

기업 맞춤형 생성형 AI 플랫폼

S-AIP

AGENDA

14:00 ~ 14:20 (20')

S2W의 AI 사업 비전
서상덕 대표

14:20 ~ 14:50 (30')

S·AIP 개요
정진우 이사/공학박사

14:50 ~ 15:00 (10')

Break Time

15:00 ~ 15:30 (30')

기업의 S·AIP 활용 방안
서현민 이사

15:30 ~ 15:50 (20')

Q&A

Part 1

Dark Web & Dark BERT & Dark Chat

S·AIP 개발 배경: 다크웹 특화 AI

표면웹

우리가 매일 사용하는 뉴스 포털, 구글 검색 등

구글 크롬이나 MS엣지 같은 일반 브라우저로 접근할 수 있는 사이버 공간



딥다크웹

일반적인 방법으로는 접근이 어려운
사용자의 익명성, 신원, 위치 등을
특정 브라우저나 접근 기술을 통해

무기 판매

Private Builder Ransomware V2

해킹정보 공유

마약 거래

데이터 유출

범죄 콘텐츠 공유

Welcome to Bitcoin Laundry
범죄자금 세탁
CLEAN YOUR BTC NOW

Anonymous 21/01/23(Sat)11:29:57 No.1171
n번방 어디서 보냐 알려줘

Anonymous 21/01/23(Sat)11:47:04 No.1173
모네로 받고 램

Anonymous 21/01/23(Sat)11:49:19 No.1175
얼마정도?

Anonymous 21/01/23(Sat)12:41:10 No.1177
병신 다크월에 견찰을 갈라서 n번방거래하면 추적당해 ㅋ

Anonymous 21/01/23(Sat)12:44:54 No.1178
모네로 받고 판다는세끼는 강 무시하고 교환으로 받던가 텔레가서 좇목해라

Anonymous 21/01/23(Sat)12:51:17 No.1181
니 세션 아이디 남겨라

S·AIP의 시작? 다크웹 AI

 **S2W** 's Darkweb DATA

 &
인공지능 AI



KAIST 전산학부 석사 졸업

KAIST 전산학부 박사 졸업

삼성전자 종합기술원 머신러닝랩



마주한 다크웹



다크웹 특화 AI



음란물 자동 필터링

다크웹 범죄 카테고리 자동 분류

해킹포럼 게시물 자동 추천

마약 포럼 내 신조어 탐지

⋮

다크웹 특화 AI



중요한 건,

음란물 자동 필터링

"도메인"에 대한 이해!

⋮

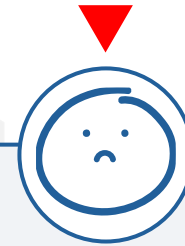
S·AIP 의 도메인 학습 예시

지식그래프 기반 식별자 관계도 파악

Q. 두 작성자가 동일한 인물일까 ?



다크웹 ID: shadowhacker




다크웹 ID: sikinawnaw

Vietnam Government Systems Network access
by shadowhacker - Monday January 30, 2023 at 08:02 PM

9 minutes ago

shadowhacker



<https://t.me/shadowhacker>

Hello i am selling a vietnamese governi and got following result, (network.txt fi price of internal network systems acce telegram me : <https://t.me/shadowhacker>

Attached Files

network.txt (Size: 4.94 KB / Downloads

Visit <https://t.me/shadowleakss>, for HQ top F

microsoft.com subdomain

sikinawnaw · ⌚ Сегодня в 20:13

Сегодня в 20:13

hello i am selling shell access into microsoft.com subdomain subdomain, price is 700\$, not negotiable

you can upload files on the subdomain etc.
normal user not root

not providing more info (dont ask anything, i wont risk my access to answer you about anything)

middle man ok

if want it, dm me on my tele
[@shadowhacker](https://t.me/shadowhacker)

NO AVATAR

sikinawnaw
HDD-drive

Пользователь

Регистрация : 19.11.2022

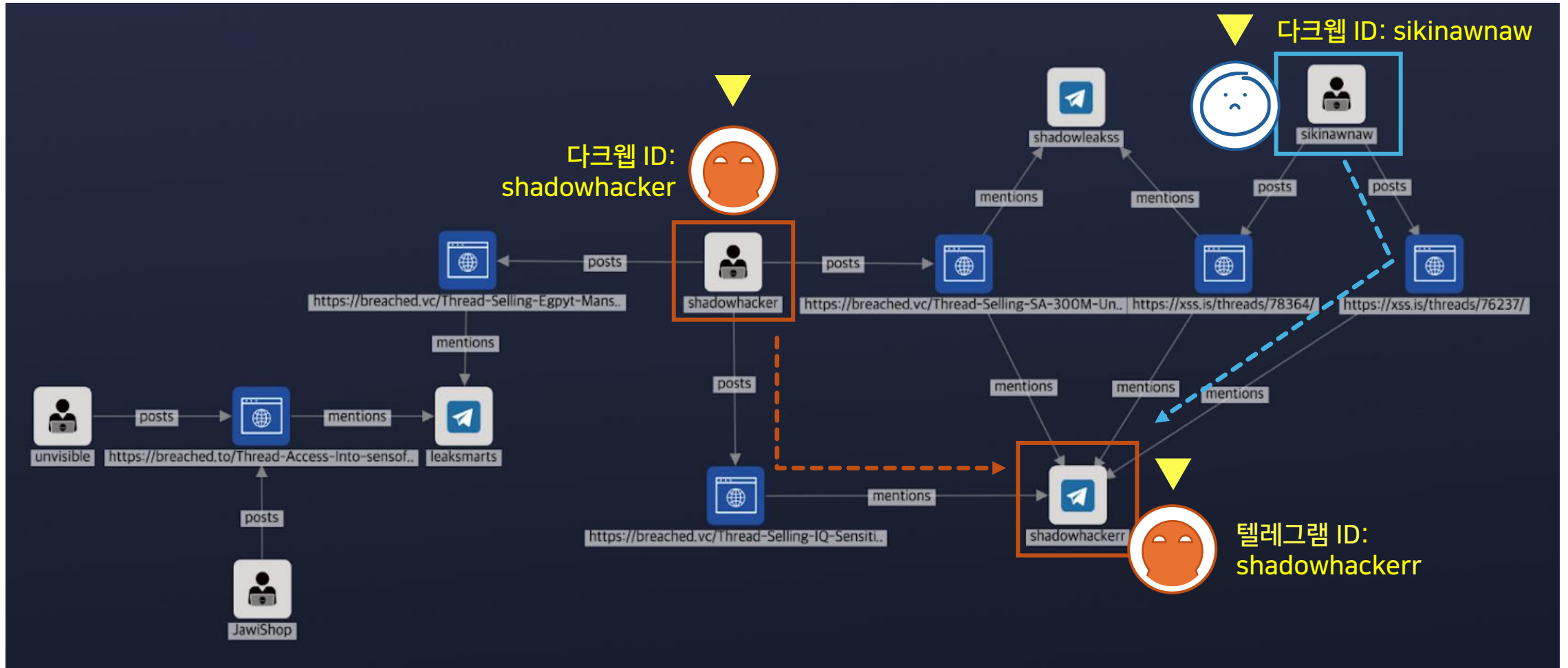
Сообщения : 45

Реакции : 2

S·AIP 의 도메인 학습 예시


지식그래프 기반 식별자 관계도 파악

Q. 두 작성자가 동일한 인물일까 ?



S·AIP 의 도메인 학습 예시

지식그래프 기반 식별자 관계도 파악



Vietnam Government Systems Network access
by shadowhacker · Monday January 30, 2023 at 08:02 PM

작성자: shadowhacker

9 minutes ago

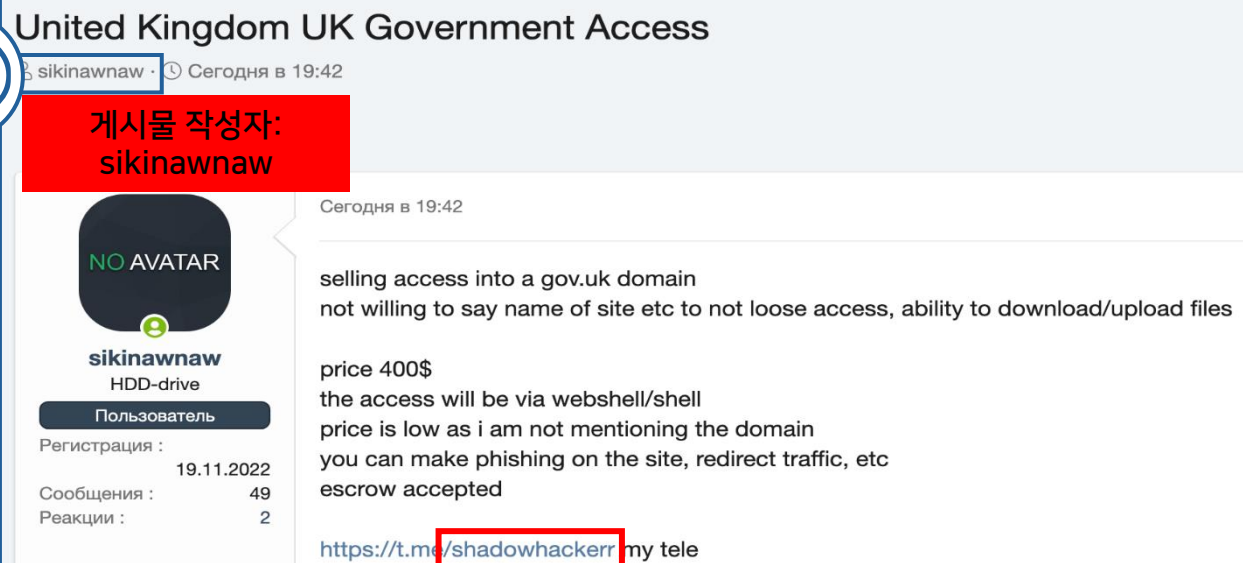
Hello i am selling a vietnamese government server, gov.vn that have over 5 g and got following result, (network.txt file) price of internal network systems access is 300\$ telegram me : <https://t.me/shadowhackerr>

Attached Files

network.txt (Size: 4.94 KB / Downloads: 2)

<https://t.me/shadowhacker>

Visit <https://t.me/shadowleakss>, for HQ top Private Free dbs, government logins.



United Kingdom UK Government Access
by sikanawna · Сегодня в 19:42

게시물 작성자: sikanawna

Сегодня в 19:42

selling access into a gov.uk domain
not willing to say name of site etc to not loose access, ability to download/upload files

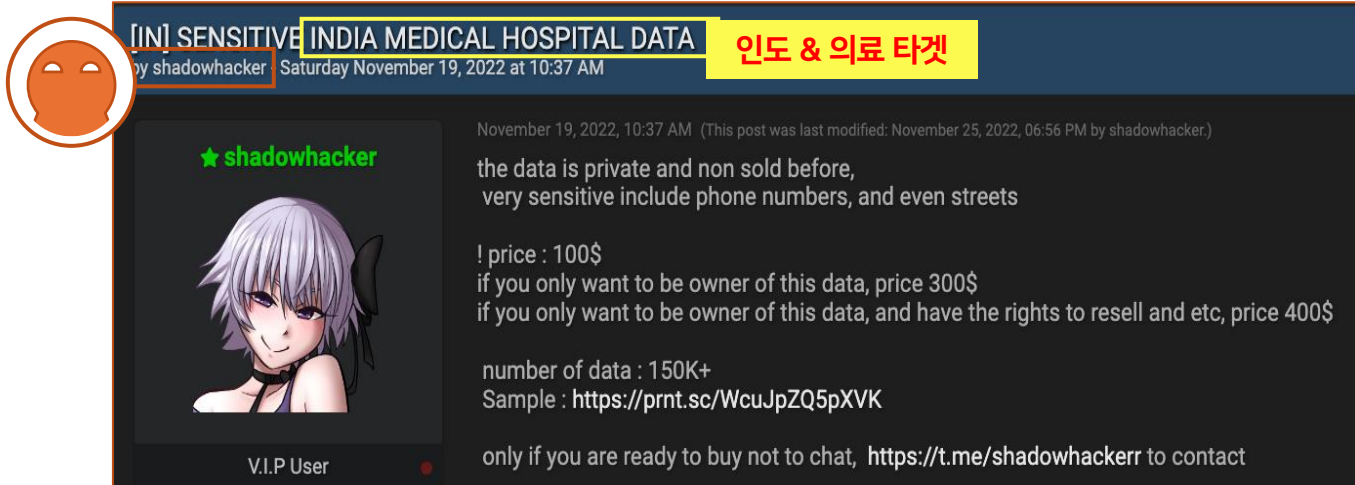
price 400\$
the access will be via webshell/shell
price is low as i am not mentioning the domain
you can make phishing on the site, redirect traffic, etc
escrow accepted

<https://t.me/shadowhackerr> my tele



S-AIP 의 도메인 학습 예시

지식그래프 기반 식별자 관계도 파악



[IN] SENSITIVE INDIA MEDICAL HOSPITAL DATA 인도 & 의료 타겟
by shadowhacker Saturday November 19, 2022 at 10:37 AM

November 19, 2022, 10:37 AM (This post was last modified: November 25, 2022, 06:56 PM by shadowhacker.)

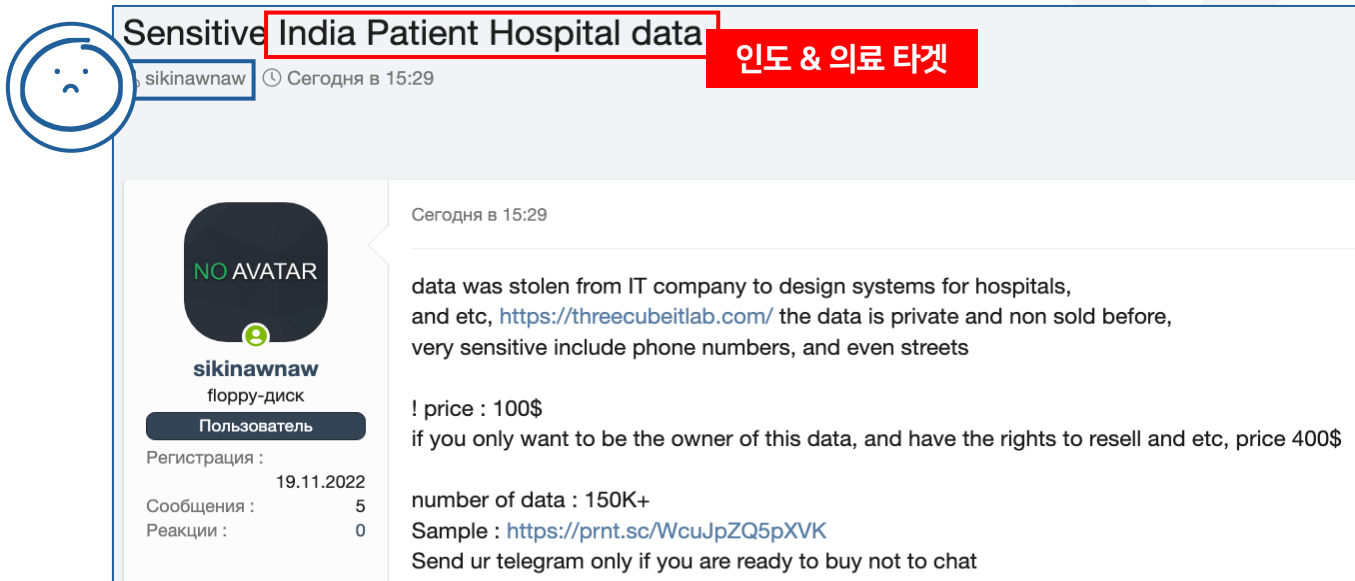
the data is private and non sold before, very sensitive include phone numbers, and even streets

! price : 100\$
if you only want to be owner of this data, price 300\$
if you only want to be owner of this data, and have the rights to resell and etc, price 400\$

number of data : 150K+
Sample : <https://prnt.sc/WcuJpZQ5pXVK>

only if you are ready to buy not to chat, <https://t.me/shadowhackerr> to contact

V.I.P User



Sensitive India Patient Hospital data 인도 & 의료 타겟
sikinawnaw Сегодня в 15:29

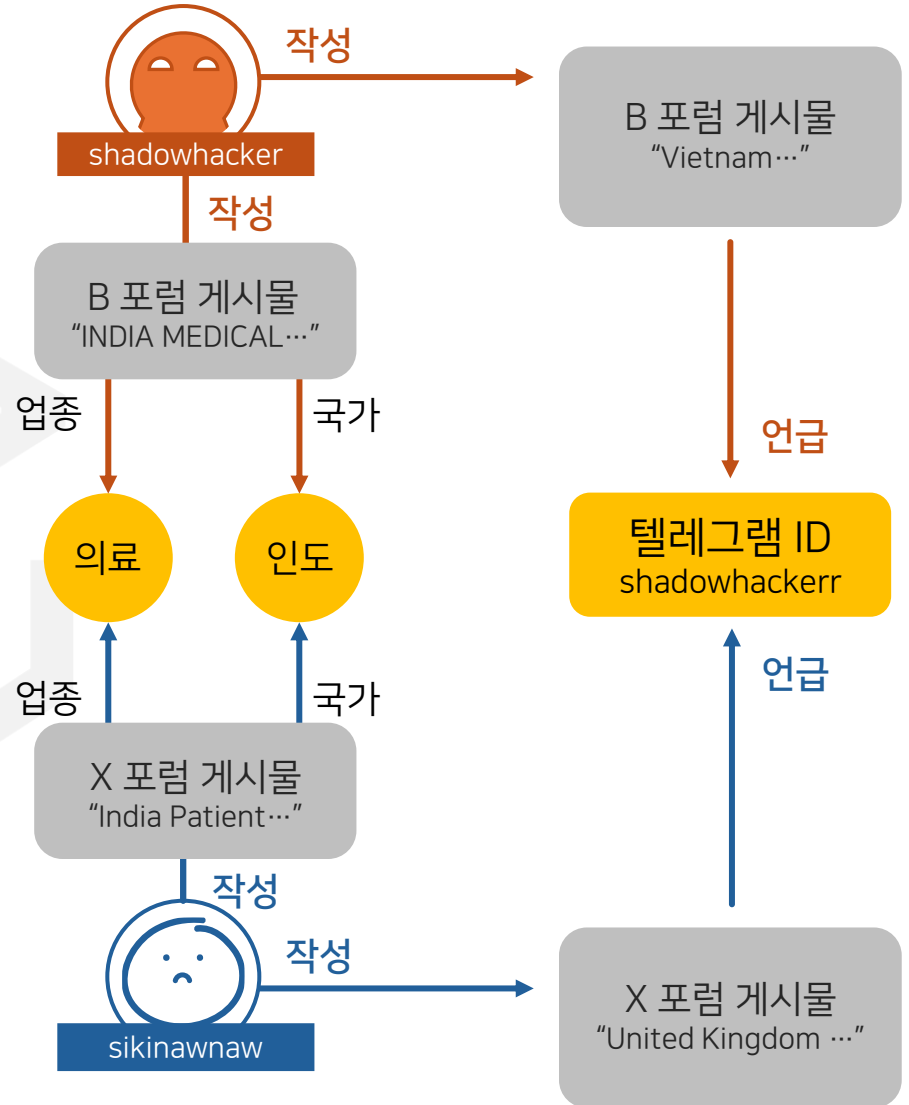
Сегодня в 15:29

data was stolen from IT company to design systems for hospitals, and etc, <https://threecubeitlab.com/> the data is private and non sold before, very sensitive include phone numbers, and even streets

! price : 100\$
if you only want to be the owner of this data, and have the rights to resell and etc, price 400\$


number of data : 150K+
Sample : <https://prnt.sc/WcuJpZQ5pXVK>

Send ur telegram only if you are ready to buy not to chat




S-AIP 의 도메인 학습 예시

지식그래프 기반 식별자 관계도 파악

 [IN] SENSITIVE INDIA MEDICAL HOSPITAL DATA 인도 & 의료 타겟
by shadowhacker Saturday November 19, 2022 at 10:37 AM

November 19, 2022, 10:37 AM (This post was last modified: November 25, 2022, 06:56 PM by shadowhacker.)

★ shadowhacker




V.I.P User

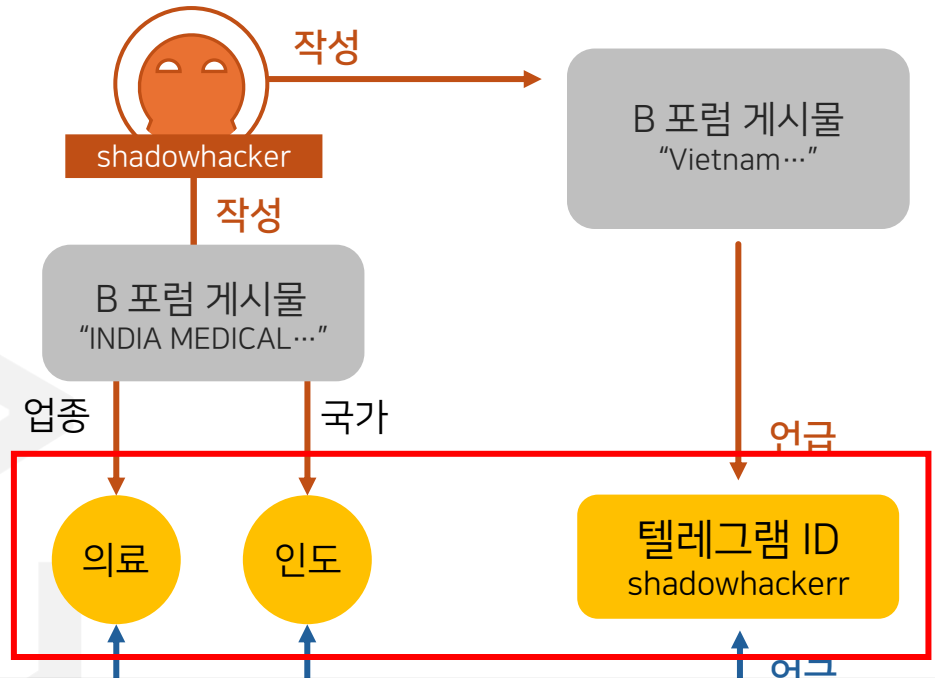
the data is private and non sold before, very sensitive include phone numbers, and even streets

! price : 100\$
if you only want to be owner of this data, price 300\$
if you only want to be owner of this data, and have the rights to resell and etc, price 400\$

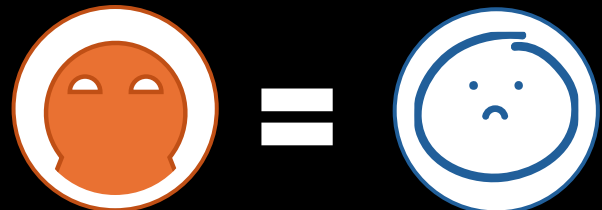
number of data : 150K+
Sample : <https://prnt.sc/WcuJpZQ5pXVK>

only if you are ready to buy not to chat, <https://t.me/shadowhackerr> to contact

 Sensitive India Patient Hospital data 인도 & 의료 타겟
sikinawhaw ⌚ Сегодня в 15:29



동일 인물
가능성이 매우 높음

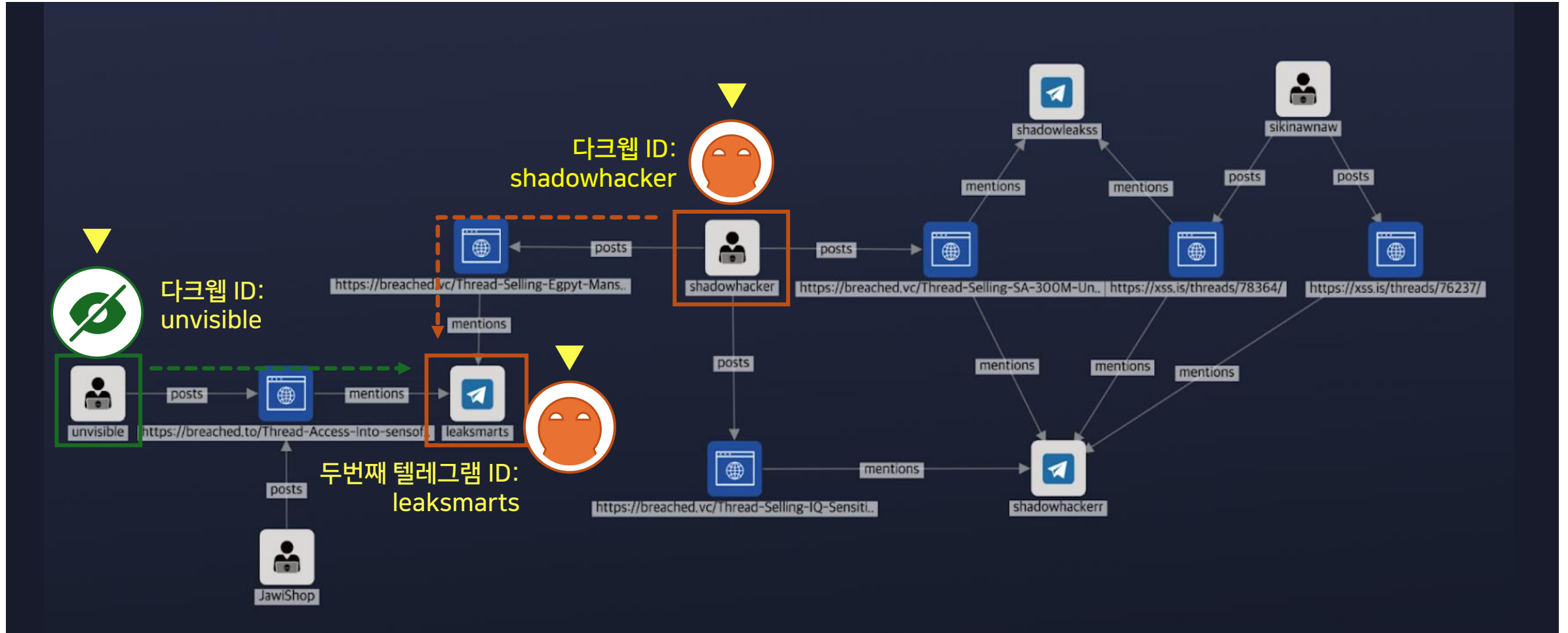


1. 동일 텔레그램 ID
2. 동일한 타겟 업권
3. 동일한 타겟 국가

S·AIP 의 도메인 학습 예시


지식그래프 기반 식별자 관계도 파악

Q. 그럼, 동일한 텔레그램 ID를 공유한다고 반드시 동일한 사용자일까?



S-AIP 의 도메인 학습 예시


지식그래프 기반 식별자 관계도 파악

 [FREE] India Students Database (인도 & 교육 타겟) by shadowhacker Thursday August 18, 2022 at 07:57 PM

shadowhacker August 18, 2022, 07:57 PM
Hello wont spam and talk alot simple as the title is
Leak is the last post on <https://t.me/leaksmarts>
Bye

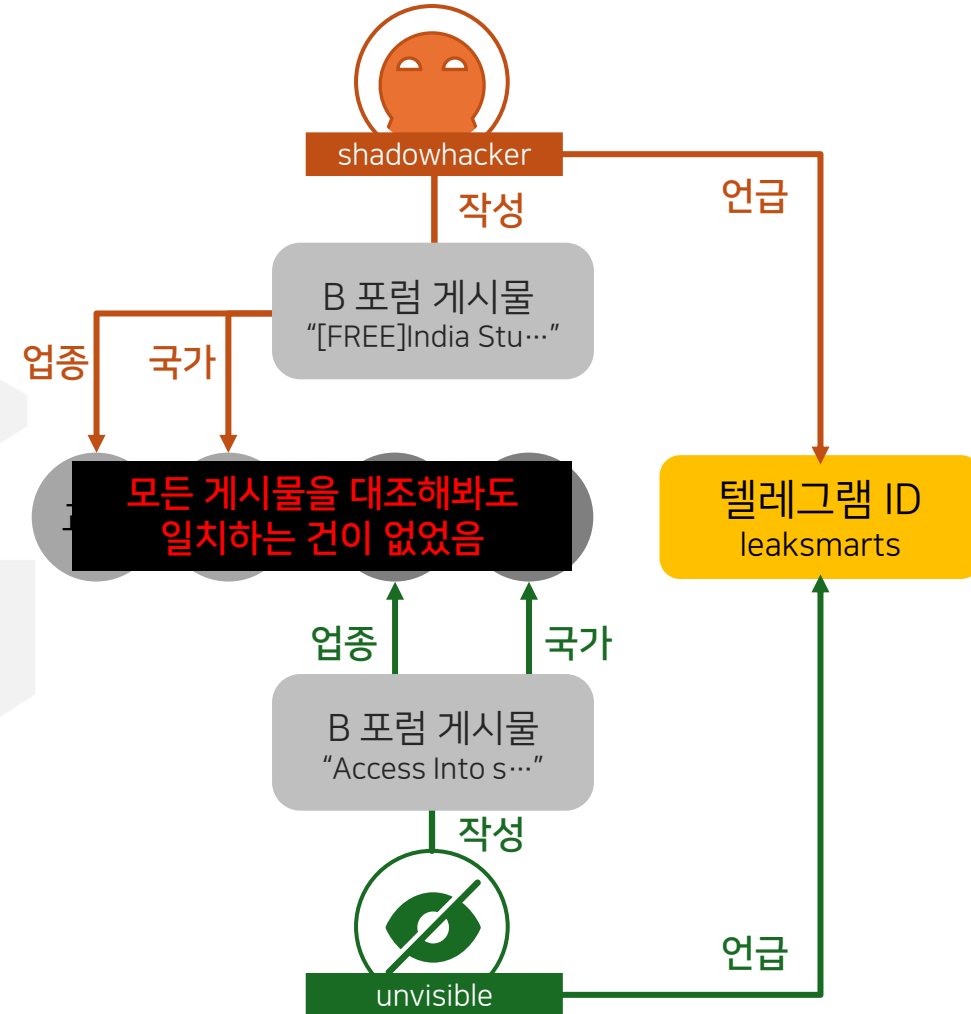
<https://t.me/shadowhacker>

텔레그램 ID

 Access Into sensoft Company [free :D] by invisible Sunday August 14, 2022 at 08:25 AM

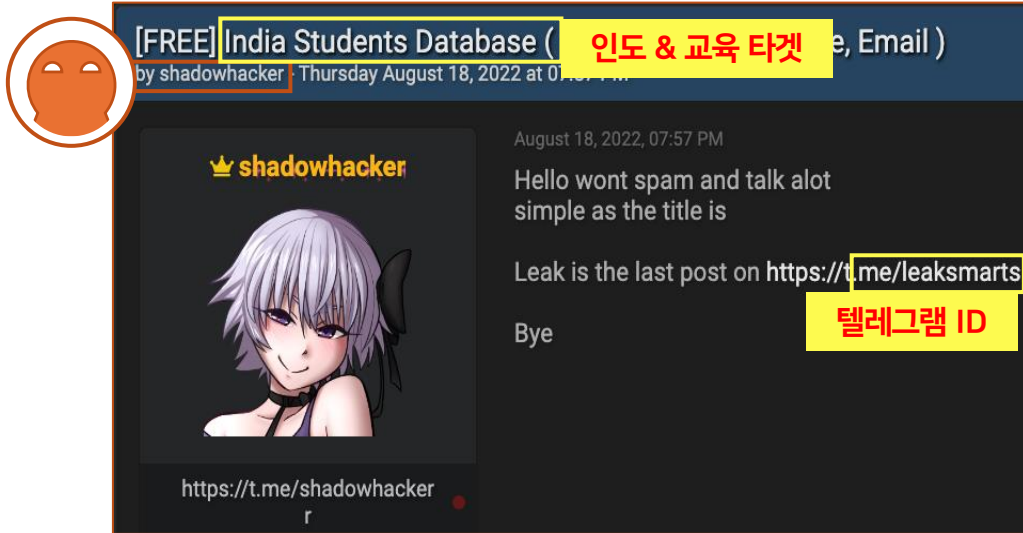
invisible August 14, 2022, 08:25 AM
enjoy babes
<https://wso2.sensoft-next.com:9443//auth...cmd=whoami>
on the date of this post, its working and fresh group responsible and credits for : <https://t.me/leaksmarts>

텔레그램 ID



S·AIP 의 도메인 학습 예시

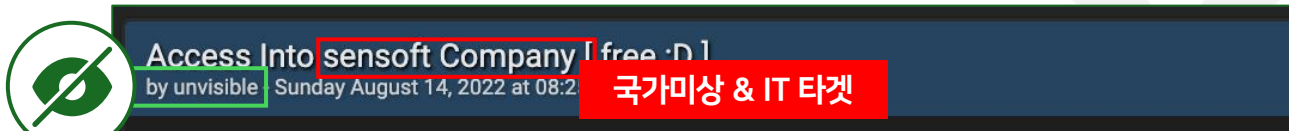
지식그래프 기반 식별자 관계도 파악



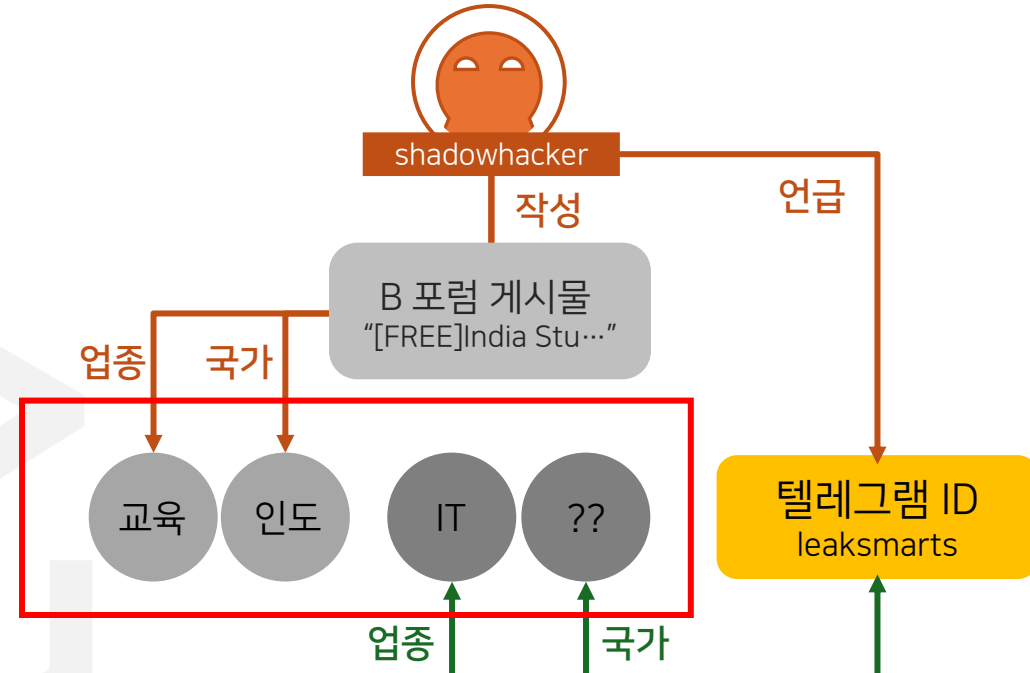
[FREE] India Students Database (인도 & 교육 타겟) by shadowhacker Thursday August 18, 2022 at 07:57 PM

shadowhacker August 18, 2022, 07:57 PM
Hello wont spam and talk alot simple as the title is
Leak is the last post on <https://t.me/leaksmarts> (텔레그램 ID)
Bye

<https://t.me/shadowhacker>



Access Into sensoft Company [free :D] by unvisible Sunday August 14, 2022 at 08:20 (국가미상 & IT 타겟)



동일인물 가능성이 상대적으로 낮음



?

=



동일한 텔레그램 ID but 일치하지 않는 타겟

온톨로지와 지식그래프

1

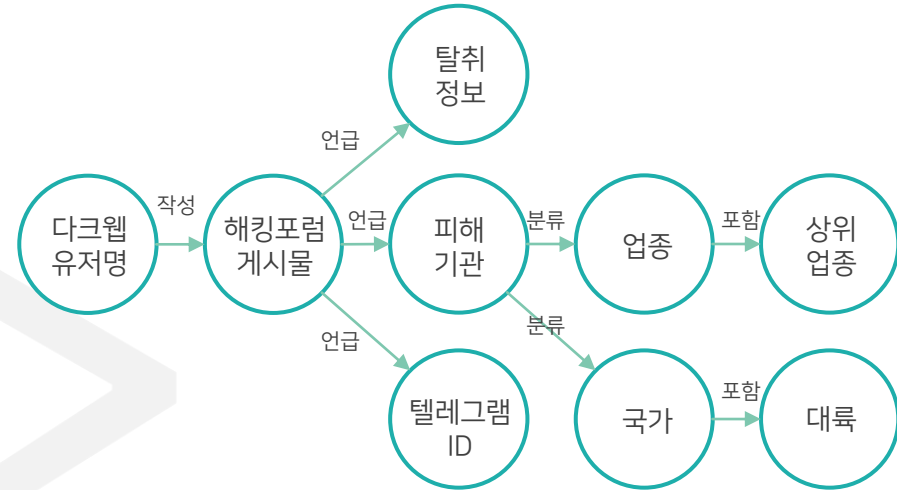
유저 프로파일링 지식그래프를
자동으로 구축하기 위해 알아야 했던 도메인 지식

1. 다크웹에서 범죄자는 해킹 포럼에 흔적을 남긴다.
2. 동일한 범죄자가 여러 해킹 포럼에 걸쳐 서로 다른 ID를 사용하여 활동하기도 한다.
3. 해킹 포럼 게시물에는 탈취한 정보의 종류와 규모, 피해 기관의 국가/업종, 그리고 거래를 위한 연락처가 언급되어 있다.
4. 범죄자마다 특정 국가/대륙/특정 업종의 기관을 노리는 경우가 많다.
5. 범죄자들은 텔레그램으로 자주 소통한다.
6. 실제 거래도 텔레그램에서 주로 이루어진다.
7. 거래를 위한 텔레그램 방에는 1:1 대화방뿐 아니라, 단체방도 있다.
- ...



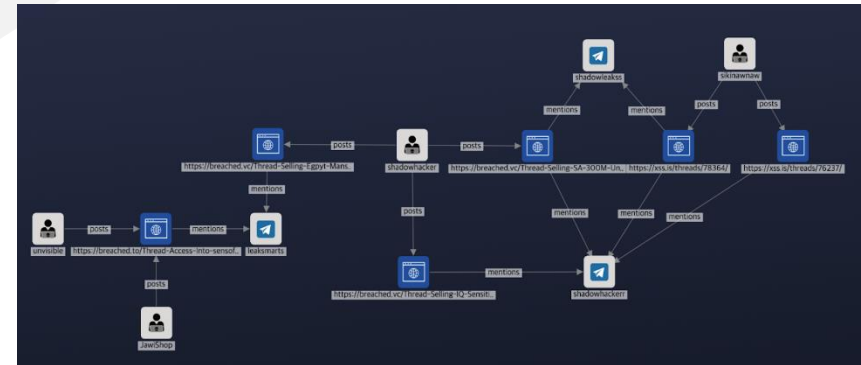
2

도메인 지식을 기반으로 **온톨로지 생성**



3

온톨로지를 기반으로 **지식그래프 생성**



도메인 특화 AI의 핵심

도메인 = 분야 ≙ **세계관**

세계관은 **온톨로지**로 표현

따라서,

AI에게 온톨로지를 통해 세계관을 알려주면
더 **전문화**된 AI가 탄생한다.



큰 세계관 = 마블



큰 세계관 = 다크웹



큰 세계관 = 마블

큰 세계관 = 다크웹



캡틴 헐크 아이언맨

퍼스트어벤저 세계관	헐크 세계관	아이언맨 세계관
캡틴 아메리카 세계관	쉬헐크 세계관	어벤저스 세계관
시빌워 세계관	어벤저스 세계관	시빌워 세계관



해킹 마약

데이터유출 세계관	마약거래 세계관
취약점 세계관	마약제조 세계관
랜섬웨어 세계관	자금세탁 세계관



큰 세계관 = 마블

큰 세계관 = 다크웹

제조 세계관



금융 세계관



바이오 세계관



이처럼 수많은 도메인이 존재한다.

마블 세계관을 온톨로지/지식그래프로 기술하려면?

NEW Logo

TIM WHEN ARE YOU GOING TO ADMIT that you LEAKED the best FOOTAGE

DEADPOOL (CONT'D)
(turns)
And I hope you're watching. (exposed SELF)
Blood-stained pavement and a katana shard mark the spot the prisoner was a few seconds ago. He's GONE. Deadpool turns back to Colossus with a LOOK that could KILL.

DEADPOOL (CONT'D)
That does it.

MEGASONIC TEENAGE WARHEAD
(fingers to temple)
This is gonna get ugly.

DEADPOOL BUM-RUSHES Colossus and POUNDS him across the face with a BRISTAL fist... CLANG... BREAKING HIS OWN WRIST.

Colossus SIGHS. Deadpool CIRCLES him, then fires an UPPERCUT with his OTHER FIST... BANG... BREAKING that wrist, too.

Gah.

DEADPOOL (CONT'D)
Deadpool holds up two limp wrists like an effeminate boxer.

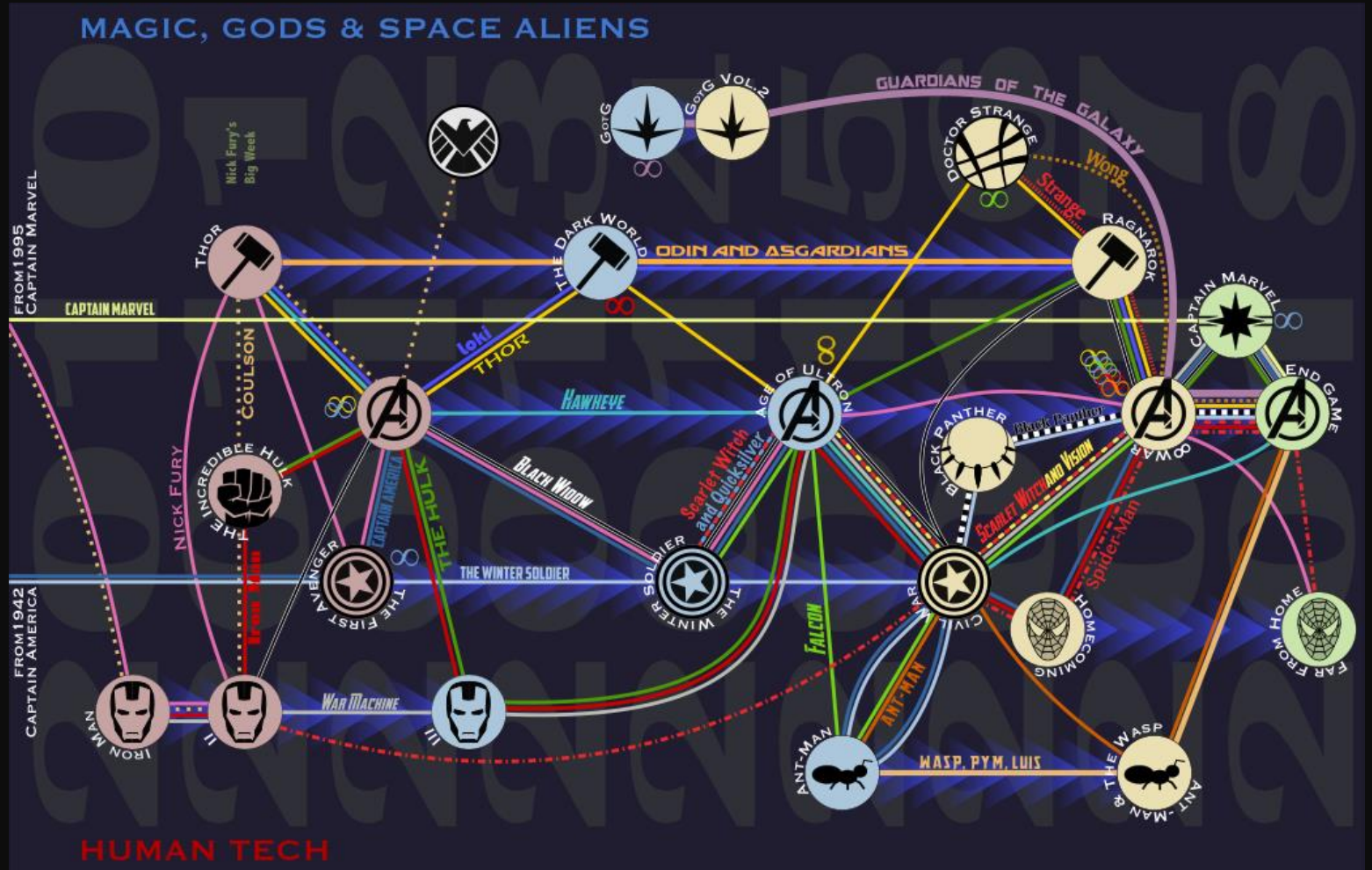
DEADPOOL (CONT'D)
(re: useless wrists)
All the dinosaurs feared the T-Rex AND HERPES

Deadpool throws a ROUNDHOUSE KICK at Colossus. CRACK! HE BREAKS HIS OWN ANKLE. IT'S suddenly BENT at a WILD ANGLE.

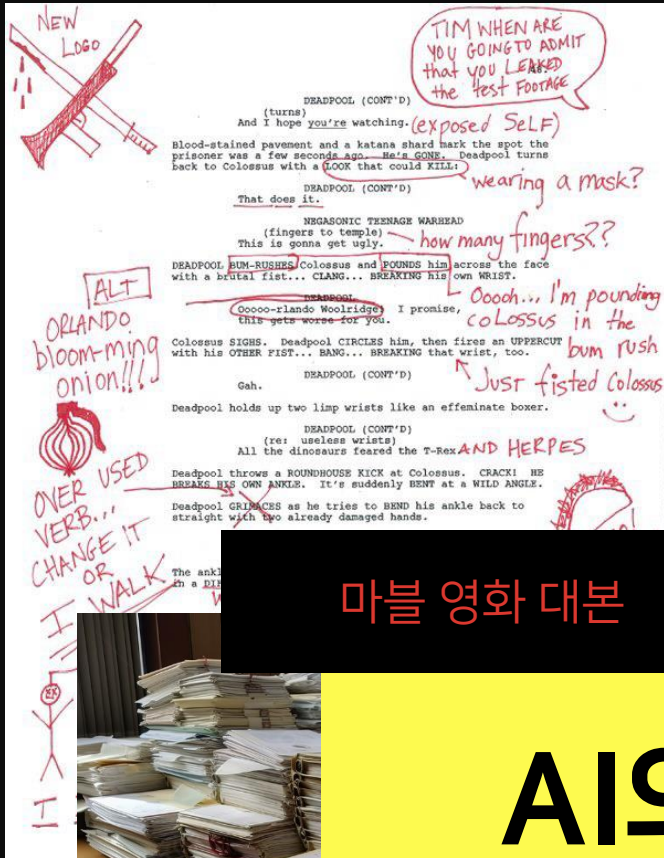
Deadpool GRIMACES as he tries to BEND his ankle back to straight with two already damaged hands.

Handwritten notes:
wearing a mask?
how many fingers??
Cooch... I'm pounding Colossus in the bum rush
Just fist-ed Colossus
ORLANDO Bloom-ing onion!!!
OVER USED VERB... CHANGE IT OR I WALK

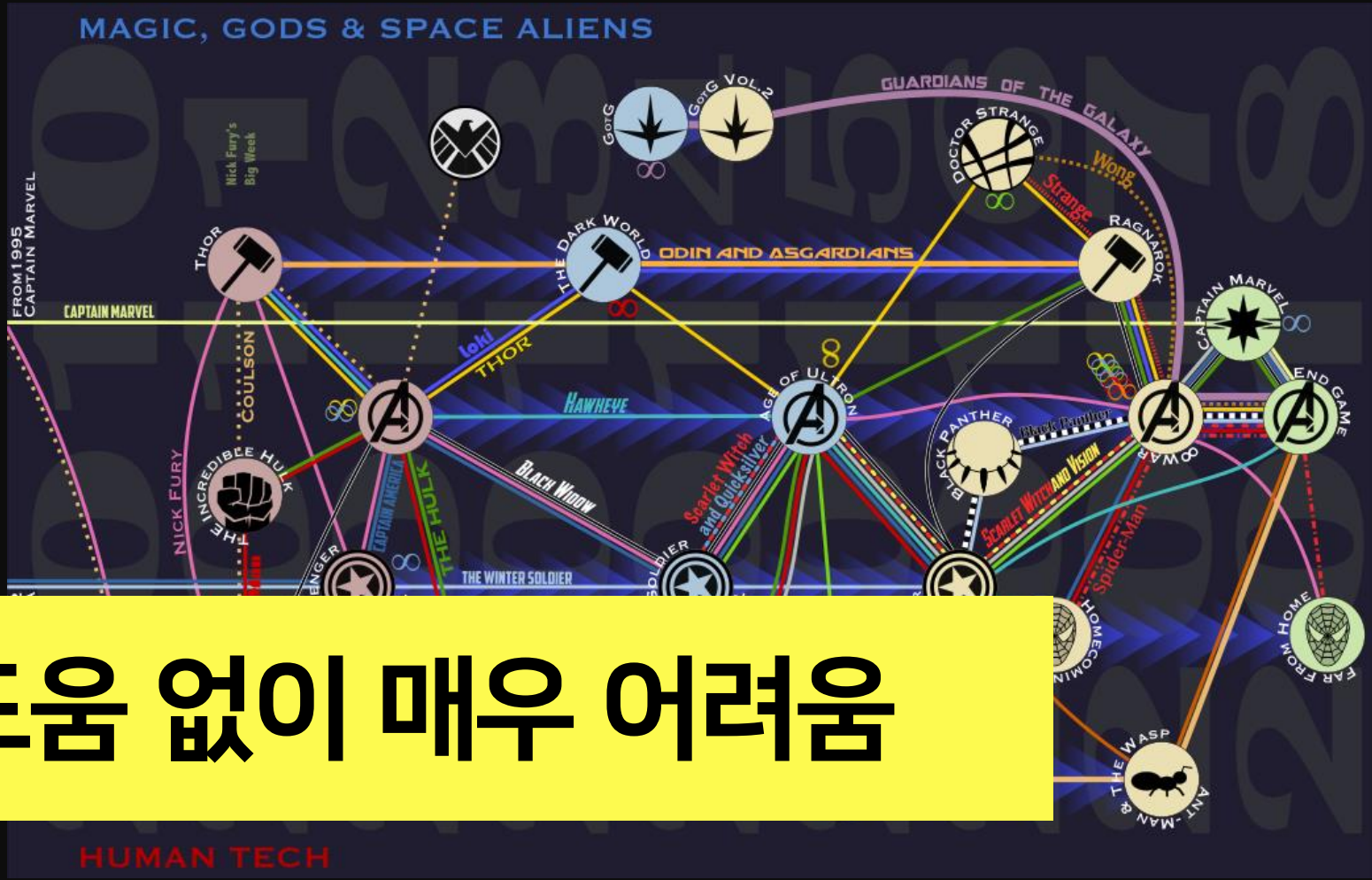
마블 영화 대본



마블 세계관을 온톨로지/지식그래프로 기술하려면?



마블 영화 대본



AI의 도움 없이 매우 어려움

온톨로지와 지식그래프를 제대로 구축, 활용하려면

도메인 특화 AI 전문성이 반드시 필요

**S2W는 다크웹 특화 AI 개발 경험을 통한
도메인 특화 AI 전문성 보유**



S2W

KAIST

공동연구



NAACL 2022

다크웹 범죄 유형 자동 탐지 및 범죄 언어 특징 분석

Shedding New Light on the Language of the Dark Web

Youngjin Jin¹ Eugene Jang² Yongjae Lee² Seungwon Shin¹ Jin-Woo Chung^{2*}

¹KAIST, Daejeon, South Korea
²S2W Inc., Seongnam, South Korea
¹{i jinjin, claude}@kaist.ac.kr
²{genesith, lee, jwchung}@s2w.inc

Abstract

The hidden nature and the limited accessibility of the Dark Web, combined with the lack of public datasets in this domain, make it difficult to study its inherent characteristics such as linguistic properties. Previous works on text classification of Dark Web domain have suggested that the use of deep neural models may be ineffective, potentially due to the linguistic differences between the Dark and Surface Webs. However, not much work has been done to uncover the linguistic characteristics of the

Ghosh et al., 2017; Jin et al., 2019; Choshen et al., 2019) and Avarikioti et al., 2018). However, not much work has been done on the linguistic analysis of the Dark Web (Choshen et al., 2019). In addition, the Dark Web has been studied and analyzed in the security research community to uncover malicious activities including phishing (Yoon et al., 2019), illicit online marketplace activity (Soska and Christin, 2015), terrorism (Chen, 2011), cryptocurrency abuse (Lee et al., 2019), and



ACL 2023

다크웹 텍스트를 이해하는 언어모델 DarkBERT

DarkBERT: A Language Model for the Dark Side of the Internet

Youngjin Jin¹ Eugene Jang² Jian Cui² Jin-Woo Chung² Yongjae Lee² Seungwon Shin¹

¹KAIST, Daejeon, South Korea
²S2W Inc., Seongnam, South Korea
¹{i jinjin, claude}@kaist.ac.kr
²{genesith, geeoon19, jwchung, lee}@s2w.inc

Abstract

are the Dark Web compared to that of the Surface Web. As studies on the Dark Web commonly require textual analysis of the domain, language models trained on the Dark Web may provide valuable insights to researchers. In this work, we introduce DarkBERT, a language model pretrained on Dark Web data. We describe the steps taken to filter and compile the text data used to train DarkBERT to combat the extreme lexical and structural diversity of the Dark Web that may be detrimental to build-

To handle the ever-changing landscape of modern cyber threats, cybersecurity experts and researchers have started to employ natural language processing (NLP) methods. Gaining evidence-based knowledge such as indicators of compromise (IOC) to mitigate emerging threats is an integral part of modern cybersecurity known as *cyber threat intelligence* (CTI) (Liao et al., 2016; Bromiley, 2016), and modern NLP tools have become an indispensable part of CTI research. As such, the use of NLP techniques has also been extended to the Dark Web (Jin et al., 2022; Yoon et al., 2019; Choshen et al., 2019; Al Nabki et al., 2017; Al-

도메인 특화 AI 주제로

자연어처리 분야 국제 탐티어 학술대회

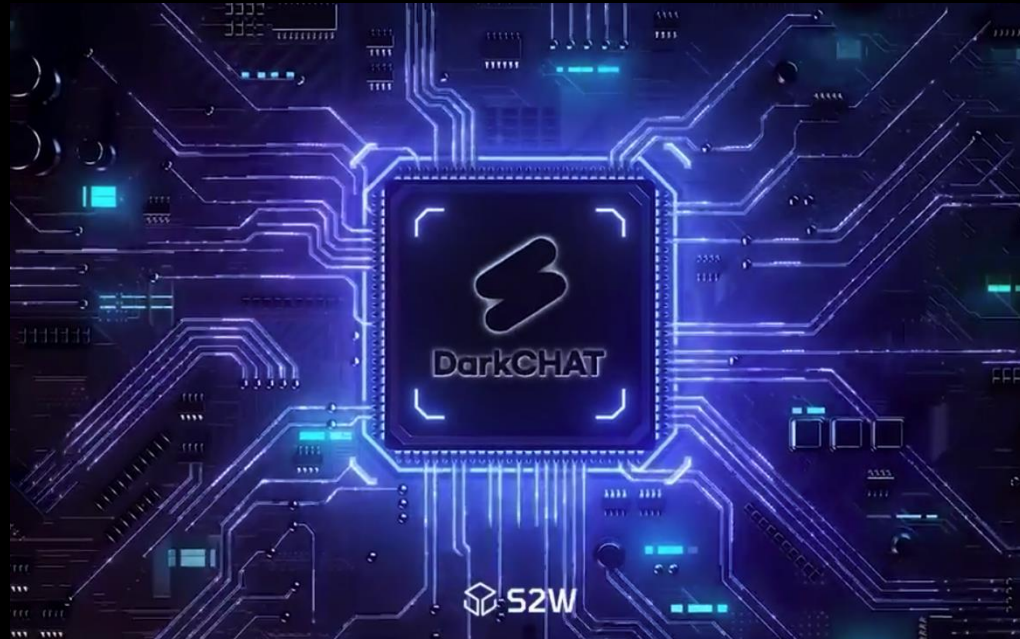
2년 연속 정규 논문 채택

비정형 다크웹 데이터
실시간 수집/정제

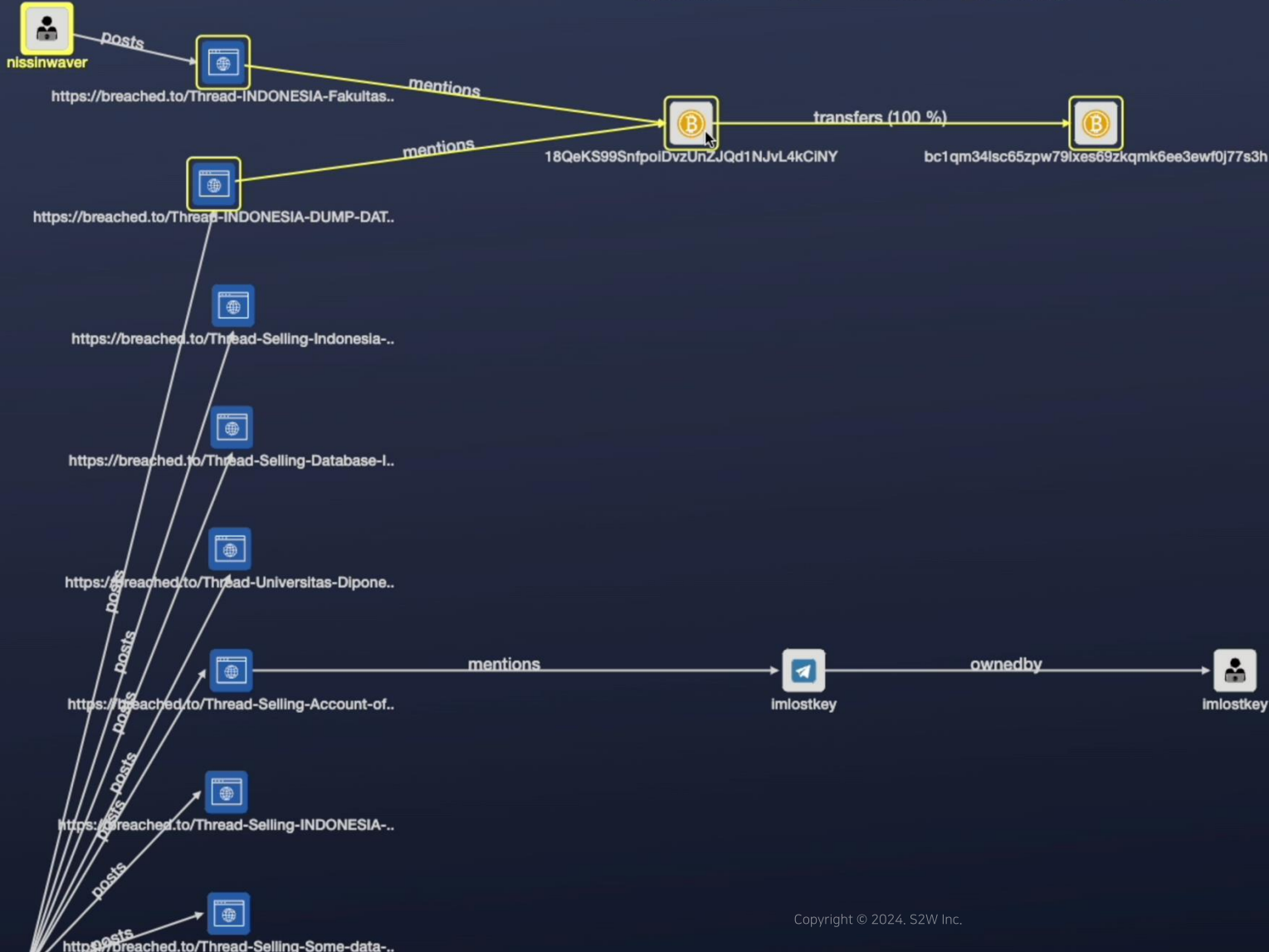
다크웹 온톨로지 및
지식그래프

다크웹 특화 AI
DarkBERT

생성형 AI 기반 다크웹 질의응답 시스템
DarkCHAT



+ Import Node Undo Redo Save Graph Share



DarkChat

extensively posted content related to Indonesian personal information in September 2022, publishing 13 posts in a short timeframe.

Please provide more details about the user lost_key

Here are additional particulars about the user known as lost_key:

- **lost_key** made their debut on the Breached Forum, with the first post appearing on September 11, 2022, and the last one on October 18, 2022.
- The user's **initial post** made their debut on the Breached Forum, with the first post appearing on September 11, 2022, and the last one on October 18, 2022.
- The user's telegram ID is 'imlostkey', and they accept BTC donations with the address '18QeKS99SnfpoiDvzUnzJQd1NjvL4kCINY'.

Tell me more about the BTC address that lost_key used

The Bitcoin address associated with lost_key is '18QeKS99SnfpoiDvzUnzJQd1NjvL4kCINY'. Interestingly, this same address has also been mentioned by another dark web user under the alias nissinwaver. It is highly likely that these two identities are, in fact, one and the same.

Send a message

기술

다크웹에서 축적한 데이터 인텔리전스 **노하우**를 다양한 도메인으로 확대

비정형 빅데이터
실시간 수집/정제

온톨로지 및
지식그래프 구축/검색

도메인 특화
AI/NLP

 **S·AIP**

유통

금융

다크웹

제조

보안

Clear

Part 2

S2W Artificial Intelligence Platform (S-AIP)

Safe and Secure LLMs for enterprise



생성형AI 도입의 허들

LLM

비즈니스 데이터를 가지고 ChatGPT처럼 말로 질의응답을 하길 원함

sLLM

기업 내부에 직접 설치해서 쓰고 싶은데
일반 LLM은 너무 무겁고, 특정 분야만 잘하는 작은 LLM을 원함

RAG

"Retrieval-Augmented Generation"
LLM이 내 데이터를 실시간으로 검색(retrieve)해서 참고하여 답을 해주기를 원함

[2024 전망] 생성형 AI 서비스 본격 상용화... "누가 먼저 깃발 꽂나"

[딜라이트닷넷] LLM 시장 본격화된 2023년, 올해는 sLLM 바탕의 '활용'에 초점

오피니언 입력 2024.01.23 00:00

[기고] 2024년, 검색증강생성 RAG 기반 LLM의 부상

홈 > 업계동향

"할루시네이션 잡고 신뢰도 ↑" 높인다" 금융·법조·공공기관 등 RAG 적극 활용 '눈길'

sLLM 과 RAG



입력한 정보가 외부 클라우드로 전송됨
(기밀문서, 개인정보 등 데이터 유출 위험)

보안성

오픈소스 sLLM을 활용한 사내 구축
(Private Cloud & On-Premise)



내부 데이터에 대한 질문에 답변할 수 없음

범용성

기업 내부 데이터를 지식그래프 형태로
변환 후 LLM 컨텍스트로 활용



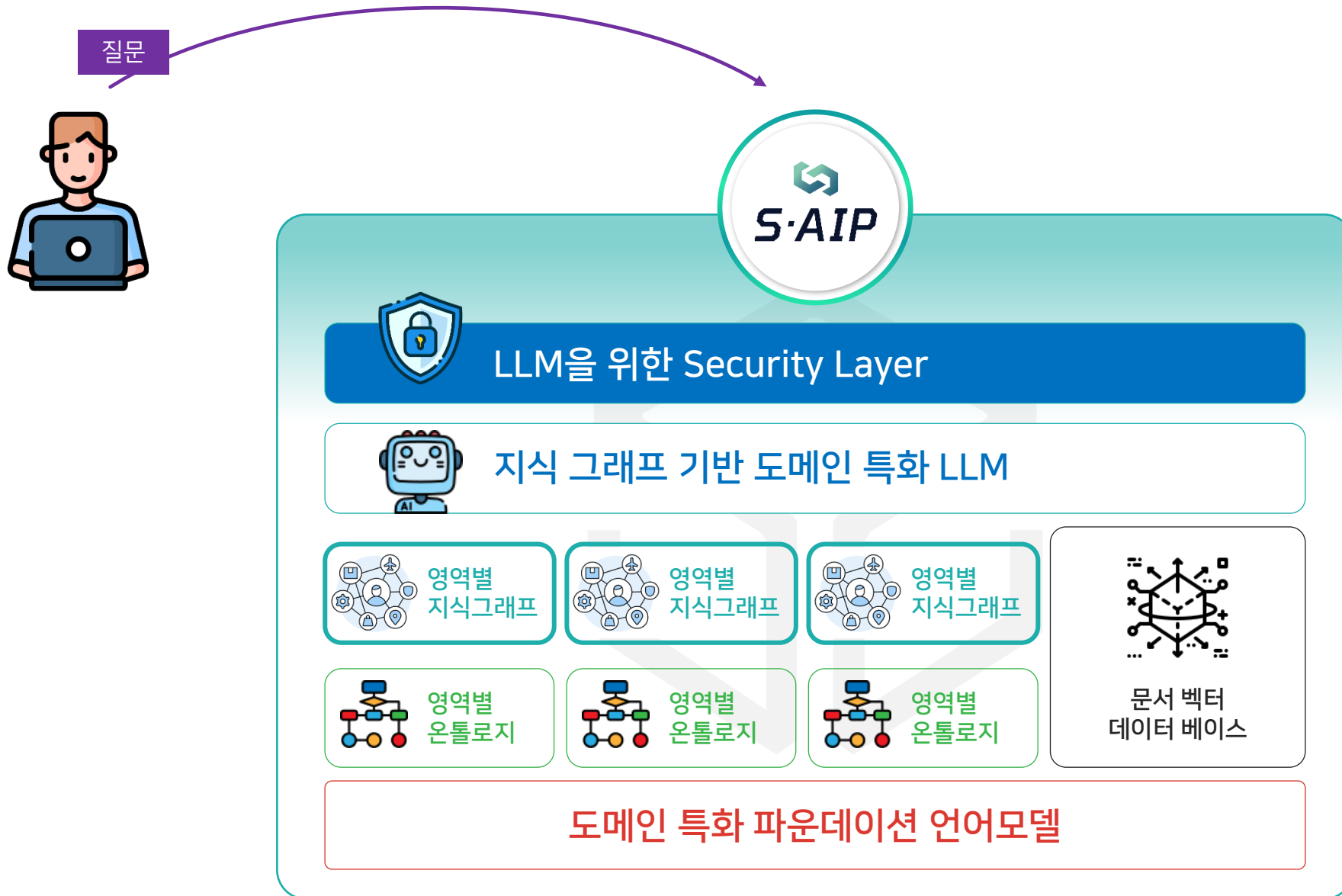
거짓 정보 등 답변을 신뢰하기 어려움 (환각)

신뢰성

답변에 대한 출처 명시로 신뢰성 보강

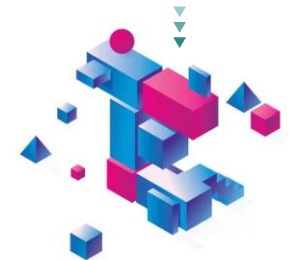


S·AIP 아키텍처

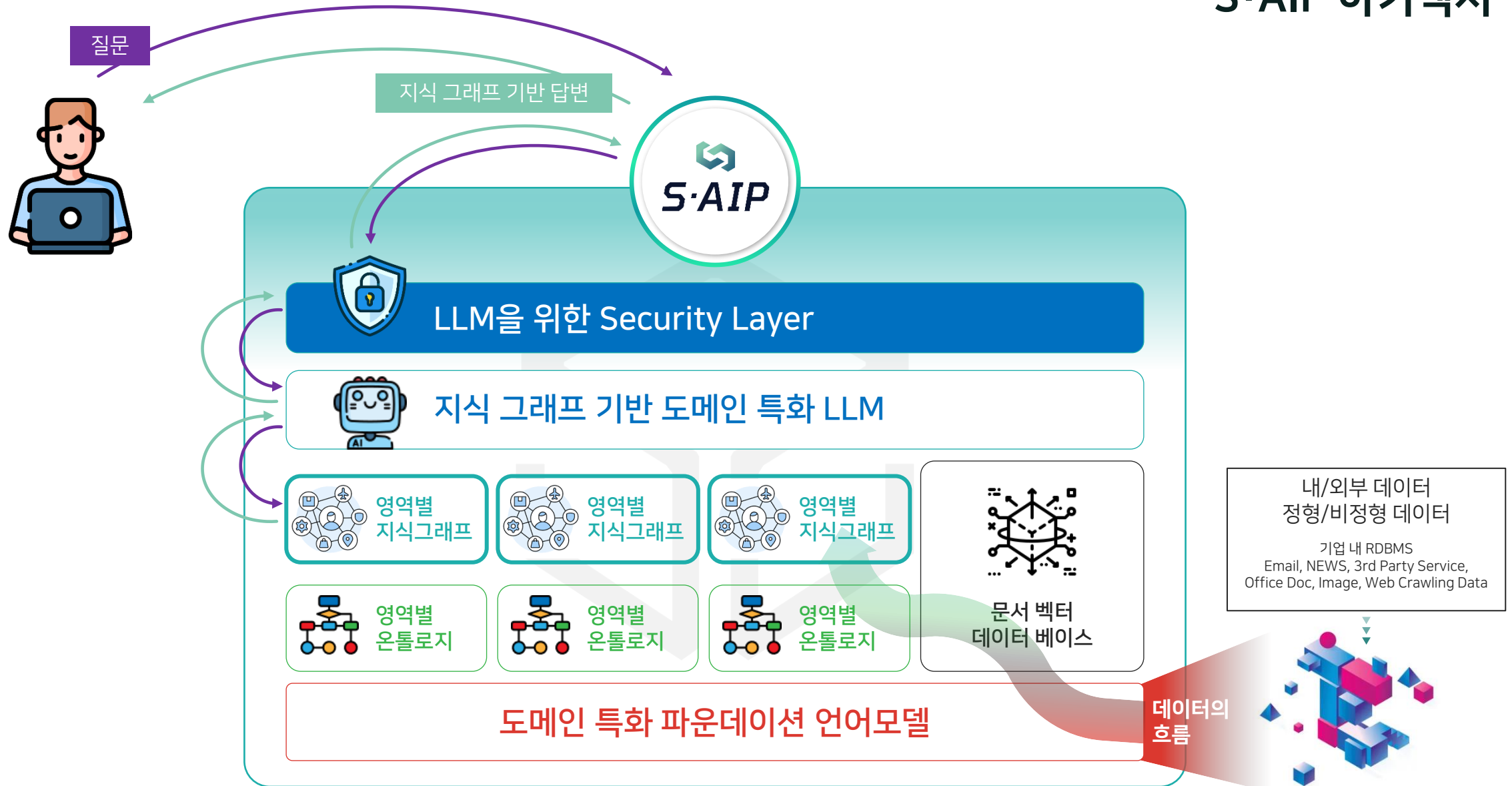


내/외부 데이터
정형/비정형 데이터

기업 내 RDBMS
Email, NEWS, 3rd Party Service,
Office Doc, Image, Web Crawling Data



S·AIP 아키텍처



국내 최고의 사이버보안 기업이 구축하는 안전하고 믿을 수 있는 기업용 LLM



S2W Artificial Intelligence Platform
데이터 인텔리전스 기업이 만든 안전한 기업용 AI 플랫폼



다양한 형식의 방대한 기업 데이터 처리
기업 외부/내부, 정형/비정형 데이터 실시간 수집/처리

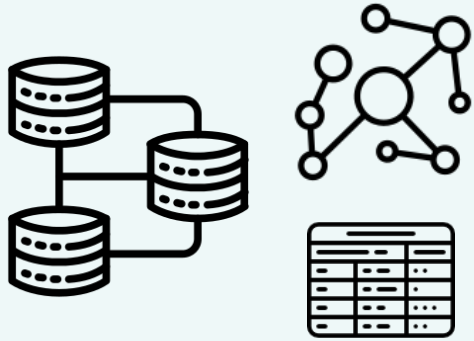


도메인 특화 지식그래프 자동 구축 및 의미 기반 검색
지식그래프, 문서 벡터 DB, 관계형 DB 통합 증강 검색



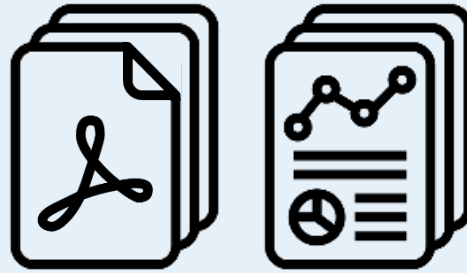
보안성 강화
사이버 보안 전문성을 기반으로 다양한 보안 테스트 수행

#1 다양한 형식의 방대한 기업 데이터 처리



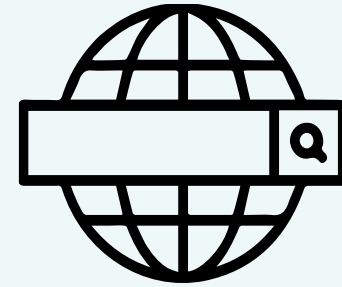
RDBMS, GraphDB, JSON 등

정형 데이터



PDF, PPT, XLS, DOC, HWP 등

비정형 데이터



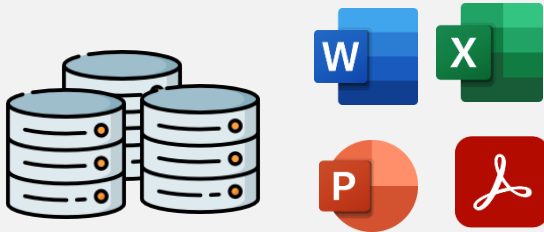
뉴스기사, 시황, 포럼 등

웹사이트 실시간 크롤링

#2 도메인 특화 온톨로지 및 지식그래프 자동 구축



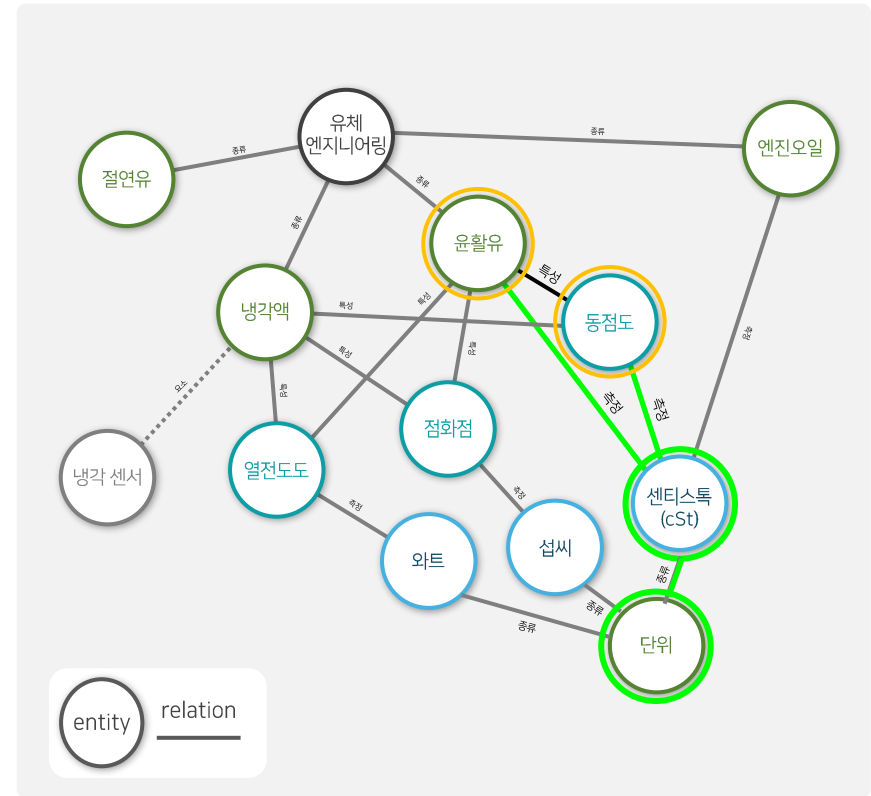
기업 외부의 공개 데이터



기업 내부의 DB 및 문서



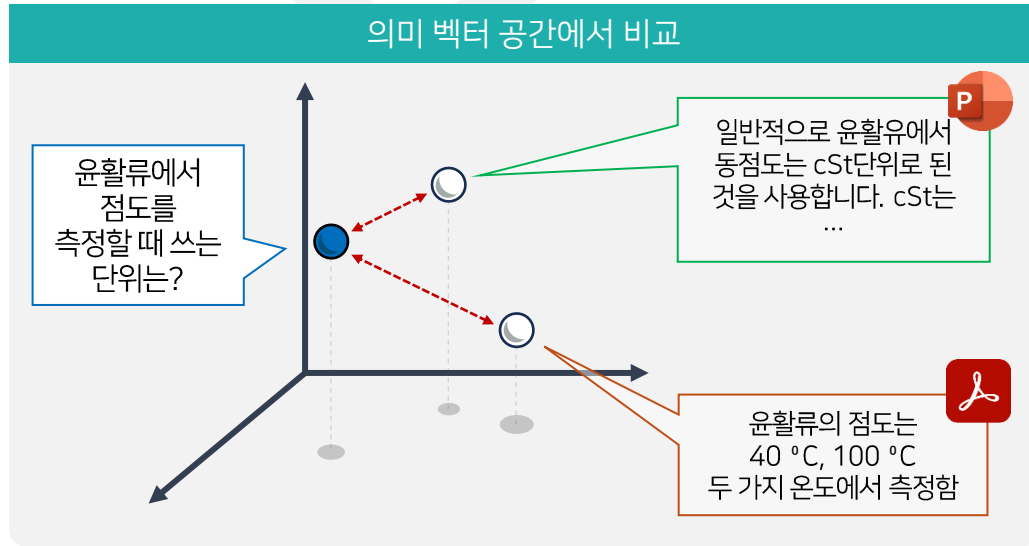
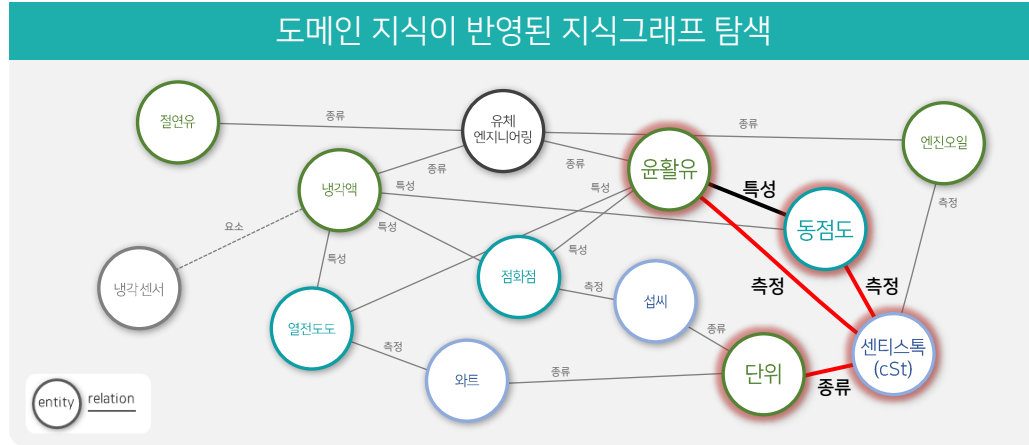
Entity, Relation
추출 엔진



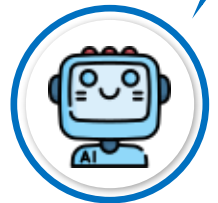
#2 도메인 특화 지식그래프 및 벡터DB 통합 검색



운할류에서 점도를 측정할 때 쓰는 단위는?



운할류에서는 점도를 측정할 때는 센티스톡(cSt)이라는 단위를 사용하며, ...

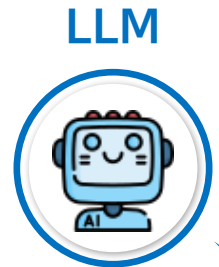
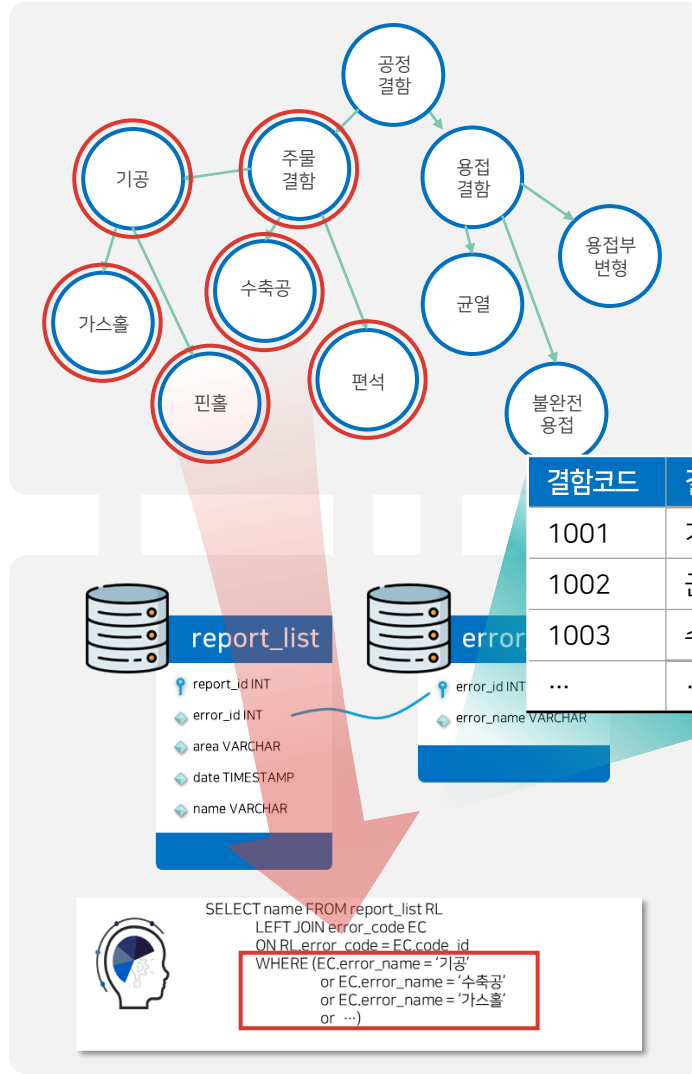


LLM

#2 도메인 특화 지식그래프 및 관계형DB 통합 검색



22년 A공장의 주물 관련 결함보고서를 결함 유형별로 정리해줘



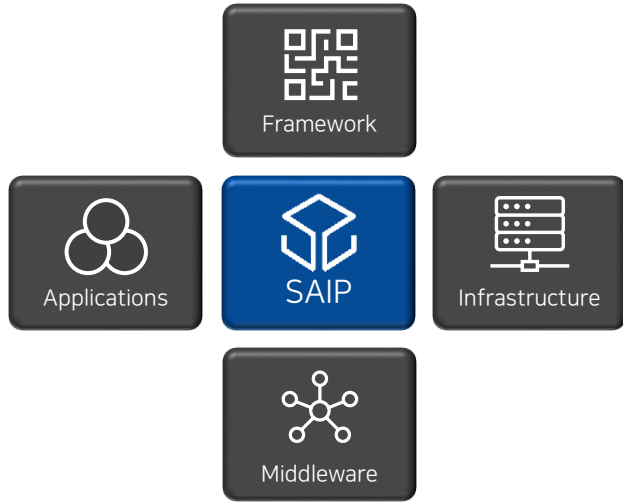
2022년 A현장에서 발생한 주물 결함에 대한 결함보고서를 결함 유형별로 분류하면 다음과 같습니다.

기공 결함 (총 4건)
 20220305_A공장_차량_주조_결함보고서_가스홀
 ...

수축공 결함 (총 7건)
 20220411_A공장_M라인_표면수축공_보고서
 ...

보안성 강화

Customized On-premise LLM Architecture



맞춤형 LLM 아키텍처

기업 보안 환경에 맞춘 안전한 LLM 아키텍처로 내부 데이터 유출 걱정 없이 내부 전용 서버에 구축

Role-Based Access Control



역할 기반 액세스 제어(RBAC)

사내 IT 시스템 내에서 사용자/직급별로 접근 권한을 다르게 설정하여 효율적인 데이터 보안 지원

Security Enhancement



모의 공격 및 취약점 테스트

사이버 보안 전문성을 바탕으로 Prompt Injection, Model DoS 등 다양한 LLM 관련 공격 및 취약점 테스트를 수행하여 잠재적인 위협 조기 탐지 및 대응



1

도메인 특화 온톨로지 구축을 통해
도메인 이해도가 높고

2

지식그래프 기반으로 비정형 데이터에 대해
정확한 답변을 도출할 수 있고

3

보안성이 강화된 생성형 AI 플랫폼