3.1.3 Service orchestration

Legacy application과 Web service interface에서 단위 웹 서비스로 제작/배포된 각각의 응용 프로그램들은 Service orchestration에서 약결합 (loosely coupled) 상태로 연결된다. 약결합 된 상태는 BPEL을 사용함으로써, 사용자가 단위 웹 서비스들을 자신이 원하는 프로세스 로직으로 설계하도록 만들어준다.

SeqWeB의 Service orchestration은 SOA를 이용한 약결합은 물론, BPEL을 이용한 사용자 정의 프로세스 로직을 구현 가능하게 하는 가장 중요한 부분이다. 사용자는 Presentation의 GUI를 통해 수월하게 프로세스 로직을 설계할 수 있다(그림 2).

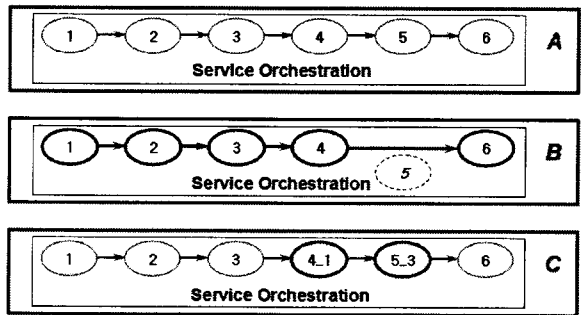


그림 3. 사용자 정의 프로세스 설계

그림 3의 A는 그림 1의 Service orchestration 부분과 동일한 그림으로, SeqWeB에서 정의한 서열 주해 과정 6단계 모두를 순서대로 설계한 그림이며, SeqWeB의 기본(default) 프로세스 로직이다.

B는 A의 프로세스 로직에서 제5단계를 제거하고 제4단계를 제6단계로 바로 연결시켜 설계한 프로세스 로직이다. 사용자는 SeqWeB에서 기본적으로 정의한 서열 주해 작업 프로세스에서 자신이 원하지 않는 불필요한 단계를 제거하거나 추가함으로써, 자신에게 최적화 된 효율적인 서열 분석 프로세스를 생성할 수 있다.

C는 동일 기능을 수행하는 다른 서비스를 연결하여 프로세스 로직을 설계한 것이다. 4\_1은 4와 동일한 기능을 수행하는 다른 종류의 서비스이며, 5\_3도 5와 동일 기능을 수행하는 서비스들 중 하나이다. 사용자는 표 1에서 제시한 대표 프로그램 외에도 자신이 선호하는 응용 프로그램을 선택하여 자신만의 프로세스 로직을 설계할 수 있다.

BPEL을 이용한 프로세스 정의는 Oracle BPEL Manager를 이용하였으며, IDE tool은 JDeveloper를 사용하였다.

SOA를 통한 약결합이기에 가능한 작업이며, 이러한 사용자 정의 프로세스 설계는 BPEL을 통해, 사용자는 시스템의 구조에 대한 이해나 시스템 일부 코드를 변경하는 작업 없이 GUI를 통해 손쉽게 자신만의 프로세스를 설계할 수 있는 것이다.

### 3.2 SeqWeb을 이용한 서열 주해

그림 4(a)는 염기 호출 단계에서 서열 주해를 원하는 크로마토그램 파일들을 파일 업로드 기능을 이용하여 업로드 할 수 있음을 보여준다. 또한 염기 호출에 필요한 각종 파라미터에 대하여 사용자가 직접 값을 입력할 수 있다.

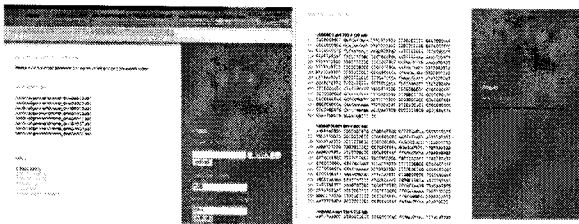


그림 4(a)

그림 4(b)

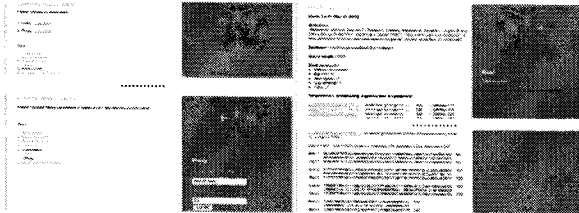


그림 4(c)

그림 4(d)

그림 4(a)와 같이 크로마토그램 파일을 업로드하고, 적절한 옵션 값을 설정한 후 실행하면, 그림 4(b)와 같이 염기 호출이 완료된 결과값을 확인할 수 있다. 그림 4(b)의 결과는 텍스트 형태의 파일로 제공될 수 있으며, 다음 단계로의 진행을 원할 경우, 파일 포맷 변경 등을 비롯한 추가 작업 없이 다음 단계로 진행할 수 있다.

SeqWeb은 서열 조립 과정에서 동일 기능을 하는 두 개의 응용 프로그램을 서비스함으로써, 사용자에게 의한 프로세스 설계를 가능하게 한다. 그림 4(c)는 서열 조립 과정에서 사용자가 CAP3와 Phrap 중 자신이 선호하는 응용 프로그램을 선택할 수 있음을 보여준다. 각 응용 프로그램은 단위 웹 서비스로 배포되어, BPEL을 통해 약결합되므로 사용자는 시스템에 대한 지식이나 별도의 작업 없이 자신이 원하는 응용 프로그램을 이용하여 자신에게 최적화된 서열 주해 작업을 수행할 수 있다.

서열 주해의 마지막 단계인 상동성 검사 결과는 그림 4(d)와 같이 나타난다. 상동성 검사는 NCBI에서 제공하는 BLAST의 결과와 동일한 형식을 포함하여 XML과 테이블 형식 등 다양한 형식을 제공한다.

## 4. SeqWeb의 검증 및 주요 특징

### 4.1 SeqWeb의 검증

표 2는 SeqWeb의 검증에 적용된 실험 환경과 검증 결과이다. 검증은 기존의 각 프로그램들을 이용하여 수작업 방식과 각 프로그램들이 서비스 레벨에서 통합된 SeqWeb을 이용한 방식을 비교를 통하여 이루어졌다.

	기존 방법	SeqWeb
검증방식	수작업에 의한 처리	자동화된 프로세스에 의한 처리
운영환경	Linux	Linux, Tomcat 5.5
사용환경	7개의 단위 응용 프로그램	통합된 Web application
입력 데이터	100개의 <i>Oryza sativa</i> (japonica cultivar-group) 뉴클레오타이드 크로마토그램 파일	
결과	입력 데이터 대한 서열 주해 결과 일치 확인	

표 2. 실험 환경 및 검증 결과

검증 결과를 통해 SeqWeb은 기존의 서열 주해 방법과 비교하여 단 한번의 서열 데이터 입력만으로 쉽고 빠르게 신뢰할 수 있는 서열 주해 결과를 얻을 수 있음을 증명하였다.

### 4.2 주요 특징 및 기대 효과

#### 4.2.1 통합된 서열 주해 웹 어플리케이션

SeqWeb은 SOA 기반의 웹 서비스 기술과 BPM 기법을 적용하여 웹 기반의 서열 주해 서비스를 제공한다. 사용자는 인터넷 환경하에서 손쉽게 최적화된 서열 주해 프로세스를 생성하여 그 기능을 예측할 수 있다. 또한 기존 응용 프로그램 기반의 서열 주해 솔루션에서 발생하는 응용 프로그램 업데이트 등의 불편을 해소한다.

#### 4.2.2 SOA 기반의 응용 프로그램간 느슨한 연결

SeqWeB은 SOA 기반의 웹 서비스 기술을 사용함으로써, 서열 주해 프로세스에 포함된 응용 프로그램들을 단위 웹 서비스로 개발하고, 모든 서비스들을 느슨하게 연결(loosely coupled)하였다. SeqWeB의 각 서비스간 느슨한 연결은 시스템의 확장 및 변경을 용이하여 사용자의 환경 변화에 유연하게 대응할 수 있는 서열 주해 솔루션을 제공한다.

#### 4.2.3 BPM 기법을 통한 사용자 정의 프로세스

SeqWeB은 서비스간 느슨한 연결에 BPM 기법을 적용하여 사용자에게 의해 프로세스 생성을 가능하게 하였다. 사용자는 SeqWeB이 제공하는 응용 프로그램들 중 자신에게 적합한 서비스들만을 선택하고 그 서비스들을 효율적인 프로세스로 정의함으로써 자신에게 최적화된 서열 주해 작업을 수행할 수 있다.

### 5. 결론 및 향후 연구

서열 주해 작업은 생명 현상 규명을 위해 필수적으로 거쳐야만 하는 과정으로, 많은 연구자들이 각종 응용 프로그램을 이용하여 정확도 높은 서열 주해 작업을 하기 위해 노력하고 있다. 하지만 여러 단계의 서열 주해 작업은 몇몇 어려움으로 인해 현재까지 하나의 통합/자동화된 프로세스로 웹 서비스에 의해 제공되고 있지 않기 때문에 연구자들은 사용의 어려움은 물론 서열 주해의 정확도 저하를 우려하고 있다.

웹 서비스와 BPM을 이용한 서열 주해 시스템인 SeqWeB은 인터넷을 통해, 여러 단계의 서열 주해 작업을 하나의 통합/자동화된 프로세스로 제공한다. 사용자는 각 응용 프로그램의 설치/사용법을 알지 않아도 인터넷이 연결된 곳이라면 어디서든 손쉽게 한번의 크로마토그램 입력만으로 서열 예측 결과를 확인할 수 있다. 또한 자신에게 최적화된 서열 분석 프로세스를 직접 설계함으로써 기존의 방법보다 훨씬 효율적인 서열 주해 작업을 할 수 있다.

본 연구는 SeqWeB을 사용함으로써, 기존의 방법보다 더 쉽고, 정확한 서열 주해는 물론, 신뢰할만한 서열 주해가 가능함을 검증하였다.

향후 더 다양한 응용 프로그램들의 사용이 가능하도록 정확도 높은 응용 프로그램들을 선정하여 Legacy application으로 추가할 예정이다. 또한 Ajax와 같은 Web2.0의 다양한 기술들을 적극적으로 활용하여, 사용자에게 직관적이고 사용하기 쉬운 서비스 제공이 가능하도록 발전시킬 계획이다.

#### 참 고 문 헌

- [1] Bioinformatics workflow builder interface (biowbi). <http://www.alphaworks.ibm.com/tech/biowbi>.
- [2] J. J. Garrett. AJAX: A new approach to web applications. <http://www.adaptivepath.com/publications/essays/archives/000385.php>, Feb. 18, 2005.
- [3] Ken Vollmer, et al. Integration In A Service-Oriented World. *Forrester Research*, 2004.
- [4] N. Milanovic and M. Malek. Current solutions for web service composition. *IEEE Intelligent Systems*, pages 46-53, March/April 2001.
- [5] Nomi L. Harris. Genotator: A Workbench for Sequence Annotation. *Genome Research*, July 1997.
- [6] Remko de Knikker., et al. A web services choreography scenario for interoperating bioinformatics applications. *BMC Bioinformatics*, May 2004.
- [7] S. Carrere and J. Gouzy. Remora: a pilot in the ocean of biomoby web-services. *Bioinformatics*, 22(7), 2006.
- [8] Wil M. P., et al. Business process management: Where business processes and web services meet. *Data & Knowledge Engineering*, Volume 61, Issue 1, April 2007.
- [9] Xiaorong Xiang. and Gregory Madey. A Service-oriented Data Integration and Analysis Environment for In Silico Experiments and Bioinformatics Research. *Proceedings of the 40<sup>th</sup> Annual Hawaii International Conference on System Sciences*, 2007.
- [10] <http://www.w3.org/2002/ws/Activity>
- [11] <http://www.phrap.org/>
- [12] <http://www.repeatmasker.org/>
- [13] <http://www.littlest.co.uk/software/bioinf/>
- [14] <http://www.ncbi.nlm.nih.gov/BLAST/>
- [15] <http://www.ensoltek.com/Products/Products.htm>
- [16] <http://www.gartner.com/>