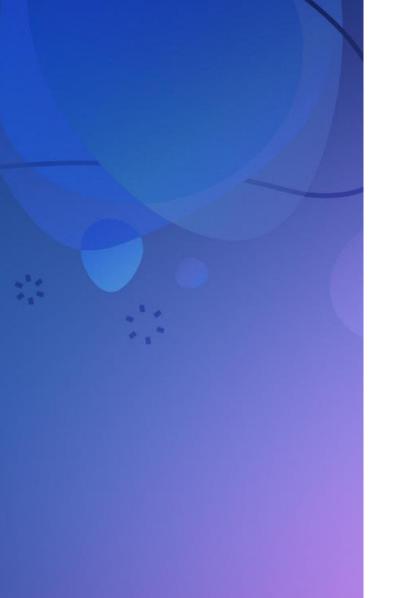


# NQA





# [NQA]

- ① QA 시스템 동향
- ② NQA의 개요
- ③ 데이터 모델 구조
- ④ 관리 도구와 플랫폼 연동
- ⑤ 실사용 예시

### QA 시스템 동향

대화 시나리오 기반 챗봇

룰 매칭 기반 챗봇

상황인지 기반 챗봇

대화 시나리오 및 단순 Keyword 기반 질의 답변 질문/답변 기계학습 기반 최적의 Rule Matching 기반

상황 인지 및 지능형 대화 생성

- 대화를 위한 시나리오 기반
- 키워드 인식 기반의 답변 추론
- 인간 경험치에 의한 룰 베이스
- 의미 추론 기반의 신경망
- 지도 학습 기반의 토픽 모델링 지도 학습 기반의 토픽 모델링
  - 의미추론 기반의 신경망
  - 인공지능 기반의 언어 모델링





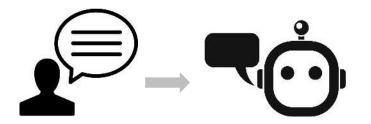


# QA 시스템 최신 기술

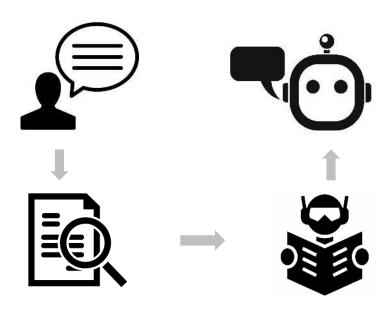




QA 시스템



MRC 응용 QA 시스템



#### TF - IDF

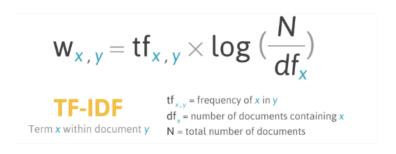
#### 상대 빈도 분석 : Term Frequency Inverse Document Frequency

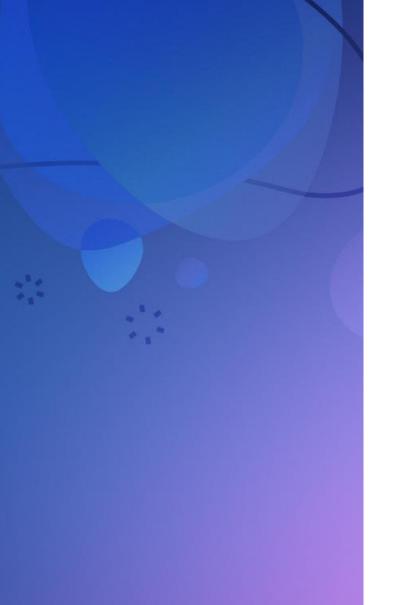
- 1. 문서의 내용을 쉽게 벡터로 표현하는 고전적 방식
- 2. Term Frequency : 해당 문서 내 Token 발생 빈도
  - (문서 Token출현 빈도)/(문서 전체 Token개수)
- 3. Inverse Document Frequency : 문서 빈도의 역
  - log (전체 문서 수 /Token포함 문서 수)

- 1. TF는 해당 문서만 있으면 바로 연산이 가능
- 2. IDF는 모집단의 Token별 통계 데이터가 필요









# [NQA]

- ① QA 시스템 동향
- ② NQA의 개요
- ③ 데이터 모델 구조
- ④ 관리 도구와 플랫폼 연동
- ⑤ 실사용 예시

#### BQA vs NQA

- 답변(Answer)에 여러가지 정보와 질문들이 연결되어 있어 원하는 답변을 한번에 찾을 수도 있고, 질문을 여러번 시도하여 원하는 답변을 추론해 나가는 방식으로 찾을 수도 있음 (대화형 진행 가능)
- 데이터를 단순하게 색인하는 구조가 아닌 계층적으로 구축하고 모델링 후 색인이 가능하여 관리도 효율적으로 할 수 있음
- 여러가지 새로운 검색기능과 옵션들을 활용하여 다양한 챗봇환경에 맞추어 개별적으로 모델링이 가능
- 단순한 질문-답(Q-A) 쌍의 구조로는 원하는 답변을 검색해 내기에 부족하여 새로운 기능 및 옵션을 추가로 개발

# 

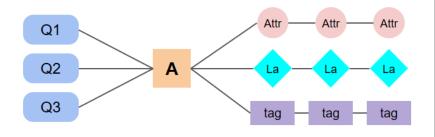
NQA Engine

#### NQA

#### **Natural Questions - Answering**

자연어로 입력한 질문에 대해 연관된 답변을 검색하고, 후보 답변이 여러 개인 경우 공통 레이어에 기반하여 사용자의 선택을 통해 후보군을 좁혀나가, 원하는 답변을 최종적으로 선택 가능하게 하는 QA 시스템

단순한 검색 엔진에 의한 답변이 아니라, layer/ tag/ attribute/ ner 등 다양한 색인과 검색을 지원



#### NQA Engine 기능

답변에 여러 가지 정보와 질문들이 연결되어 있어, 원하는 답변을 한번에 찾을 수 있다.(Score)

질문을 여러 번 시도하여 원하는 답변을 추론해 나가는 방식으로 찾을 수 있다.(Layer 선택)

데이터 구축 시, 질문/답변 색인 외에 layer/ tag/ attribute/ ner 등을 활용하여 계층적으로 색인 및 모델링 한다.

검색 시 질문/ 답변 검색 외에, layer/ tag/ attribute/ ner 검색을 지원하며, 후보 답변이 여러개인 경우 layer 선택을 통해 답변을 좁혀나갈 수 있다.

원하는 칼럼 (질문 / 답변 / layer / attribute / tag)에 가중치 적용 가능

동의어 처리, 생략어 복원, 불용어 적용 가능

#### NQA 특장점

- 1) 정답률이 높고, 데이터 계층에 따른 재질의 구조
  - a) 기존 BQA의 싱글턴 대화방식을 탈피하고 정답률이 높음
  - b) 데이터 계층을 구축하면 코딩 없이 멀티턴을 유지 할 수 있음.

	NQA	대화 시나리오 모델	기존 Q-A 모델
멀티턴 가능	0	0	х
코딩 불필요	Δ	Х	0
높은 정답률	0	0	Х

#### NQA 특장점

2) 단순 (질문-답)쌍의 구조로 질문만을 검색하지 않고 다양한 방식의 검색과 옵션 사용 가능

#### a) 검색

- ) 질문 검색 (형태소, string)
- ii) 답변 검색 (형태소, string)
- iii) NER 검색
- iv) Tag 검색
- v) Layer 검색
- vi) Attribute(발화자 속성) 검색

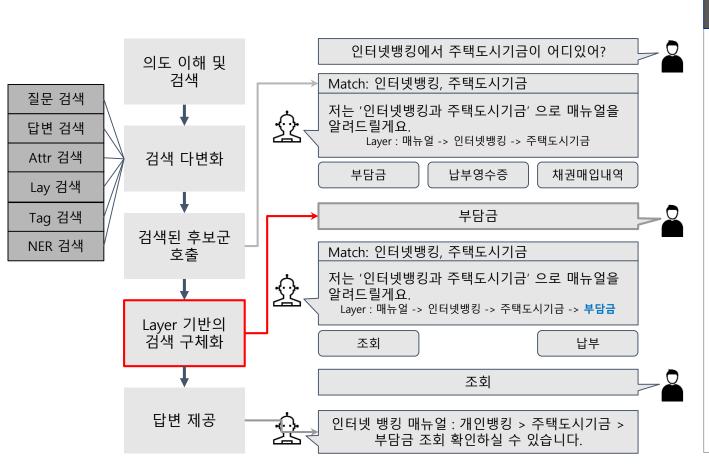
#### a) 추가 옵션 기능

- i) 가중치 적용 기능
  - 원하는 컬럼(Q,A,Attr,Layer,Tag)에 가중치 적용 가능
- i) Score 필터링 기능(TF-IDF, Morph mm, Morph size)
  - Score로 정렬하여 원하는 점수 이상만큼만 탐색 가능
- i) 동의어(대표어) 적용 기능
  - 대표, 사장, CEO -> 대표
- i) Filter out 기능
  - 검색에 포함되지 않길 원하는 단어는 제외 가능

#### NQA 특장점

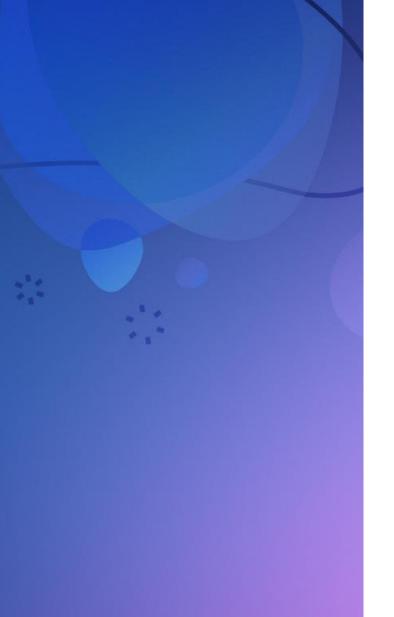
- 3) 분석된 데이터의 구조가 Meta data에 보이기 때문에, 데이터의 로직을 볼 수 있음
  - meta data의 details에 TF-IDF의 score을 보여줌
- 4) 제외된 검색 조건이 보여짐.
  - 검색 턴마다 어떠한 검색 결과가 잡혀서 결과가 나왔는지 보여줌
- 5) UI 화면에 select card(버튼) 추가
  - select card 를 통한 사용자 편의성 증가
- 6) 질문에서 검색이 안될 시 다른 곳을 탐색
  - 현재 질문에서 찾지 못했을때, 답변에서 검색하는 로직.
  - custom N => Q => A => AA => LA => T => N => SQ => SA (변경 가능)
- 7) 불용어 처리
  - 중요하지 않지만 검색에 영향을 주는 단어 제외
- 8) 형태소 및 동의어 사전 구축

#### 여러 조건을 이용한 검색형 QA엔진



#### 특징 및 장점

- 질문을 여러 번 시도하여 원하는 답변을 추론해 나가는 방식으로 찾을 수 있음 (Layer 선택)
- 데이터 구축 시, 질문/답변 색인 외에 layer/ tag/ attribute/ ner 등을 활용하여 다층적으로 색인 및 검색
- 원하는 칼럼 (질문 / 답변 / layer / attribute / tag)에 가중치 적용 가능
- 답변은 좀 더 세밀하게 하기 위한 정보로 정확한 정보에 대한 대응 가능 (Attr 기능)
- 검색 시 특정 QA 가중치를 올리는 방식으로 검색 구체화



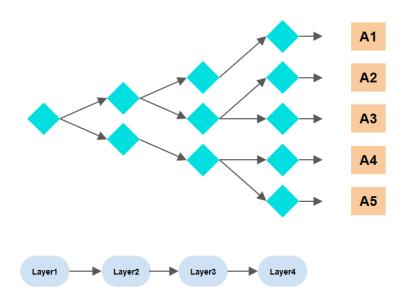
# [NQA]

- ① QA 시스템 동향
- ② NQA의 개요
- ③ 데이터 모델 구조
- ④ 관리 도구와 플랫폼 연동
- ⑤ 실사용 예시

#### 1) Layer 계층 구조 적용 (Tree)

- 질문을 입력하였을때 여러 개의 Answer 후보가 나오면 여러 개로 나뉠 때의 layer를 선택할 수 있으며 재질의를 통해 추론하는 형식으로 Answer를 찾아갈 수 있음
- Tree 구조는 **하나의 'Layer 순열'에 한 개의 답변**만 매핑되어야 하며, 지식의 계층 체계가 잘 갖춰진 사이트에서 개발할 경우 빠른 구축과 높은 정답률이 나옴

#### < Layer 계층 Tree 구조 >

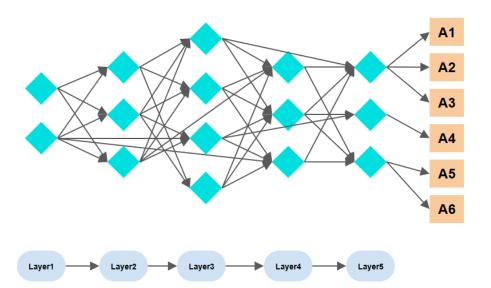


#### < 사용 가이드 >

- ●Q A 데이터가 적을 때 사용.
- ●전체 Layer 층이 낮거나, 한 Layer 층이 좁을 때 사용.
- ●'Layer 순열: Answer = 1:1' Mapping 구조.
- ●Score에 영향을 적게 받으며, 확실한 답변을 보내줄 수 있음.
- ●초기에 Layer 계층 구조를 숙지한 사람이 써야 하는 단점이 있음.

- 2) Layer 계층 구조 적용 (FeedForward Network)
- 질문을 입력하였을때 여러 개의 Answer 후보가 나오면 여러개로 나뉠때의 layer를 선택할 수 있으며 재 질의를 통해 추론하는 형식으로 Answer를 찾아갈 수 있음
- FeedForward Network Layer 구조는 하나의 **'Layer 순열'에 여러 개의 답변이 매핑**될 수 있으며, Layer 선택이 끝난 후 Top Rank 답변을 뿌림.

< Layer 계층 구조 >FeedForward Network (1:N)

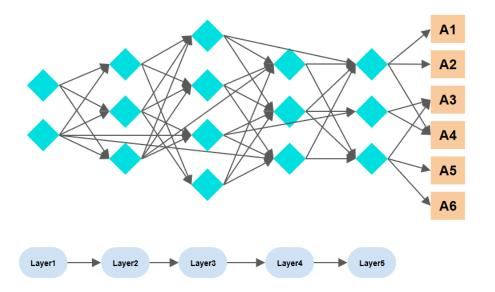


#### < 사용 가이드 >

- ●Q A 데이터가 많을 때 사용.
- ●전체 Layer 층이 깊거나, 한 Layer 층이 넓을 때 사용.
- ●'Layer 순열: Answer = 1: N' Mapping 구조.
- ●Score에 영향을 크게 받으며, Tag를 통해 Score 가중치를 높여 줘야 하는 단점이 있음.
- ●초기에 Layer 계층 구조를 숙지하지 않아도 사용 가능하다는 장점이 있으며, 적은 Layer 계층으로도 많은 데이터를 처리 가능하다는 장점이 있음.

- 3) Layer 계층 구조 적용 (FeedForward Network)
- 질문을 입력하였을때 여러 개의 Answer 후보가 나오면 여러개로 나뉠때의 layer를 선택할 수 있으며 재 질의를 통해 추론하는 형식으로 Answer를 찾아갈 수 있음
- FeedForward Network Layer 구조는 하나의 '**Layer 순열**'에 **여러 개의 답변이 매핑**될 수 있으며, Layer 선택이 끝난 후 Top Rank 답변을 뿌림.

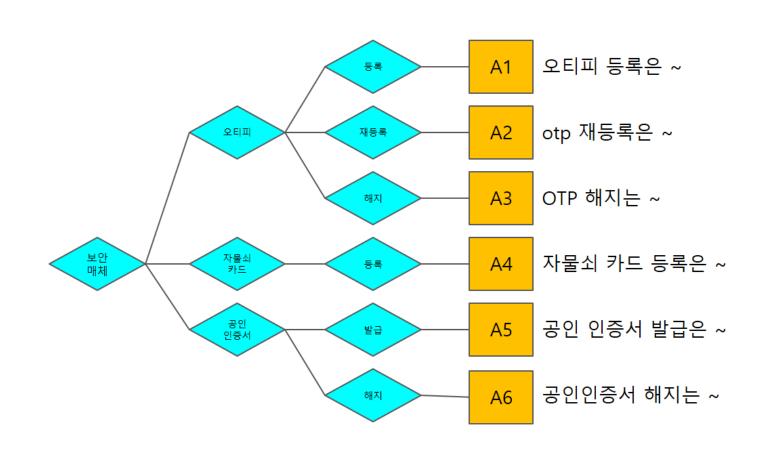
< Layer 계층 구조 >FeedForward Network (N:N)



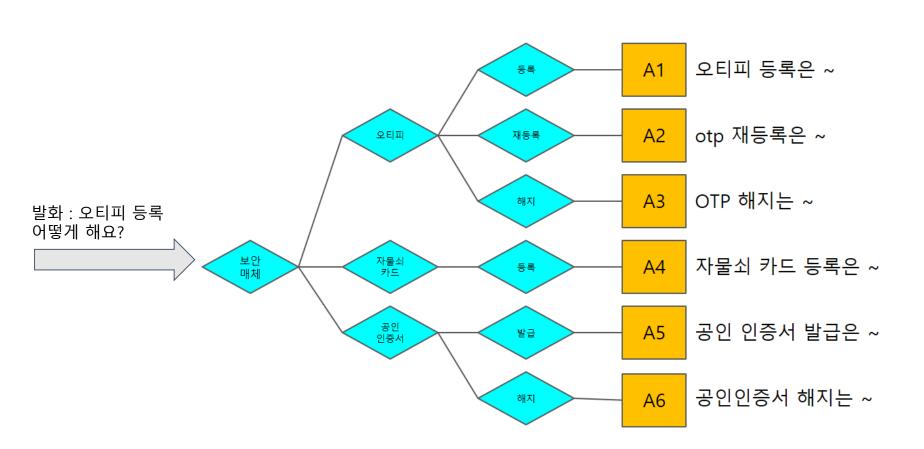
#### < 사용 가이드 >

- ●Q A 데이터가 많을 때 사용.
- ●전체 Layer 층이 깊거나, 한 Layer 층이 넓을 때 사용.
- ●'Layer 순열: Answer = N: N' Mapping 구조.
- ●Score에 영향을 크게 받으며, Tag를 통해 Score 가중치를 높여 줘야 하는 단점이 있음.
- ●초기에 Layer 계층 구조를 숙지하지 않아도 사용 가능하다는 장점이 있으며, 적은 Layer 계층으로도 많은 데이터를 처리 가능하다는 장점이 있음.
- <u>직전 모델의 변형</u>으로, 한 층에서 Layer가 중복되는 경우 사용

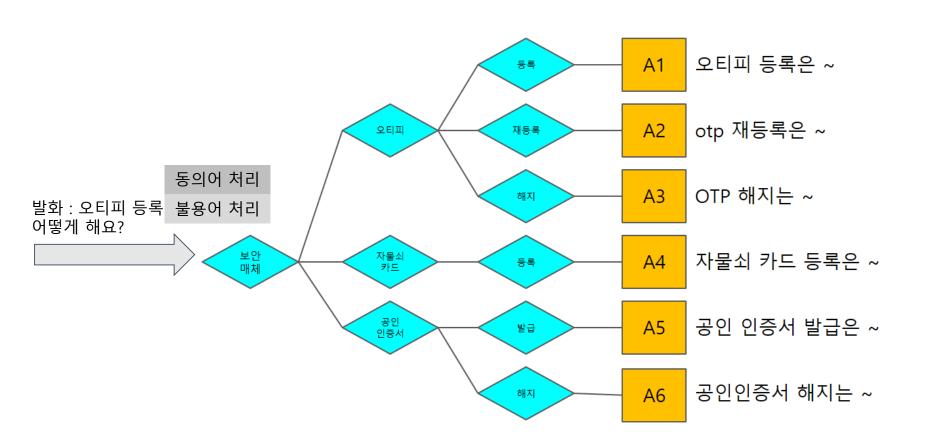
Layer 모델 구조와 발화 인입 – 인덱싱 상태



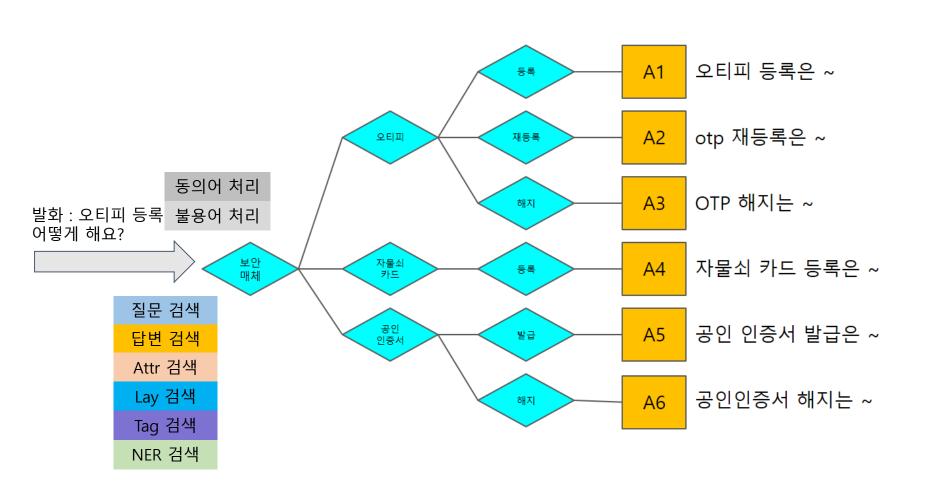
Layer 모델 구조와 발화 인입 – 발화 인입



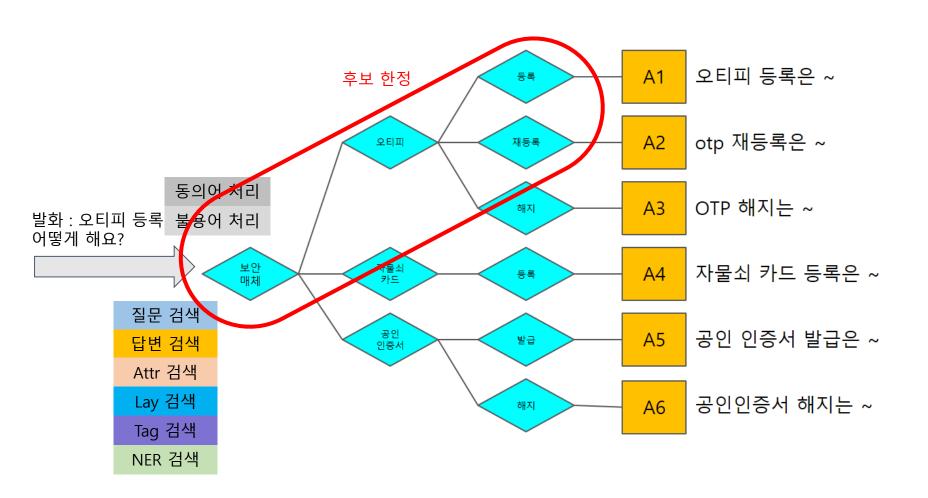
Layer 모델 구조와 발화 인입 – 동의어 처리, 불용어 처리



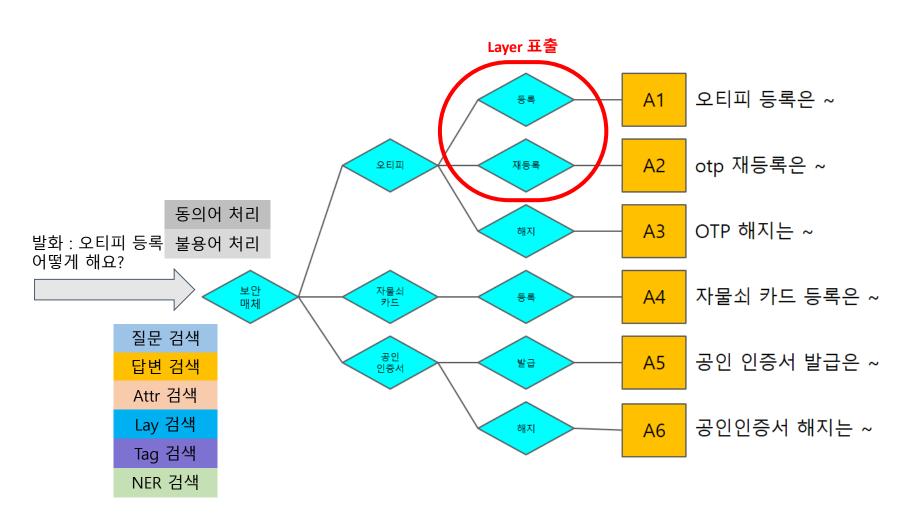
표출 답변 후보군 한정과 검색 – 검색 다양화



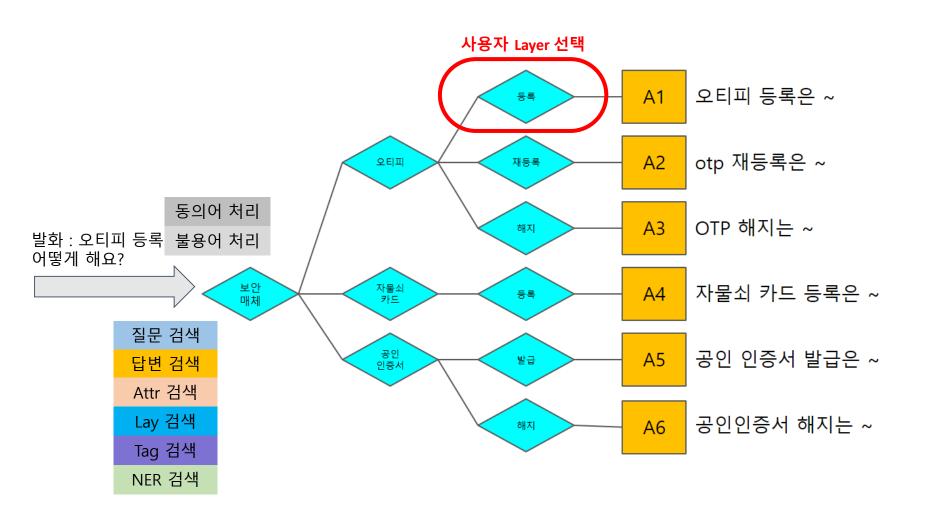
표출 답변 후보군 한정과 검색 – 답변 후보 한정



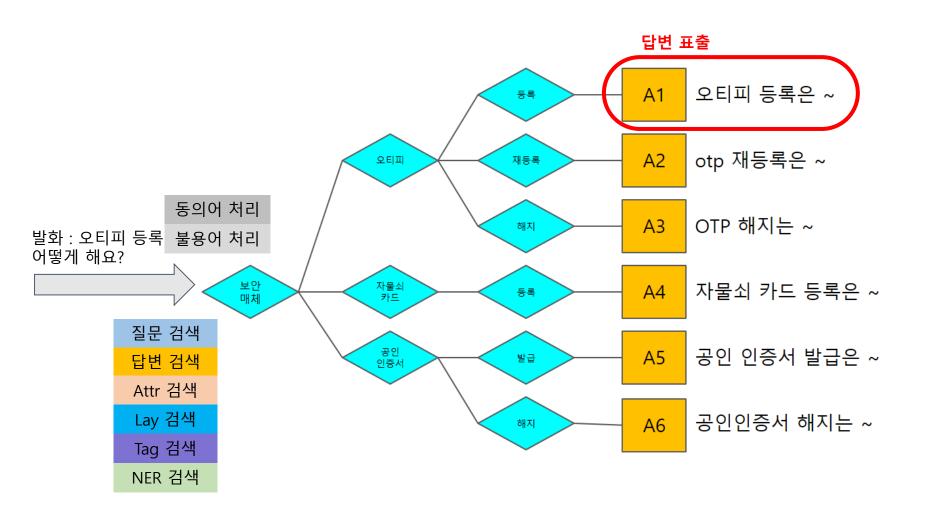
표출 답변 후보군 한정과 검색 – 화면에 Layer 표출



표출 답변 후보군 한정과 검색 – Layer 선택에 따른 질의 구체화



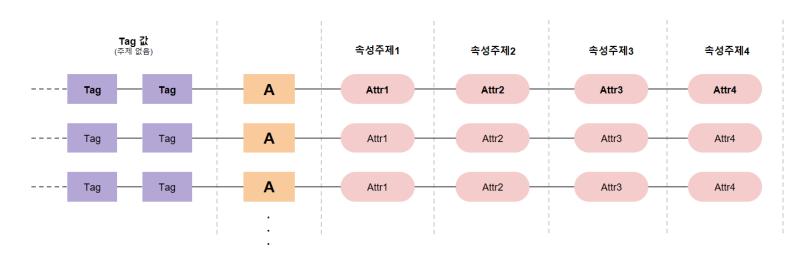
표출 답변 후보군 한정과 검색 – 사용자 선택에 따른 답변 표출



#### Attribute 및 Tag 개념 적용

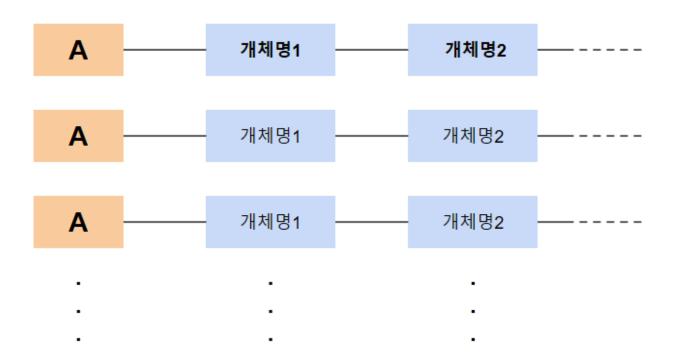
- Attribute는 주제를 정해놓은 속성값으로, 답변을 좀 더 세밀하게 하기 위한 정보
  - 원장 정보, 고객 정보 등에 따라 고객에게 **답변 표출 여부를 선택** 가능(키워드 아님)
- Tag는 인스타그램의 해시태그처럼 의미있는 단어를 달아놓는 기능
  - Q나 A에 있는 키워드를 입력하여, Q A 검색 Score의 Weight를 올림(키워드; 동의어 처리 가능)

#### < Attribute & Tag >



개체명 사전 적용

- 해당 상황에 맞는 개체명을 따로 색인함

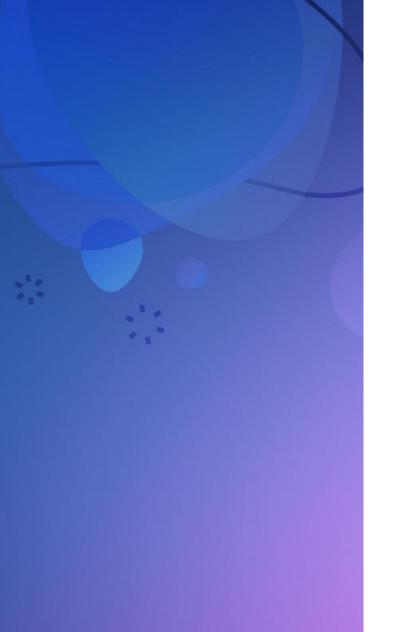


공통점과 차이점

Layer : 계층 구조를 가진 데이터 집합 답변을 추론하며 찾아가기 위한 정보

Attribute : 주제를 정해놓은 속성값 답변은 좀 더 세밀하게 하기 위한 정보 관리 용이 미리 알아야 하는 정보 ex) 19금, DNN(XDC) 활용, 부가적 정보 활용...

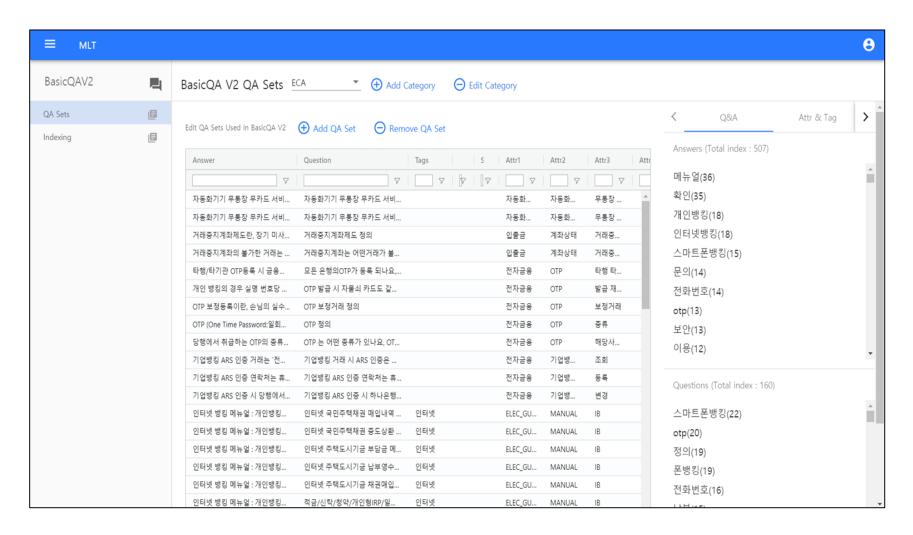
Tag: 주제를 명확히 정해놓지 않은 속성값 답변을 좀 더 잘 찾아내기 위한 정보 연관된 키워드들을 달아놓고 마지막에 검색하여 답변 추출 ex) 유사 키워드 검색 Q - A 안에 있는 키워드 들의 Score Weight를 올림



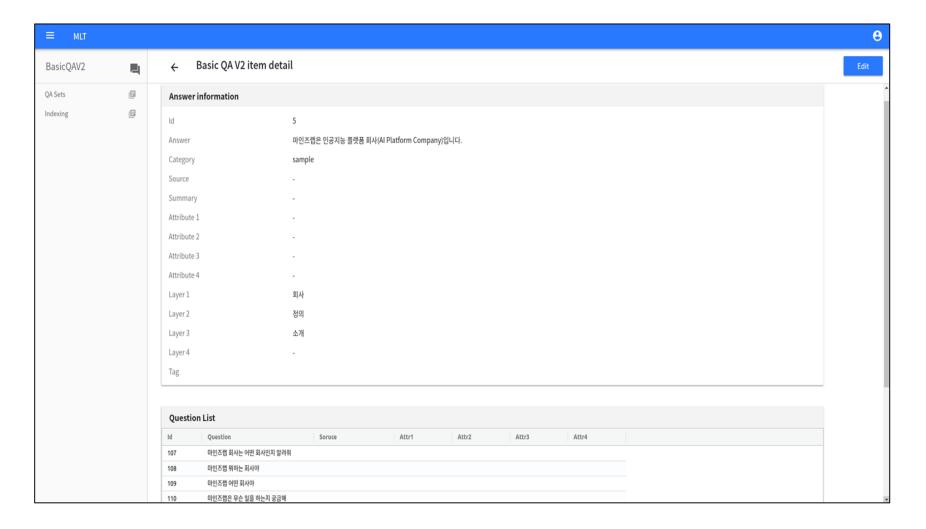
# [NQA]

- ① QA 시스템 동향
- ② NQA의 개요
- ③ 데이터 모델 구조
- ④ 관리 도구와 플랫폼 연동
- ⑤ 실사용 예시

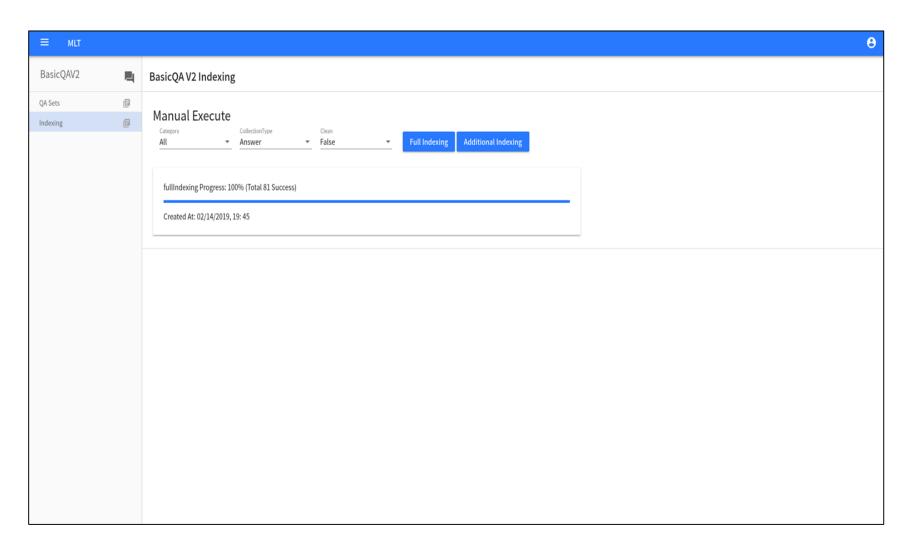
NQA 관리도구 (QA Sets List)



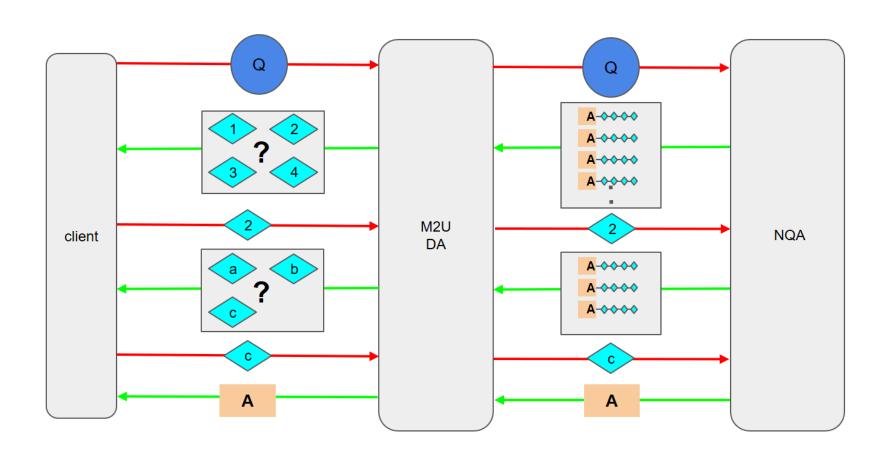
NQA 관리도구 (QA Sets Detail)

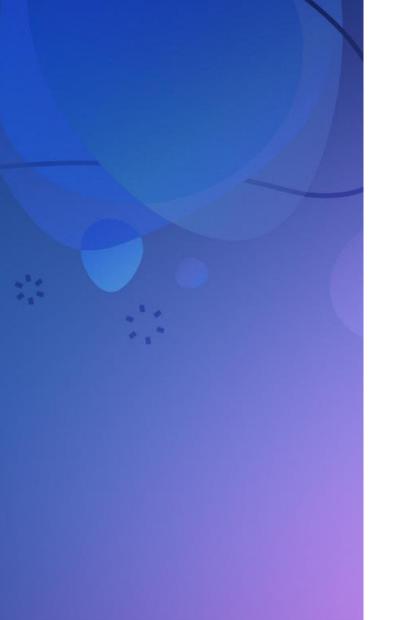


NQA 관리도구 (Indexing)



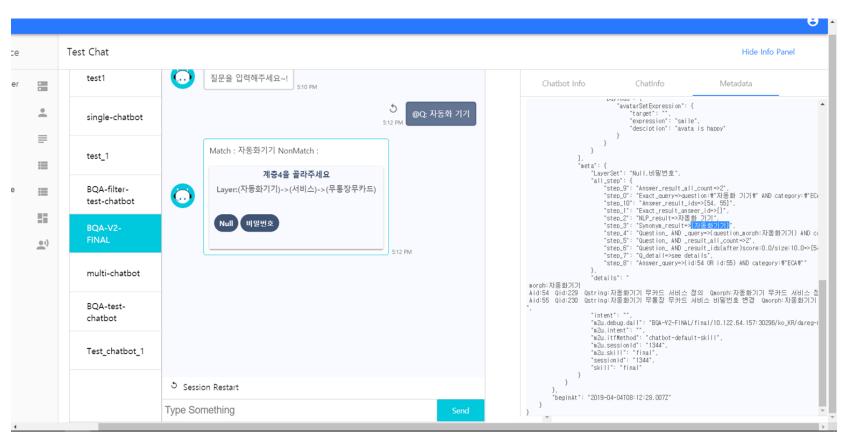
NQA 관리도구 (QA Sets Detail)





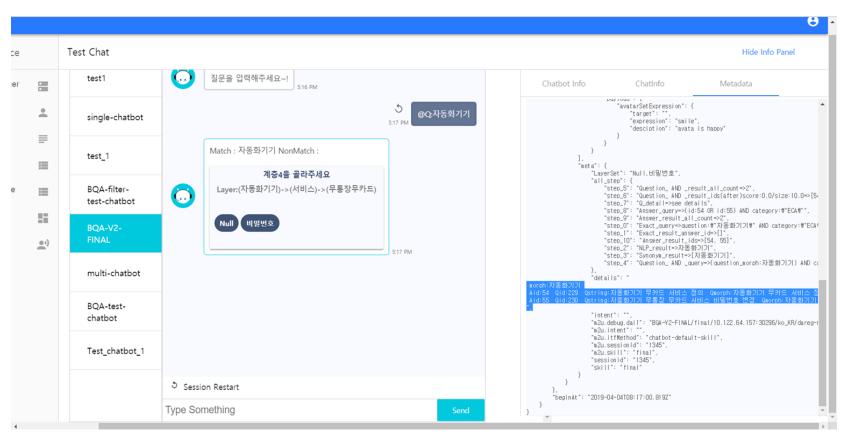
# [NQA]

- ① QA 시스템 동향
- ② NQA의 개요
- ③ 데이터 모델 구조
- ④ 관리 도구와 플랫폼 연동
- ⑤ 실사용 예시



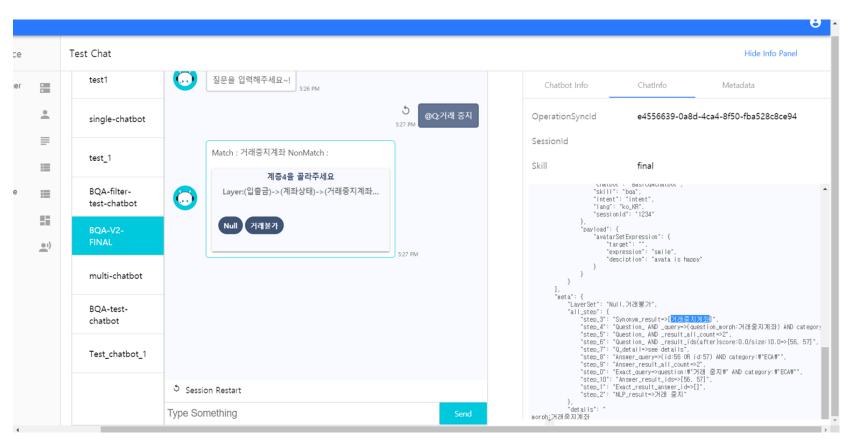
@Q:자동화 기기 로 검색

자동화기기로 동의어 치환 됨. (indexing 때문)



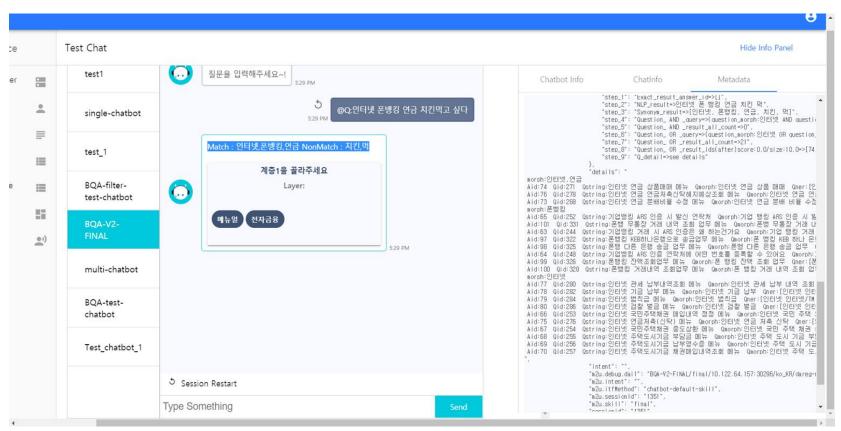
@Q:자동화 기기 로 검색

자동화기기 형태소가 있는 Answer id가 두개 있고, layer 선택을 통해 최종 답변을 선택가능

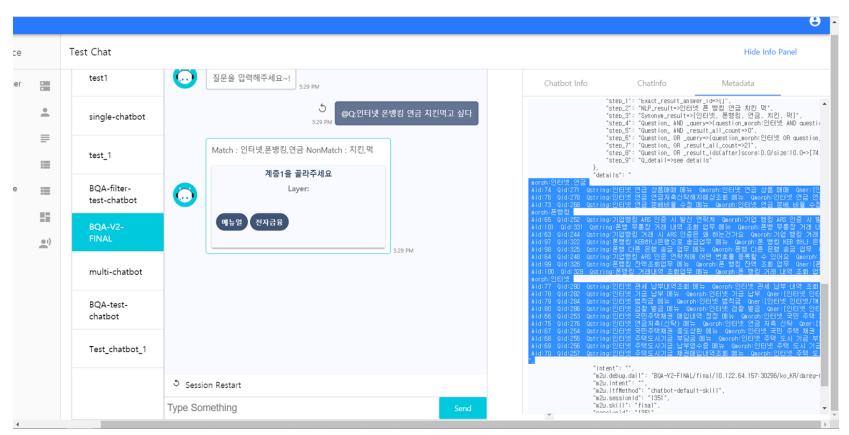


@Q:거래 중지 로 질문 검색

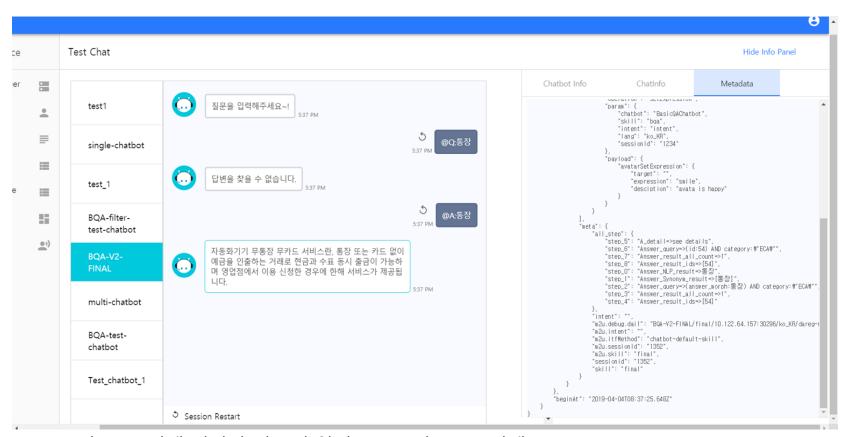
거래 중지의 생략어 계좌를 동의어 처리를 통해 복원시킴 : 동의어 처리 결과 - 거래중지계좌



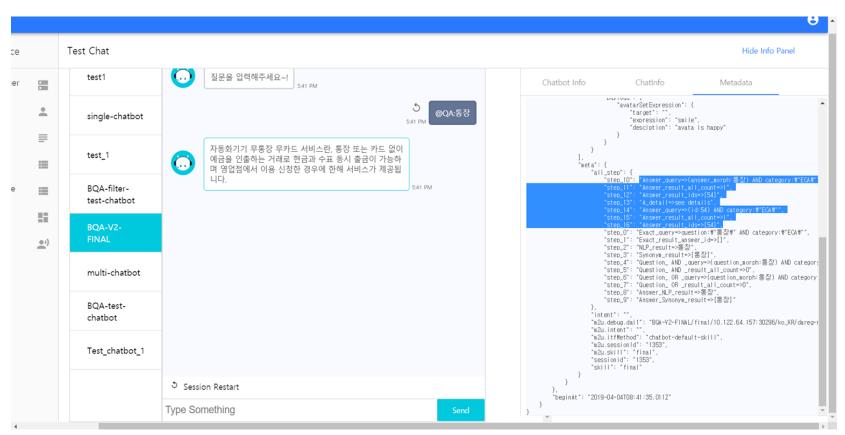
@Q:인터넷 폰뱅킹 연금 치킨먹고 싶다 로 검색질문 데이터 set에 어떤 형태소가 매칭되고 매칭이 되지 않았는지 보여줌. 매칭된 형태소: 인터넷 폰뱅킹 연금 / 매칭되지 않은 형태소: 치킨 먹



@Q:인터넷 폰뱅킹 연금 치킨먹고 싶다 로 검색 형태소가 걸린 질문과 답변의 id, 형태소 등을 meta데이터로 보여줌. "인터넷, 연금 / 폰뱅킹 / 인터넷" 으로 걸린 질문들이 보여지고 있음



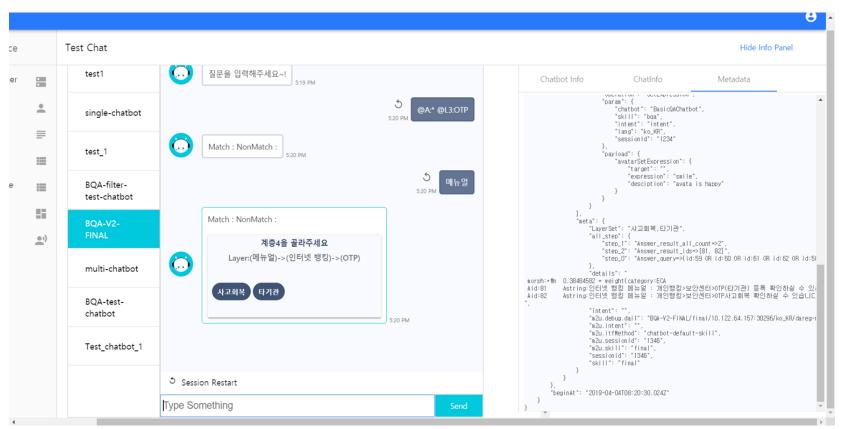
@Q:통장 으로 검색 결과가 나오지 않아 @A:통장 으로 검색 답변 검색 결과 답이 나옴.



@QA:통장 으로 검색

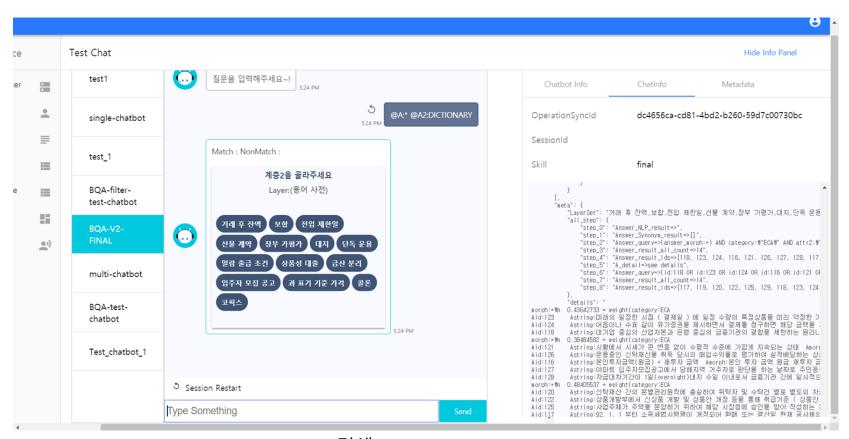
답변 검색 결과 답이 나옴.

시연을 위해 API 별로 쪼갠 것일 뿐, 현재 NQA구조는 질문 검색으로 답이 없을 경우 답변을 검색함.

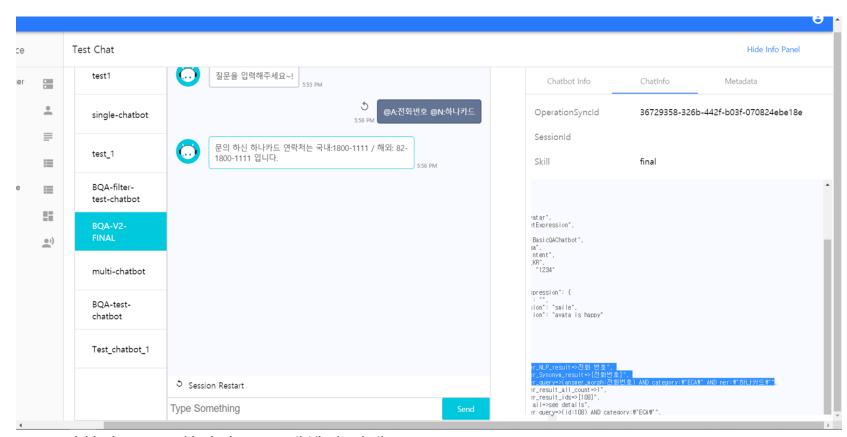


@A:\* @L3:OTP 로 layer검색

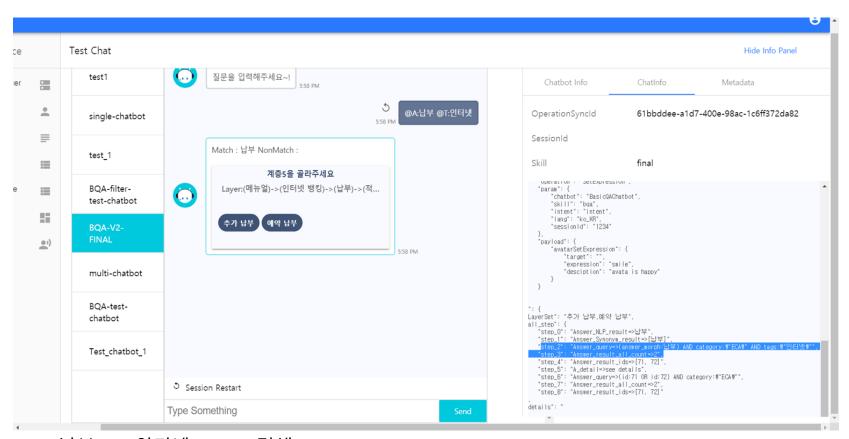
layer 검색한 결과로 otp를 찾음. 같은 layer인 "메뉴얼 -> 인터넷뱅킹 -> OTP"는 자동으로 선택됨.



@A:\* @A2:DICTIONARY 로 attribute 검색 attribute 검색한 결과로 사전 찾음. 같은 attribute 인 "용어사전 -> DICTIONARY"는 자동으로 선택됨.



@A:전화번호 @N:하나카드 로 개체명 검색 개체명 검색한 결과로 답변을 찾음.



@A:납부 @T:인터넷 로 tag 검색 tag 검색한 결과로 선택할 layer을 보여줌.