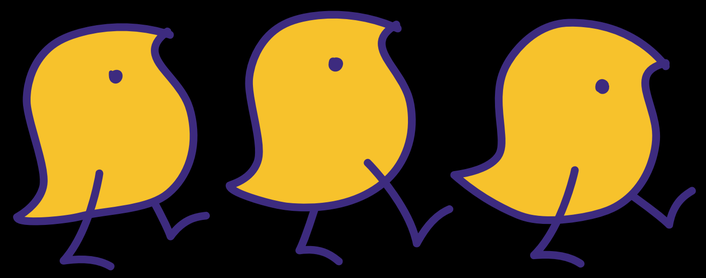
LangCon 2019

D.I.Y ourself 챗봇 v0.2

without Chatbot Builder & Deep Learning

박 혜 웅

bage79@gmail.com



without Chatbot Builder & Deep Learning

LangCon 2019

Why?

- Chatbot Builder (=Dialogue Manager)

  엄밀히 말하면 다르지만, 우선 같다고 생각하셔도 이 발표를 이해하는데 지장이 없습니다.

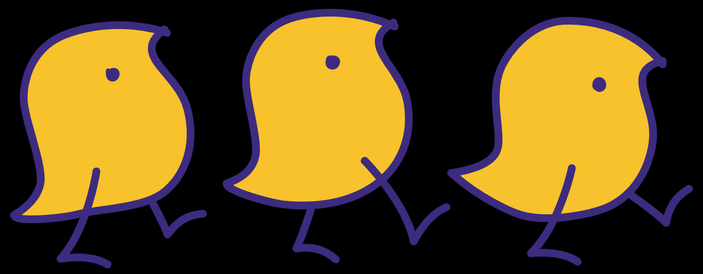
  - **We need different chatbot builders** for various chatbot services.
  - Chatbot builders **can't call some external APIs.**
- Deep Learning (=DL)
  - DL **can't perfectly control** NLU outputs(predictions).
  - I don't know DL. DL is **hard to learn**.

How?

- Messaging Platform
  - Slack, Facebook, …

이 발표에서 다루는 내용

- **Dialogue Manager (DM)**
  - **Rule based (Scenario based)**
- **Natural Language Understanding (NLU)**
  - **Intent Classification (IC)**
    - **Example based (Search-engine based)**
  - **Named Entity Recognition (NER)**
    - **Dictionary based (Regular-expression based)**
- Natural Language Generation (NLG)
  - Template based

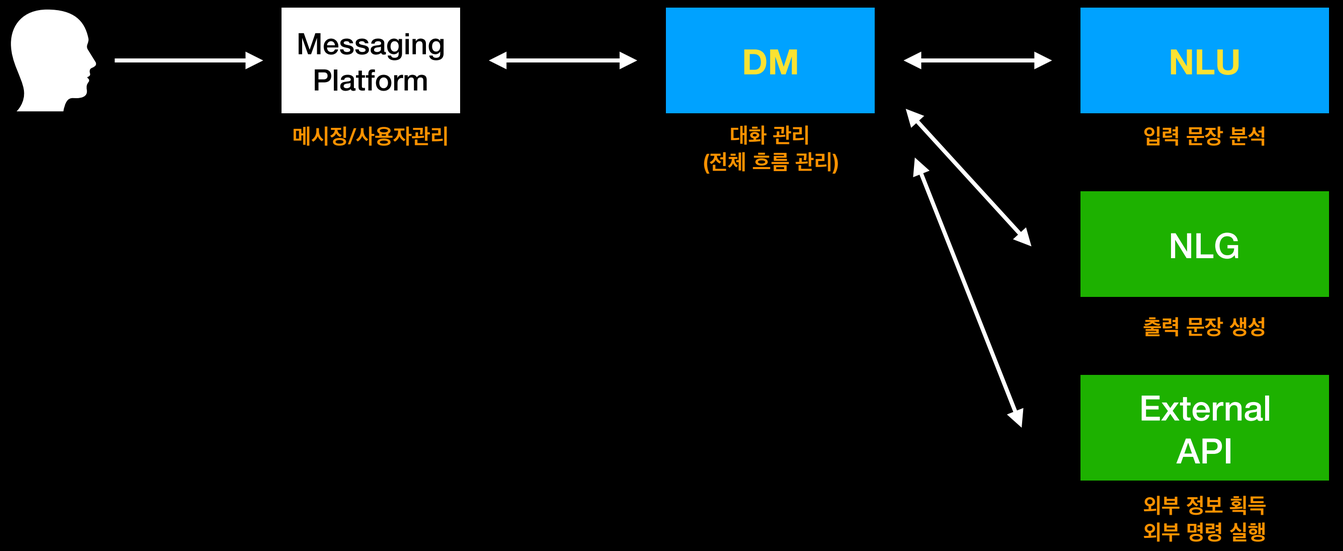


Chatbot System

LangCon 2019

Chatbot System

DIY 챗봇



Chatbot System

Messaging Platform

userid: bage

DM

NLU

NLG

External API

Chatbot System

DIY 챗봇

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

userid: bage

DM

NLU

entity: T멤버쉽
intent: check.remain

NLG

External API

Chatbot System

DIY 챗봇

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

DM

NLU

entity: T멤버쉽
intent: check.remain

userid: bage

state: T멤버쉽포인트조회

NLG

External API

Chatbot System

DIY 챗봇

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

DM

NLU

entity: T멤버쉽
intent: check.remain

userid: bage

state: T멤버쉽포인트조회

NLG

template: “[name]님의 잔여 포인트는 [point]점입니다.”

External API

Chatbot System

DIY 챗봇

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

DM

NLU

entity: T멤버쉽
intent: check.remain

userid: bage

state: T멤버쉽포인트조회

NLG

template: “[name]님의 잔여 포인트는 [point]점입니다.”

External API

name: 박혜웅
point: 1000

Chatbot System

DIY 챗봇

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

DM

NLU

entity: T멤버쉽
intent: check.remain

userid: bage state: T멤버쉽포인트조회

output: “박혜웅님의 잔여 포인트는 1000점입니다.”

NLG

template: “[name]님의 잔여 포인트는 [point]점입니다.”

External API

name: 박혜웅
point: 1000

Chatbot System

DIY 챗봇

이 발표에서 다루는 내용

input: “T멤버쉽 잔여 포인트 알려줘.”

Messaging Platform

DM

NLU

entity: T멤버쉽
intent: check.remain

userid: bage

state: T멤버쉽포인트조회

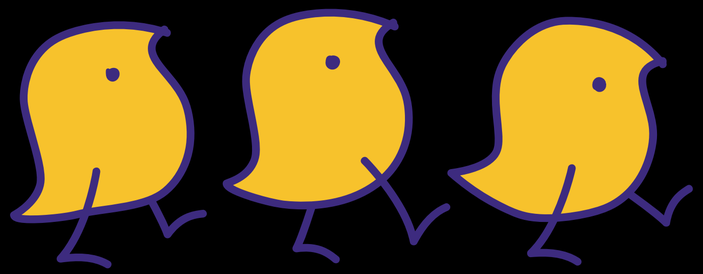
output: “박혜웅님의 잔여 포인트는 1000점입니다.”

NLG

template: “[name]님의 잔여 포인트는 [point]점입니다.”

External API

name: 박혜웅
point: 1000

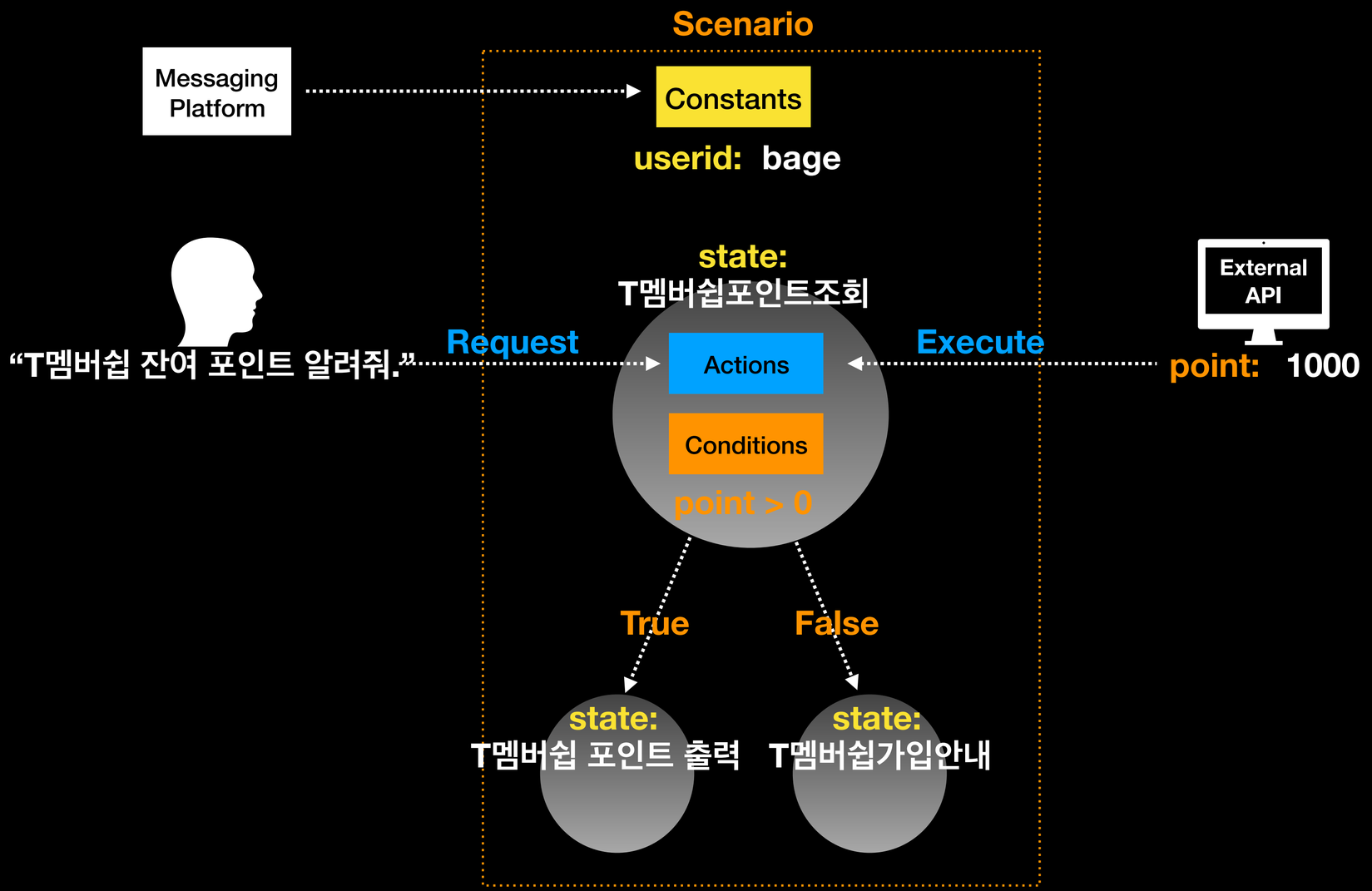


DM

Dialogue Manager

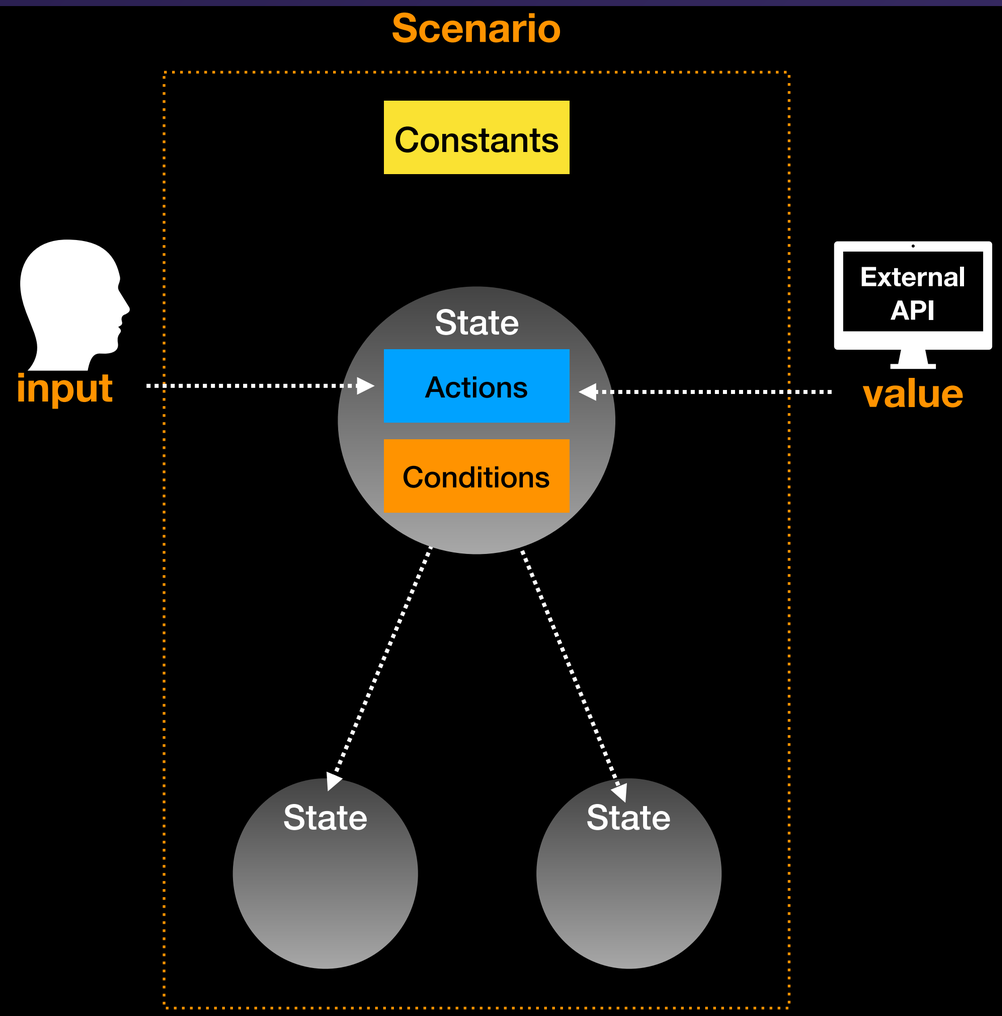
LangCon 2019

Scenario


Scenario
Messaging Platform
Constants
userid: bage
state:
T멤버쉽포인트조회
"T멤버쉽 잔여 포인트 알려줘."
Request
Actions
Conditions
point > 0
Execute
External API
point: 1000
True
False
state:
T멤버쉽 포인트 출력
state:
T멤버쉽가입안내

Scenario

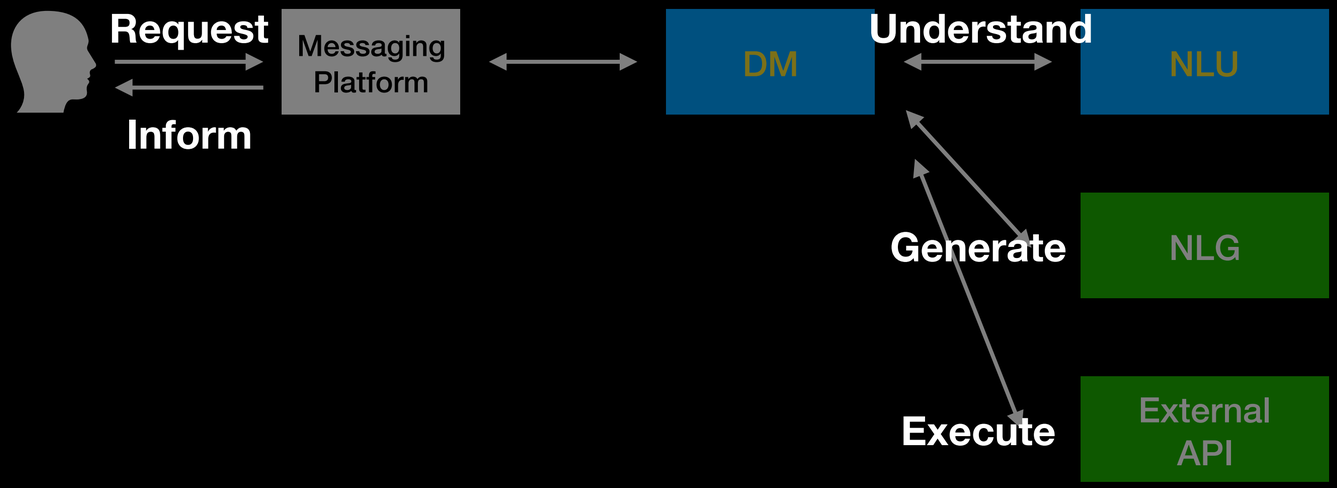
- Data Structure
  - Linked List
- Constants
  - Global
- Actions
  - for variables
- Variables
  - from User, API, NLU, NLG …
- Conditions
  - **equal**, greater, little, exists, …

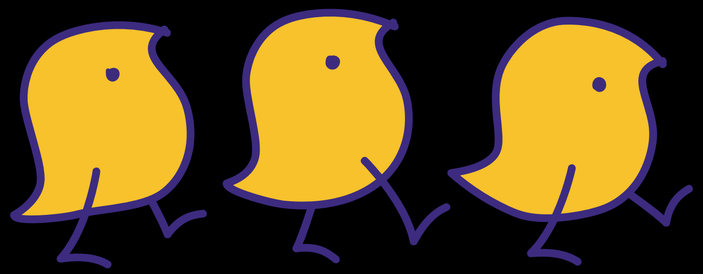


프로그래밍의 구성요소와 비슷합니다.

Actions

- Request
  - from a user
- Inform
  - to a user
- Understand
  - NLU
- Generate
  - NLG
- Execute
  - external API



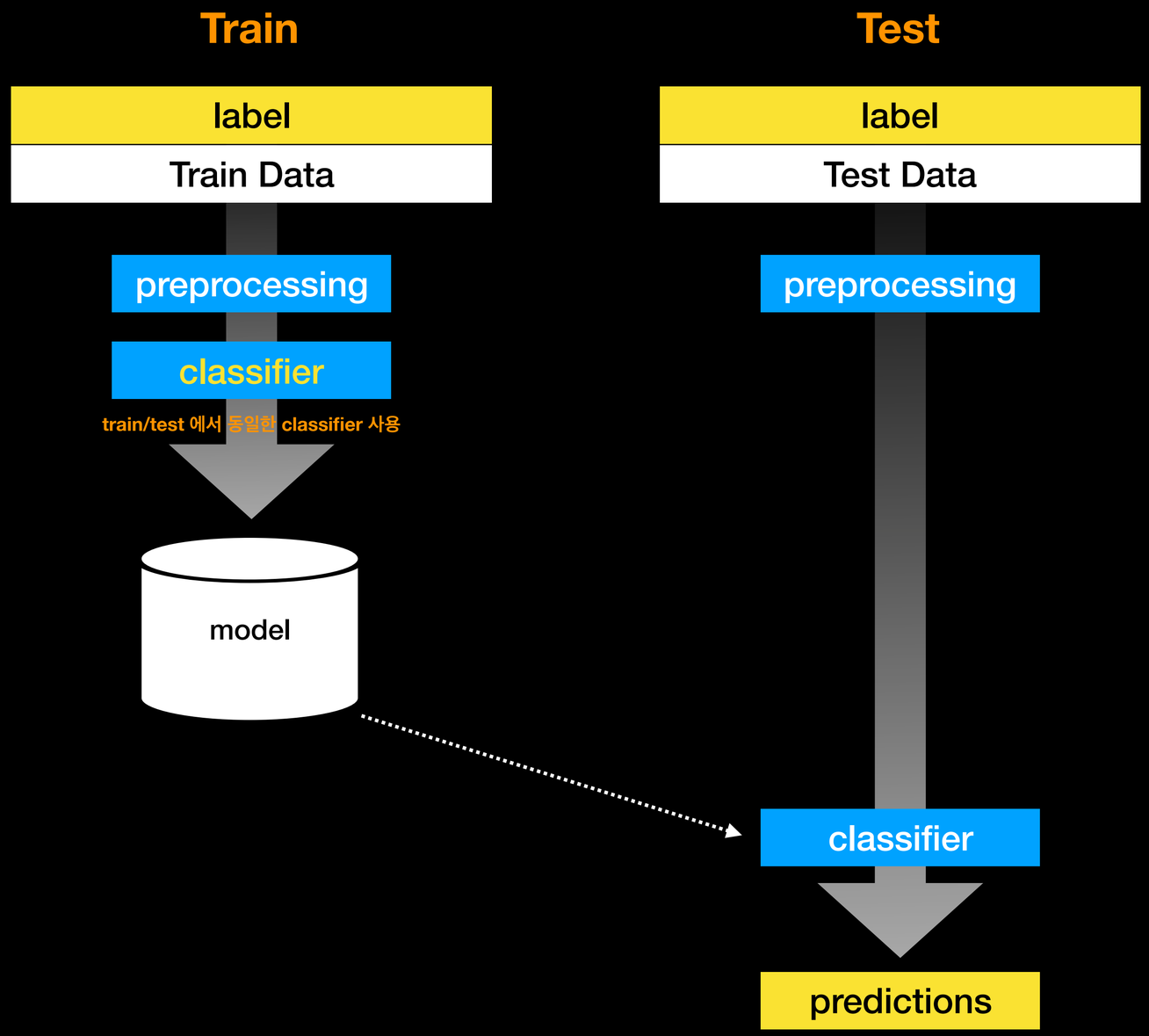


NLU

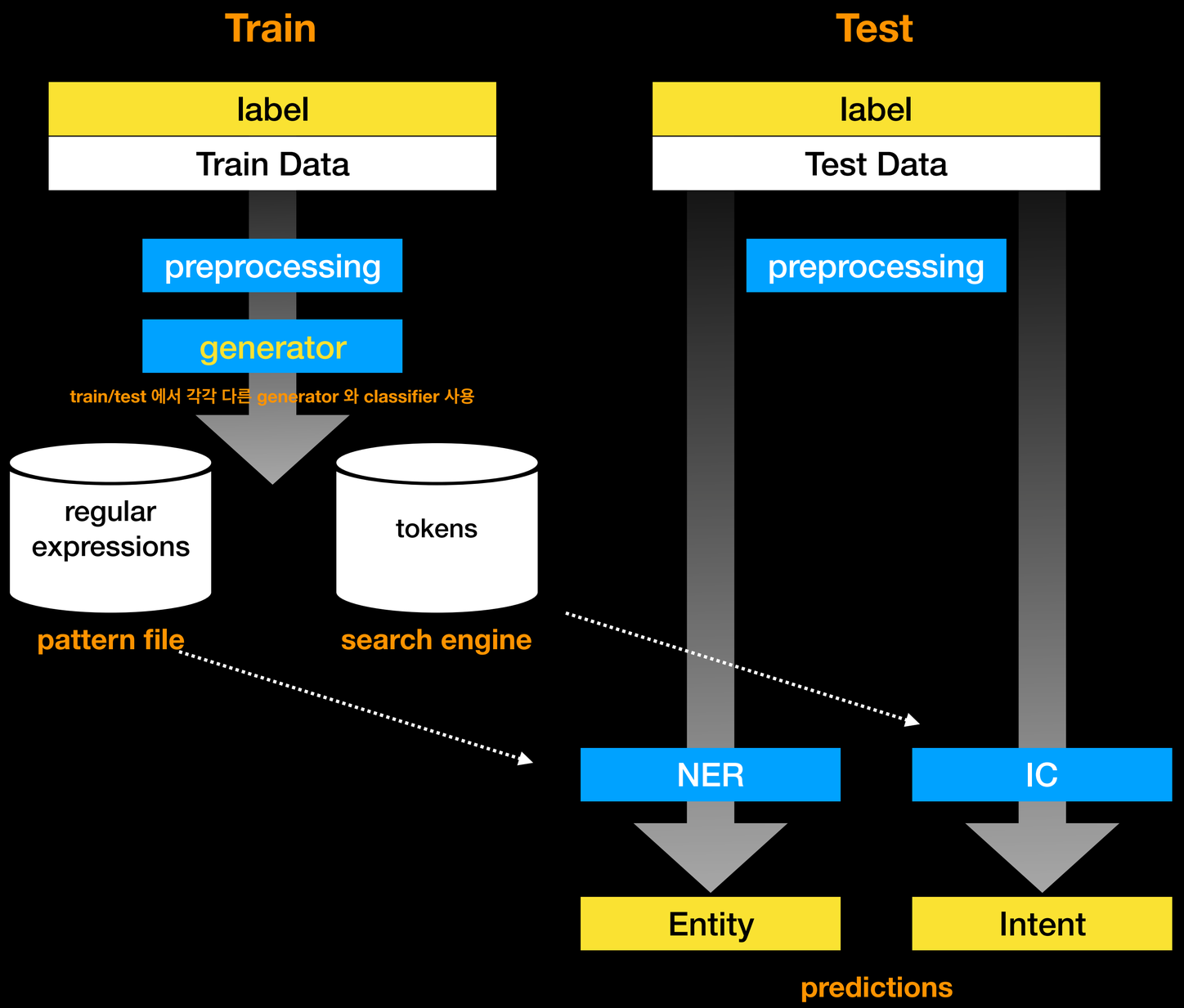
Natural Language Understanding

LangCon 2019

NLU with Deep Learning

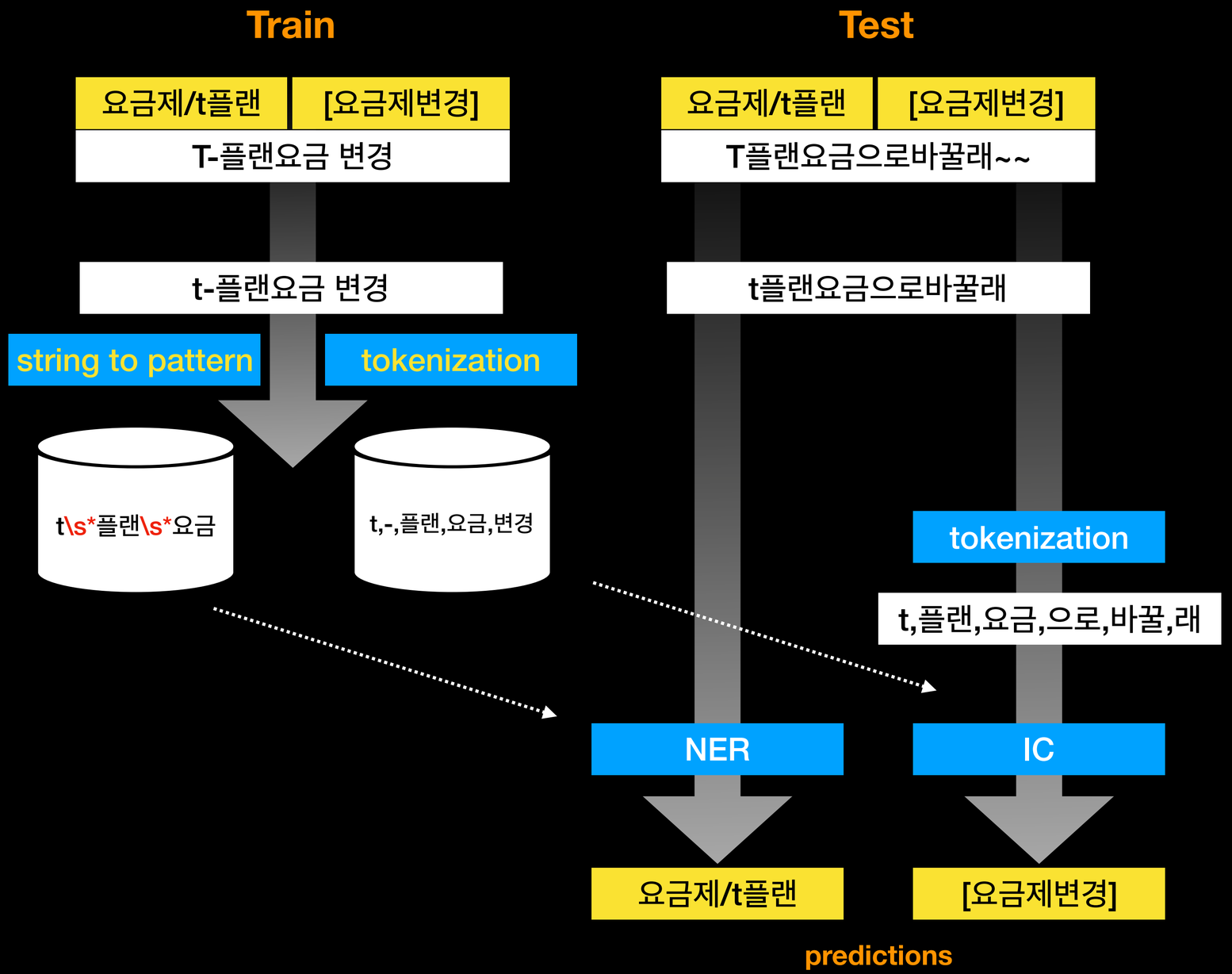


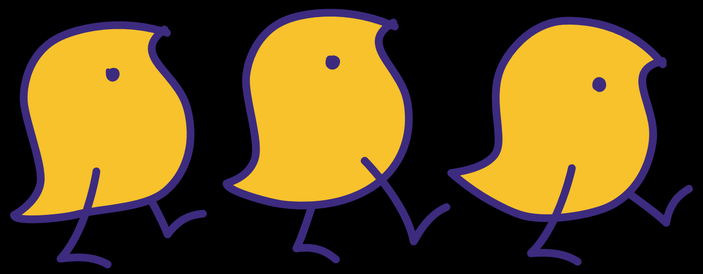
NLU without Deep Learning



NLU without Deep Learning

DIY 챗봇



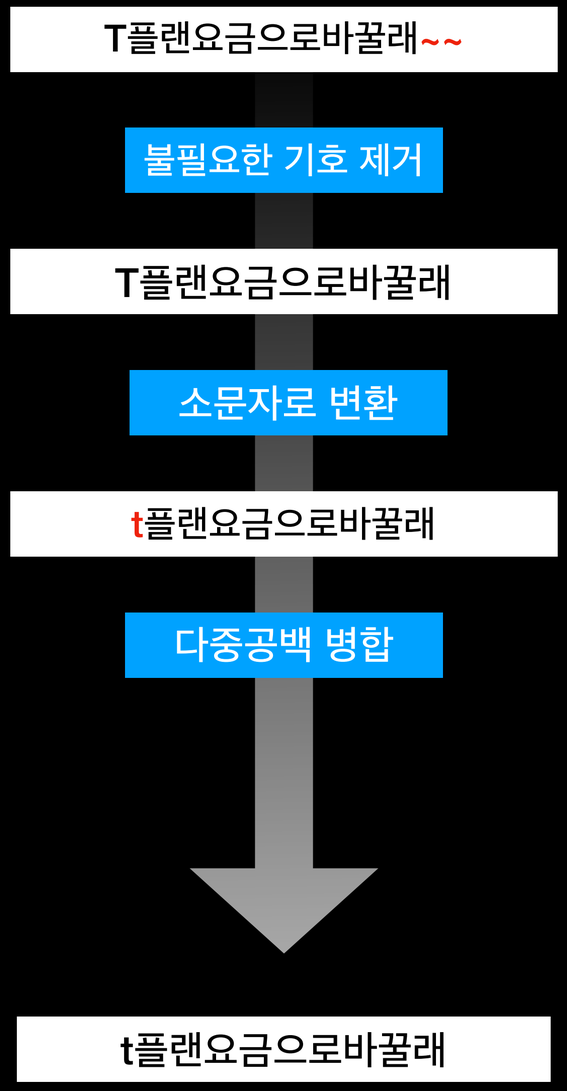


IC

Intent Classification

LangCon 2019

Preprocessing



Tokenization

t플랜요금으로바꿀래

word spacing

t,플랜,요금,으로,바꿀,래

word tokens

형태소분석기를 띄어쓰기용으로 사용하여, 어미나 조사도 분리된 경우

word spacing

morph analyzing

t,플랜,요금,으로,바꾸,ㄹ,래

morph tokens

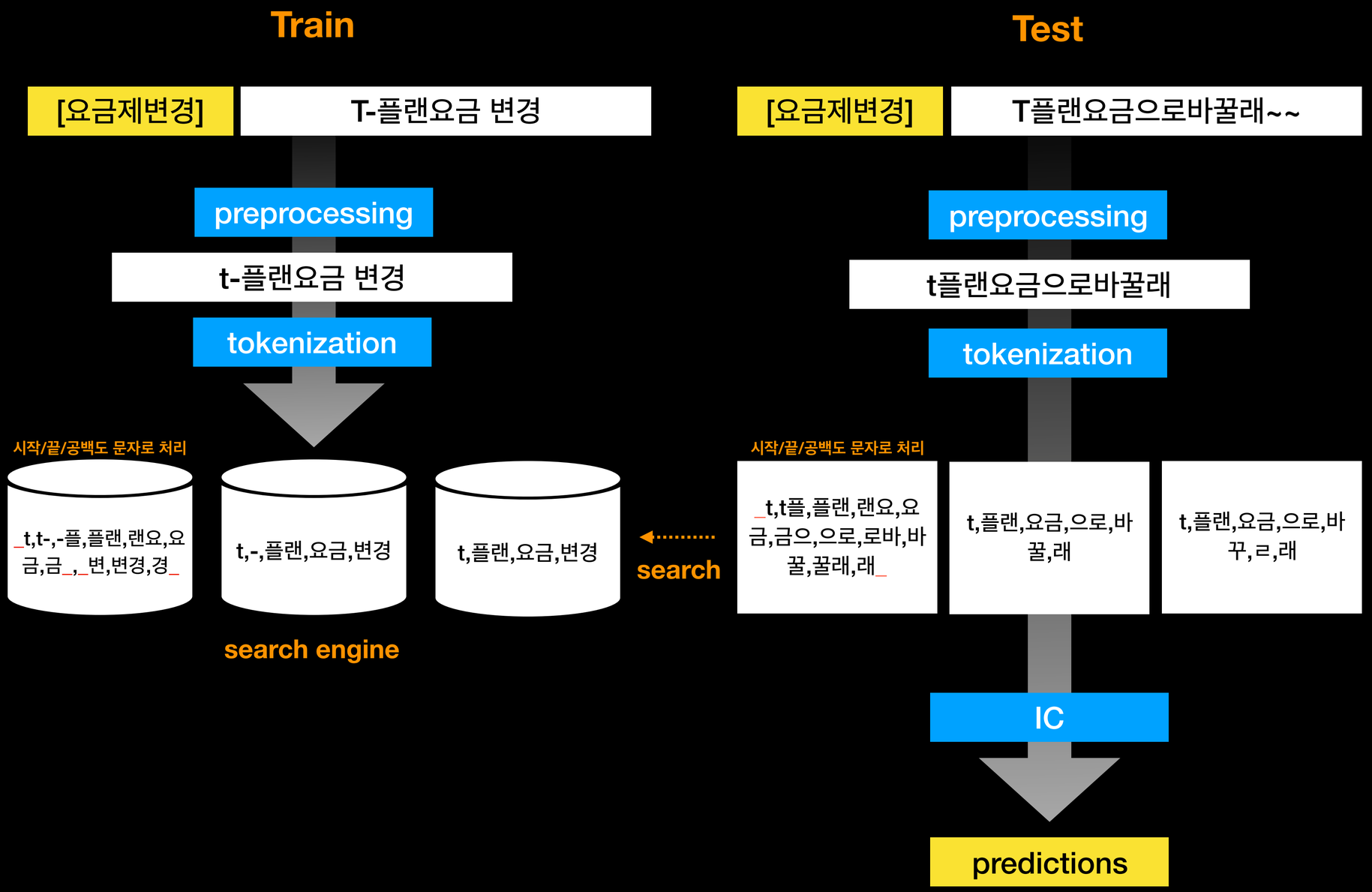
시작/끝/공백도 문자로 처리

_t,t플,플랜,랜요,요금,금으,으로,로바,바꿀,꿀래,래_

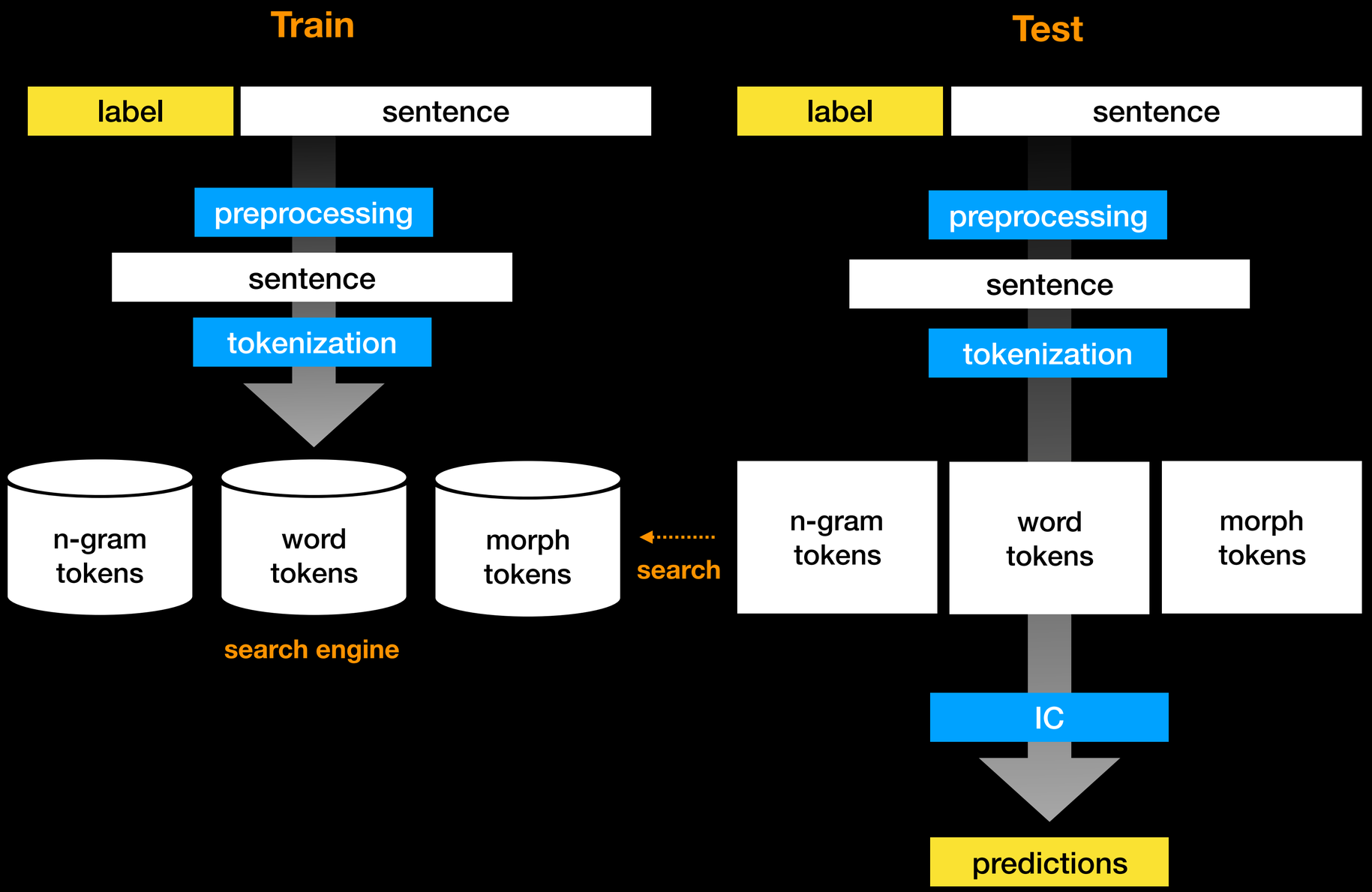
n-gram tokens

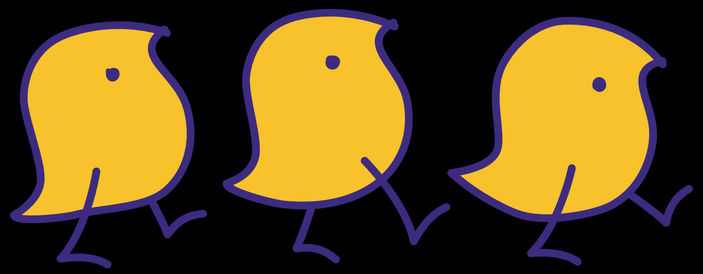
Intent Classification

DIY 챗봇



Intent Classification



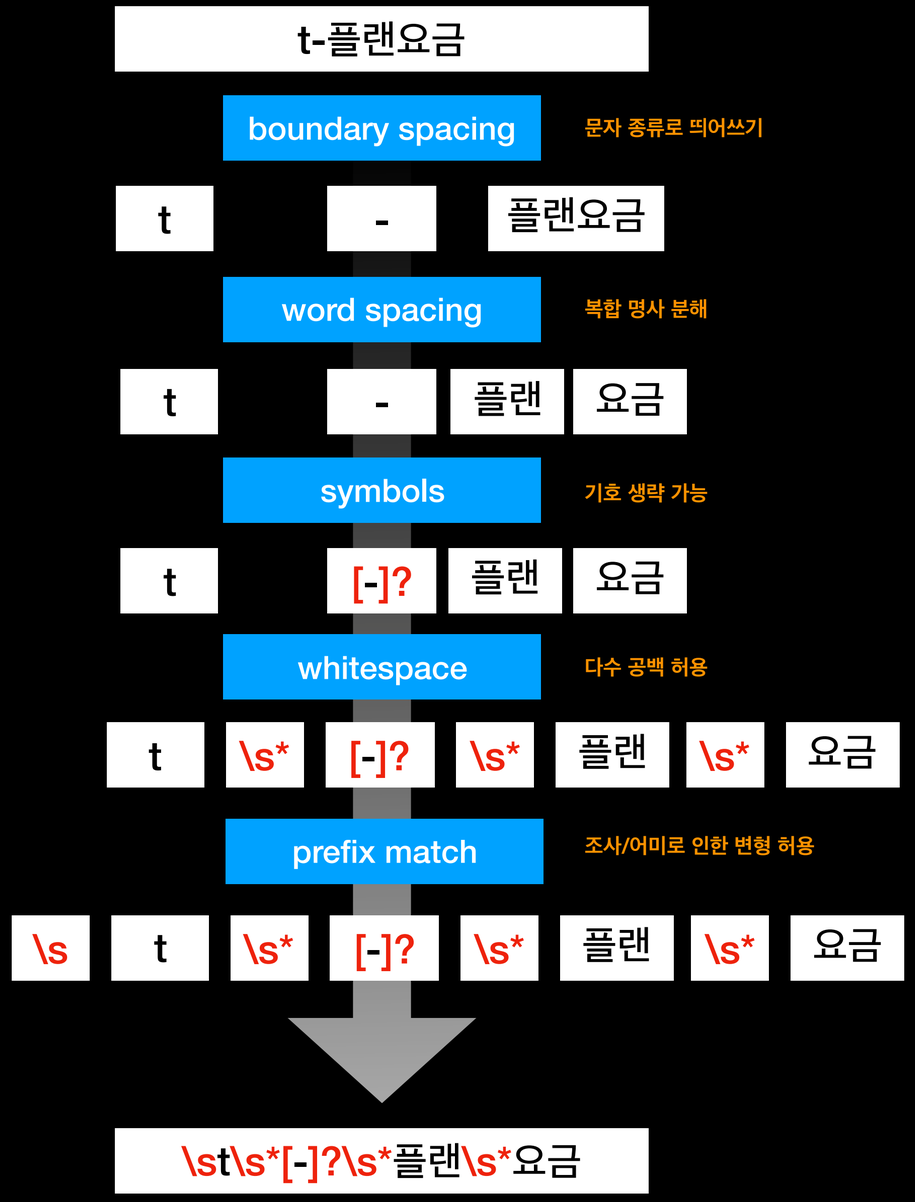


NER

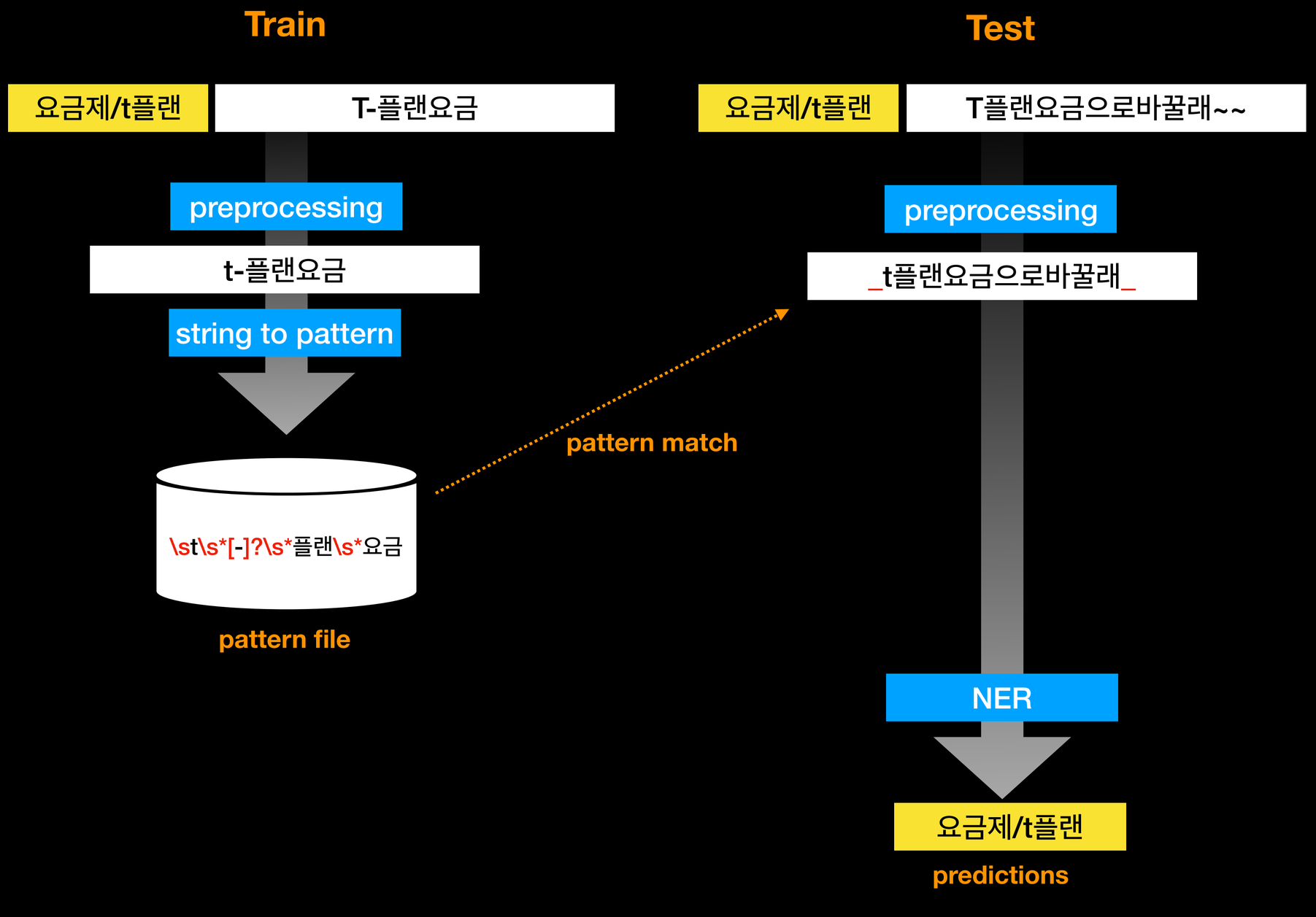
Named Entity Recognition

LangCon 2019

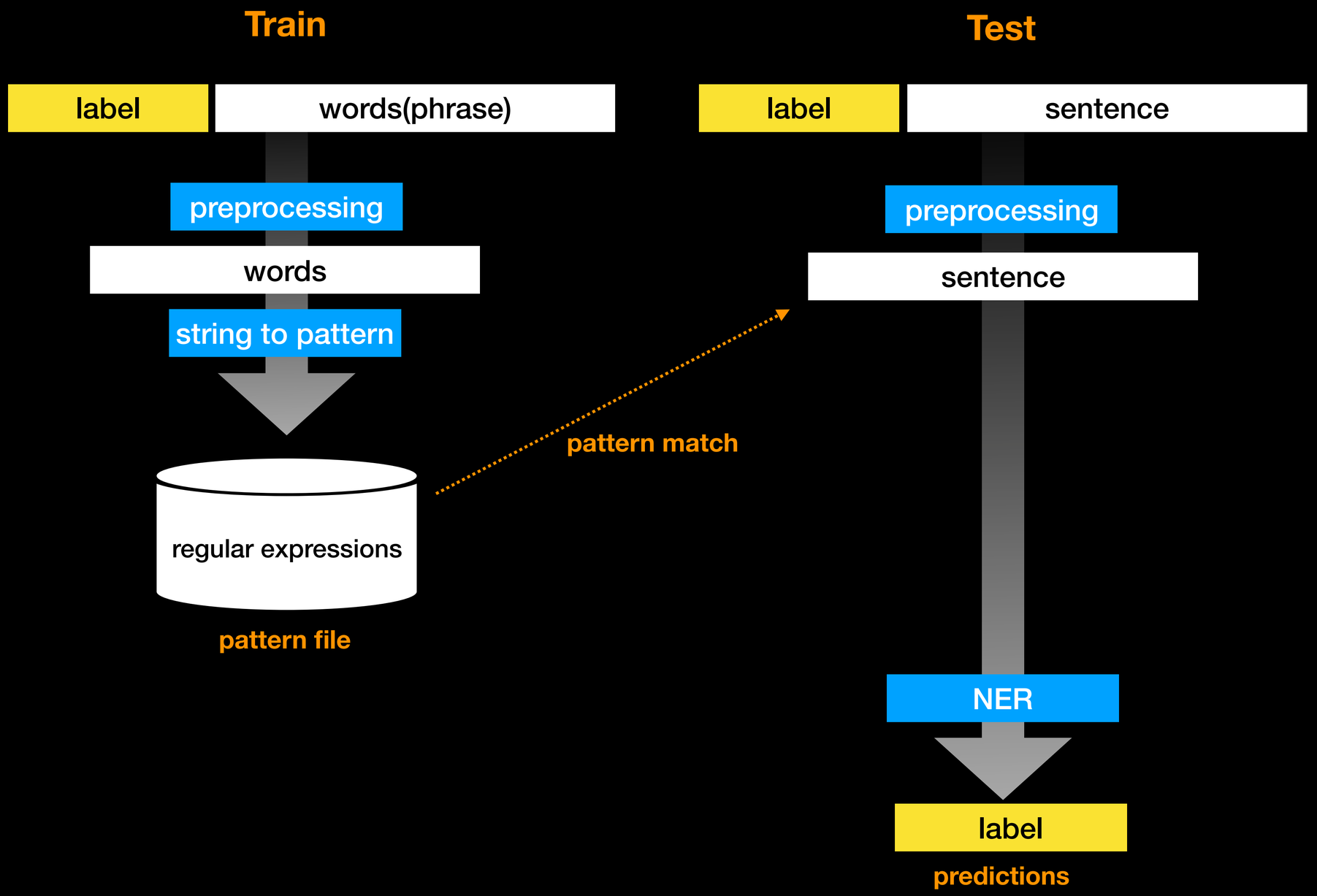
String to Pattern

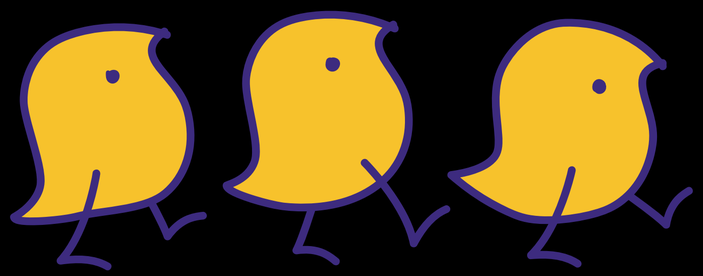


NER


Train
요금제/t플랜
T-플랜요금
preprocessing
t-플랜요금
string to pattern
\st\s*[-]?\s*플랜\s*요금
pattern file
pattern match
Test
요금제/t플랜
T플랜요금으로바꿀래~~
preprocessing
t플랜요금으로바꿀래
NER
요금제/t플랜
predictions

NER


Train
label
words(phrase)
preprocessing
words
string to pattern
regular expressions
pattern file
pattern match
Test
label
sentence
preprocessing
sentence
NER
label
predictions



One more thing... v0.2

LangCon 2019

Overcome Spelling Errors - IC

t플랜요금

_t,t플,플랜,랜요,요금,금_

n-gram
tokens

v0.1

t플랜요금

tㅍㅡㄹㄹㅐㄴㅇㅛㄱㅡㅁ

_t,tㅍ,ㅍㅡ,ㅡㄹ,ㄹㄹ,ㄹㅐ,ㅐㄴ,
ㄴㅇ,ㅇㅛ,ㅛㄱ,ㄱㅡ,ㅡㅁ,ㅁ_

n-gram
tokens

v0.2

Overcome Spelling Errors - IC

DIY 챗봇

t플렌요금

user input

t플렌요금

character based
67%

Similiarity

alphabet based
85%

철자 오류 허용 안함

_t,t플,플렌,렌요,요금,금_

=4/6

n-gram
tokens

v0.1

철자 오류 허용함

_t,tㅍ,ㅍㅡ,ㅡㄹ,ㄹㄹ,ㄹㅐ,ㅐㄴ,
ㄴㅇ,ㅇㅍ,ㅍㄱ,ㄱㅡ,ㅡㅁ,ㅁ_

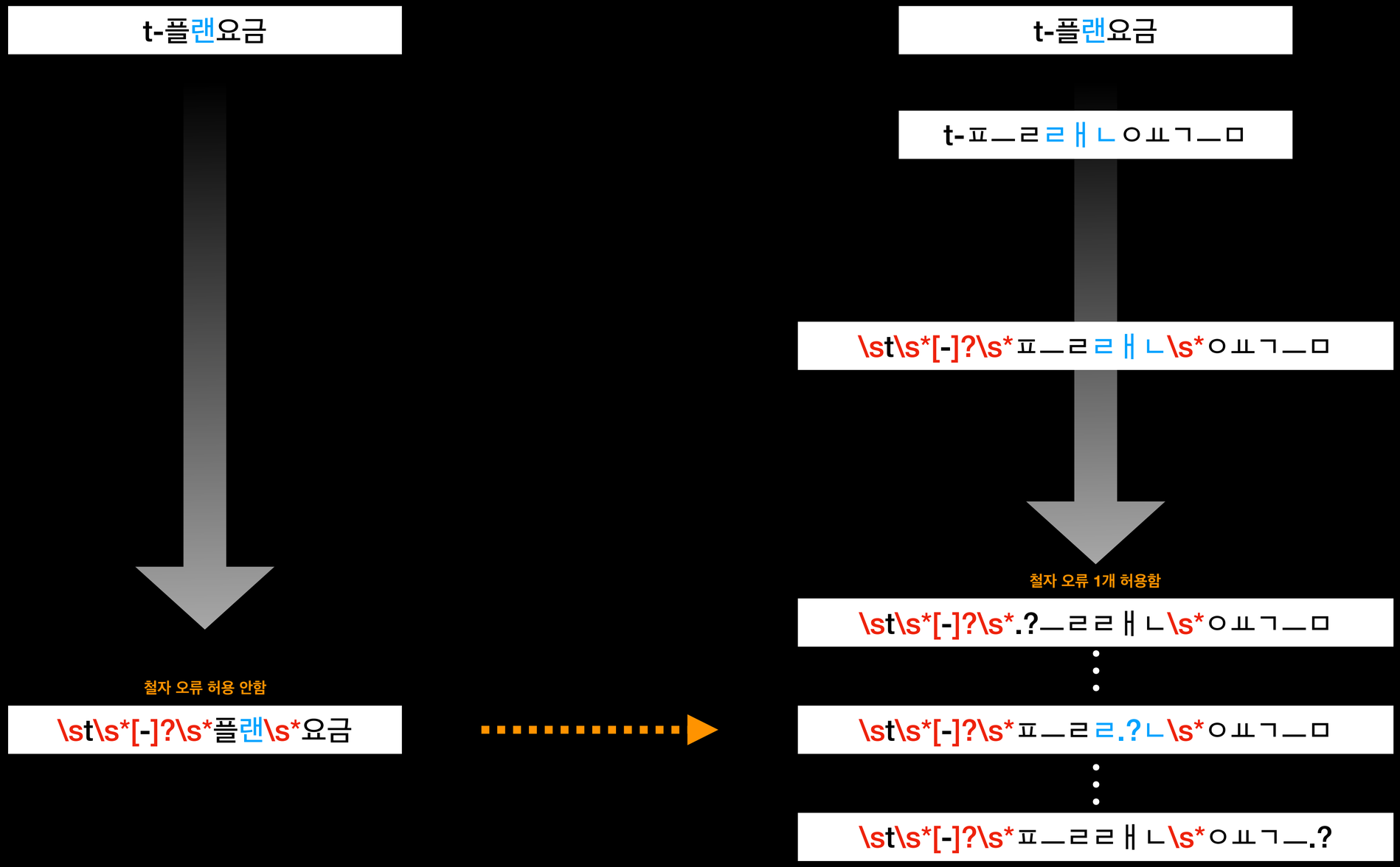
=11/13

n-gram
tokens

v0.2

Overcome Spelling Errors - NER

DIY 챗봇



Overcome Spelling Errors - NER

t-플렌요금

user input

t-플렌요금

failed

pattern matching

success

철자 오류 허용 안함

\st\s*[-]?\s*플랜\s*요금

철자 오류 1개 허용함

\st\s*[-]?\s*ㅍㅡㄹㄹ.?ㄴ\s*ㅇㅛㄱㅡㅁ

v0.1

v0.2

Overcome Typing Errors - NER

DIY 챗봇

t-플랜요금

t-ㅍㅡㄹㄹㅐㄴㅇㅛㄱㅡㅁ

오타 범위

철자 오류 1개 허용함

\st\s*[-]?\s*ㅍㅡㄹㄹ.?ㄴ\s*ㅇㅛㄱㅡㅁ

v0.2

t-플랜요금

t-ㅍㅡㄹㄹㅐㄴㅇㅛㄱㅡㅁ

오타만 허용함

\st\s*[-]?\s*ㅍㅡㄹㄹ[ㅐㅑ\(\)ㅔㅏㅣ]?ㄴ\s*ㅇㅛㄱㅡㅁ

v0.21