

MPEG IoMT 에서의 자연어 인터페이스 표준화

최미란
한국전자통신연구원
miranc@etri.re.kr

Natural Language Interface for MPEG IoMT

Miran Choi
Electronics and Telecommunication Research Institute

요 약

본 논문에서는 최근 인공지능 기반의 자연어이해기술을 활용한 자연어 인터페이스 표준화 현황을 소개하고 사물기반의 미디어 사물간의 기능들을 표준화하고 있는 MPEG IoMT 표준에서의 자연어 인터페이스 구현 내용을 소개한다. 자연어 인터페이스에는 음성인식 기술, 음성합성 기술, 언어처리 기술, 질의응답기술, 음성 자동통역 기술등이 포함되며 언어지능으로서의 자연어 인터페이스를 사물 인터넷 환경에서 구현하기 위해 MPEG IoMT 의 표준화된 포맷과 활용 방식을 소개한다.

1. 서론

최근 MPEG 에서는 IoT(Internet of Thing) 및 웨어러블 환경에서의 효율적인 미디어 소비를 위한 표준으로 IoMT(Internet of Media-Things)의 표준화를 진행하고 있다. IoMT 에서는 센서, 액추에이터(actuator) 및 미디어 분석기(media analyzer) 간의 API(Application Programming Interface) 및 인터페이스 데이터를 서술하기 위한 메타데이터 등에 대한 표준을 포함하고 있다

근래에 인간의 이해를 바탕으로 인간 친화적 서비스를 제공하기 위한 지능형 시스템의 발전이 꾸준히 이루어지고 있다. 이러한 인공지능 시스템의 주요 기술로 사용자가 손쉽게 원하는 지식을 대화하면서 서비스 받거나 언어 제한 없이 다양한 정보를 획득하고 의사 소통할 수 있는 언어지능 기술을 들 수 있다. 또한 언어 지능을 위한 핵심 기술로서 질의응답 기술과 자동통역 기술이 상용화된다면 다양한 언어로 만들어진 많은 정보를 한국어로 검색하여 원하는 정보를 획득할 수 있고 다른 언어를 사용하는 사람들과도 자유롭게 소통할 수 있게 된다. 원활한 의사소통과 지식 획득을 위한 언어지능기술의 실현으로 인간 중심의 편리한 세상은 더욱 앞당겨 질 것으로 기대된다.

본고에서는 이러한 자연어 기반의 언어지능기술의 핵심 기술인 음성 인터페이스 기술, 자동통역 기술, 그리고 자연어 질의응답 기술을 MPEG IoMT 표준으로 표현하고 활용하는 방식을 소개하고 관련된 주요 표준화 동향을 기술한다.

본 논문의 구성은 다음과 같다. 2 자연어 인터페이스를 위한 각종 기술과 표준화 현황에 대해 살펴본 후, 3 절에서는 이러한 자연어 인터페이스를 사물 인터넷 환경에서 처리하고 활용하는 MPEG-IoMT 표준 방식을 소개한다. 마지막으로

6 절에서는 본 논문에 대한 결론을 맺는다.

2. 자연어 인터페이스 표준화

최근 인공지능 기반의 자연어이해기술의 현황은 머신러닝/딥러닝을 기반의 언어 처리의 어휘와 문장의 문법분석기술, 빅데이터를 활용한 자연어 질의응답 기술이 상업적 성공가능성이 높아 관련 연구소와 기업에서 활발히 연구되고 있다. 국내의 현황은 기업을 중심으로 기초적인 지능정보 응용기술을 개발하여 자사 서비스에 적용하고 있으며 채팅 로봇 등을 활용하여 사용자가 말을 걸었을 때 사람이 대꾸하는 것처럼 대화 형태로 답변하는 서비스를 개발하는 등 다양한 언어지능 기반의 시스템과 서비스가 개발 되고 있다.

자연어 질의응답 기술은 사용자의 자연어 질문에 맞는 정답 후보들을 추출하여 그 중에 최적 정답을 제시하는 기술로 다양한 인공지능 응용 시스템에서 언어지능 핵심 모듈로 활용이 가능한 기술이다.

자연어 질의응답 기술 관련 표준화는 국내에서는 TTA 메타데이터 표준분과에서 자연어처리 표준화 활동이 진행되어 한국어 형태소 표준, 개체명 태깅과 구문분석 표준, 다의어 태깅 표준, 질의분석 표준이 개발 완료되었으며 의미분석 관련 표준화는 ISO 의 의미분석 국제표준을 기반으로 한국어에 적용, 확장한 의미역 태그셋 표준을 국가표준으로 제안하여 표준화 진행 중이다.

국제 표준화는 ITU-T SG16 의 멀티미디어 서비스 분과에서 한국에서 발의한 “지능형 질의응답 서비스 프레임워크” 표준화가 진행되어 F.746.3(2015)과 F.746.7(2018) 이 표준 승인되었다. 후속 표준으로 지능형

질의응답 서비스를 위한 인터페이스 표준화가 진행 중이다.

음성인터페이스 기술은 인간의 자연스러운 통신 수단인 음성을 인식하여 기계와의 인터페이스를 가능하게 해 주는 기술이며 사람의 말을 이해하고 대화를 통해 사용자의 의도와 상황에 맞는 서비스 및 응답을 제공하는 기술인 대화처리 기술과 같이 사용되어 자연스러운 사용자 인터페이스 (NUI: Natural User Interface)의 근간이 되는 기술이다. 이를 위한 주요 기술로 자연어 음성인터페이스 기술과 자연어 대화처리 기술이 있다

음성인터페이스 관련 국제 표준화 활동은 크게 W3C, ITU-T, ISO 등에서 수행되고 있다. ITU-T 에서는 “IPTV 를 위한 고기능 사용자 인터페이스” 표준은 2016 년에 개발 완료 되었다. ISO JTC SC35 사용자 인터페이스 표준화 그룹에서 음성명령어 표준이 완료되었으며 향후 자연어 음성인터페이스와 대화형 음성인터페이스와 같은 고기능의 음성인터페이스 표준이 추진될 예정이다. ITU-T SG16 멀티미디어 서비스 그룹에서는 “음성/자연어처리 기술 기반의 언어 이터닝 서비스 프레임워크” 표준이 F.746.5 로 승인되었다. 또한 음성인식이 주요 기술로 들어가는 “실내 대화로봇 (H.ICR)” 표준이 ITU-T SG16 멀티미디어 서비스 그룹에서 2019 년에 F.746.9 로 개발 완료되었다. 또한 음성인터페이스의 주요 기술인 대화처리 기술에 대한 표준화가 한국의 주도로 진행되고 있다.

자동통역 기술은 크게 다국어 자동번역 기술과 다국어 음성 자동통역기술로 구성되며 다국어 자동번역 기술은 다국어 문서 또는 대화를 자국의 언어로 자동으로 번역하는 기술로써 입력 텍스트 및 대화체 문장에 대한 언어분석, 변환/생성 기술이 기본적으로 필요하며, 번역 성능 향상 및 언어 확장성을 위해 지식 추출/학습 기술과 하이브리드 자동번역 기술이 필요하다. 이를 위한 주요 기술로 다국어 언어분석 기술, 다국어 생성 및 변환 기술, 지식추출 및 학습 기술, 하이브리드 자동번역 기술 등이 주요 구성 요소이다.

자동통역시스템은 주로 여행/관광 분야 의사소통이 1 차 목표이며, 일상, 비즈니스 등 다양한 영역으로 확대될 전망이며 웨어러블 컴퓨터의 발달에 따라 스마트폰이 아닌 착용형 안경, 착용형 시계에 통역기가 장착되어 사용성의 편리함이 극대화 될 것으로 예상된다.

국제 표준화 현황은 음성통역 표준화가 ITU-T 와 ISO 에서 진행 중인데 ITU-T SG16 의 인적요소 그룹인 Q24 에서는 편의성과 접근성을 강조하는 대면상황의 음성통역 서비스의 인적요소를 고려한 음성통역서비스 평가에 관한 표준화가 한국의 제안으로 시작되었고 표준화가 진행되고 있다. 또한 ISO JTC1 SC35 에서 음성 자동통역 서비스의 사용자 인터페이스 표준화가 진행되어 2017 년 10 월에 표준 발간되었고 이 표준의 후속으로 같은 그룹에서 그림 1 과 같은 동시통역 시스템 표준화가 2018 년에 시작되어 신규 표준화 항목으로 승인되었으며 향후 3~4 년에 걸쳐 국제표준으로 개발될 예정이다.

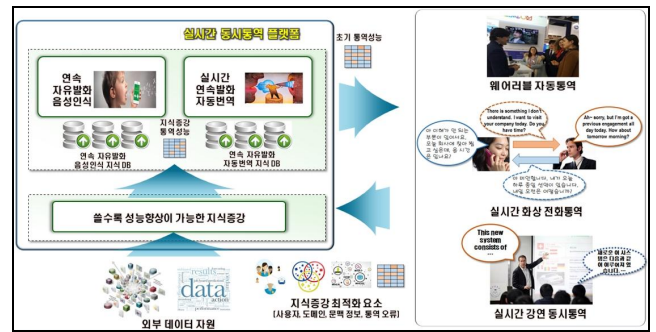


그림 1. 실시간 동시통역시스템

본 절에서는 인공지능이 중심이 되는 지식산업혁명 시대에 편리하게 필요한 지식을 찾고 이용할 수 있는 자연어처리 기반의 응용 기술인 자연어 질의응답 기술, 음성 인터페이스 기술, 자동통역기술에 대하여 소개하고 기술 현황과 표준화 현황에 대하여 기술하였다.

3. MPEG-IoMT 의 자연어 인터페이스 기능

MPEG 에서 개발되고 있는 사물 인터넷 환경에서의 효율적인 미디어 소비를 위한 IoMT(Internet of Media-Things) 표준에서는 사물간의 음성인터페이스의 기능인 음성인식과 음성합성 호출 시나리오를 그림 2 와 같이 제시하고 있다.

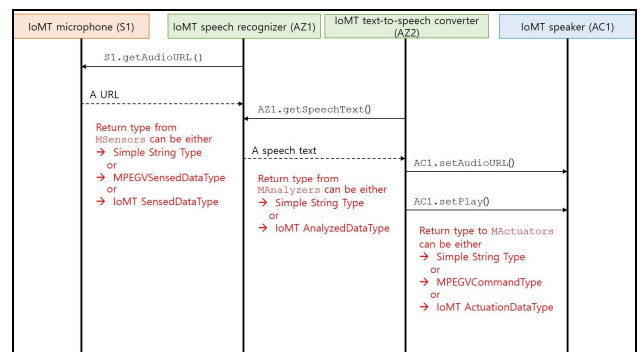


그림 2. MPEG-IoMT 에서의 음성인터페이스

종래의 질의응답 기술은 질문자가 직접 입력한 질문 문장에만 의존하여 정답을 찾기 때문에 다양한 사용자의 요구에 대해 해결이 어려웠다. 최근에 Wearable Device 를 비롯한 IoT 기기들이 많이 등장을 하는데 단순한 질문만 해결하는 질의응답시스템은 한계가 있다. 이에 대한 불편함을 해소하기 위해 질문자의 발화에 대해 기기에서 미리 발화를 분석하여 질문자의 의도를 파악하는 것이 필요하다. MPEG 에서는 이러한 IoT 환경에서의 멀티미디어 기술을 구현하기 위하여 MPEG IoMT 그룹에서 표준을 만들고 있으며 여기에 질의응답 UI 를 포함하고 있다. 이를 위하여 사용자의 발화 내용을 분석하여 적절한 IoT 기기에서 처리할 수 있게 하는 기술을 반영하여 적용하였으며 IoMT 환경에서의 질의응답 UI 기술을 제안하였다.

기존의 질의응답시스템은 사용자가 질문을 하면 음성이나 텍스트의 형태로 질의응답 시스템이 있는 서버로 전달되어 한정된 질문에 대해서 응답을 제공하는 방식이었으나 사물 인터넷 환경에서는 다양한 디바이스를 통하여 질문과 명령 등의 다양한 발화에 대한 처리가 가능하도록 하는 기술이

필요하다. 또한 자연어에만 한정되지 않은 미디어의 입력 방식의 한계를 극복하여 발화를 처리하기 위하여, 사용자가 제시하는 사진 등의 미디어를 자연어와 통합시켜서 질의응답시스템에서 답을 구하는 기능이 가능하게 하며 이를 통해 정확한 정답을 찾을 수 있도록 하는 질의응답 장치 및 방법이 연구되고 있다.

아래의 예제는 IoMT 표준에서 질의응답을 위한 인터페이스를 구현하기 위하여 미디어 타입을 IoMT 질문 분석기 형태로 구성하는 방법을 보여준다. 음성통역을 비롯한 다른 자연어 인터페이스 구성요소 기술도 아래의 질의응답 예시와 비슷한 방식으로 구현할 수 있으며 현재 IoMT 유즈케이스에 포함되어 있으므로 구체적인 내용은 향후 추가 IoMT 표준에 포함될 예정이다.

IoMT 질문 분석기

IoMT 질문분석기에 의하여 생성되는 결과를 표현하기 위하여 아래와 같은 데이터 포맷이 표준으로 정의된다. 질문분석기는 사용자의 텍스트 질문을 입력으로 받아서 질문분석 결과를 생성한다. 질문 분석 결과는 사용자의 사물과의 상호작용을 위한 명령어를 발생시키기 위한 입력으로 사용될 수 있으며 또는 사용자에게 답변이나 정보를 제공하기 위해 질의응답 (QA) 서버로 보내는 입력으로 사용될 수 있다. 아래에 정의된 사용자 질문 타입은 분석된 사용자 질문을 표현하기 위해 사용되며 이 질문에 대한 응답은 질의응답 서비스에 의하여 제공될 수 있다. 만일 사용자 질문이 제어 명령으로 분석되면 그 명령은 액추에이터(actuator)에게 보내져서 실행된다. 아래의 질문 분석 타입 정의와 질문 분석 타입의 의미 내용은 MPEG IoMT DIS 문서에서 발췌하였다.[1]

- 질문 분석 타입 정의

질문 분석 타입은 분석된 질문 요소와 언어 요소로 구성되어 있으며 분석된 질문 요소는 사용자 질문 타입으로 되어 있다. 사용자 질문 타입은 토픽, 초점, 의미, 그리고 도메인 요소로 구성되어 있다.

```
<!-- ***** Definition of Question Analysis Type ***** -->
<!-- Definition of Question Analysis Type -->
<!-- ***** -->
<complexType name="QuestionAnalysisType">
  <complexContent>
    <extension base="mtdl:AnalyzedDataBaseType">
      <sequence>
        <element name="analyzedQuestion"
          type="maov:UserQuestionType" minOccurs="0"/>
        <element name="language" type="language" minOccurs="0"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>
.
<complexType name="UserQuestionType">
  <sequence>
    <element name="qtopic" type="string" minOccurs="0"/>
    <element name="qfocus" type="mpeg7:termReferenceType" minOccurs="0"/>
    <element name="qCsemantic" type="mpeg7:termReferenceType" minOccurs="0"/>
  </sequence>
  <attribute name="qdomain" type="string"/>
</complexType>
```

- 질문 분석 타입의 의미 내용

IoMT 환경에서 사물 간에 주고 받는 미디어를 분석하여 제공하는 미디어분석기는 각각의 개별 분석기(M-Analyzer)로 구성되며 그 중의 하나가 질문 분석기이다. 질문을 분석하여 결과를 산출하는 질문 분석기는 질문 분석 타입으로 되어 있으며 이는 분석된 질문 요소와 언어 요소로 구성되어 있다. 분석된 질문 요소는 질문 분석의 결과이며 사용자 질문 타입으로 되어 있다. 사용자 질문 타입 사용자 질문을 추상화하는 정보를 제공한다. 사용자 질문은 QA 서버로 보내져서 응답을 제공할 수 있다. 제어명령으로 분석된 사용자

질문은 액추에이터로 보내져서 해당 명령을 수행한다. 질문 토픽은 그 질문의 대상이 되는 사물이나 이벤트이다. 질문 초점은 5W1H 로 분류되는 질문 형식 중의 하나를 의미한다. 질문의 의미는 질문의 의미나 목적에 기반하여 분류된 질문으로 정보 요청, 명령 요청, 용어 요청 등이 이에 해당된다. 질문 도메인 요소는 “과학”, “기후”, “역사”와 같은 질문의 분야를 지칭한다.

Name	Definition
QuestionAnalysisType	Provides an abstract of description of question analysis, which can be done in a processing unit.
analyzedQuestion	Describes analyzed question resulting from the question analysis.
language	Indicates the language of the input question. If present, the Language element should take precedence over other language indications present within the input question.
UserQuestionType	Provides abstracts of User Question description. Describes user's utterance that is the output of speech recognition process. User Question is sent to QA server for providing answers to the user. If it is a control command, it is sent to the actuator.
qtopic	Describes topic of the question. Question topic is the object or event that the question is about. Ex. Qtopic is King Lear in "Who is the author of King Lear?".
qfocus	Describes the focus of the question, which is one of 5W1H. The type of the focus shall be described using the mpeg7:termReferenceType defined in ISO/IEC 15938-5:2003, 7.6. A classification scheme that may be used for this purpose is the QfocusCS defined in A.4.4. Ex. What, where, who, what policy.
qCsemantic	Describes the question classification based on the meaning/purpose of the question. The type of the question classification shall be described using the mpeg7:termReferenceType defined in ISO/IEC 15938-5:2003, 7.6. A classification scheme that may be used for this purpose is the QCsemanticCS defined in A.4.5. Ex. What does MPEG stand for? (Request for terminology). Could you please turn on the TV? (Request for command)
qdomain	Describes the domain of the question such as "science", "weather", "history". Ex. Who is the third king of Yi dynasty in Korea? (qdomain: history)

- 질의 분석 타입의 사용 예

아래 예제는 “리어왕의 저자는 누구입니까?” 라는 사용자의 질문에 대한 분석 결과를 보여준다. 예제의 질문 분석 결과는 분석된 질문과 (analyzedQuestion) 언어 종류가 미국식 영어라는 것을 포함한다. 예제의 질문 분석 결과는 질문의 분야가 “문학”이고 질문의 주제는 “리어왕”이며 질문의 초점은 “WHO_QUESTION”(사람에 대한 질문) 그리고 이 질문의 목적은 “REQUEST_FOR_INFORMATION”(정보 요청)이라는 것을 보여준다.

```
<mtdl:analyzedData xsi:type="maov:QuestionAnalysisType">
  <maov:analyzedQuestion qdomain="Literature">
    <maov:qtopic>King Lear</maov:qtopic>
    <maov:qfocus>WHO_QUESTION</maov:qfocus>
    <maov:qCsemantic>REQUEST_FOR_INFORMATION</maov:qCsemantic>
  </maov:analyzedQuestion>
  <maov:language>en-us</maov:language>
</mtdl:analyzedData>
```

아래 예제는 “김치를 만드는 방법은?” 이라는 질문에 대한 분석 결과를 보여준다. 예제의 질문 분석 결과는 질문의 분야가 “요리”이고 질문의 주제는 “김치”이며 질문의 초점은 “HOW_QUESTION”(방법 질문) 그리고 이 질문의 목적은 “REQUEST_FOR_METHOD”(방법 요청)이라는 것을 보여준다.

```
<mtdl:analyzedData xsi:type="maov:QuestionAnalysisType">
  <maov:analyzedQuestion qdomain="Cooking">
    <maov:qtopic>Kimchi</maov:qtopic>
    <maov:qfocus>HOW_QUESTION</maov:qfocus>
    <maov:qCsemantic>REQUEST_FOR_METHOD</maov:qCsemantic>
  </maov:analyzedQuestion>
  <maov:language>en-us</maov:language>
</mtdl:analyzedData>
```

4. 결론

본고에서는 현재 인공지능이 중심이 되는 지식산업혁명 시대에 편리하게 필요한 지식을 찾고 이용할 수 있는 자연어처리 기반의 인터페이스를 위한 응용 기술인 음성 인터페이스 기술, 자동통역기술, 자연어 질의응답 기술에 대하여 소개하고 기술 현황과 표준화 현황에 대하여 기술하였다. 또한 이러한 자연어 인터페이스를 사물 인터넷 환경에서 활용할 수 있는 방식을 MPEG IoMT 표준을 중심으로 소개하였다.

참 고 문 헌

- [1] DIS of ISO/IEC 23093-3:2018 Information technology- Internet of media things-Part 3: Media data formats and API.