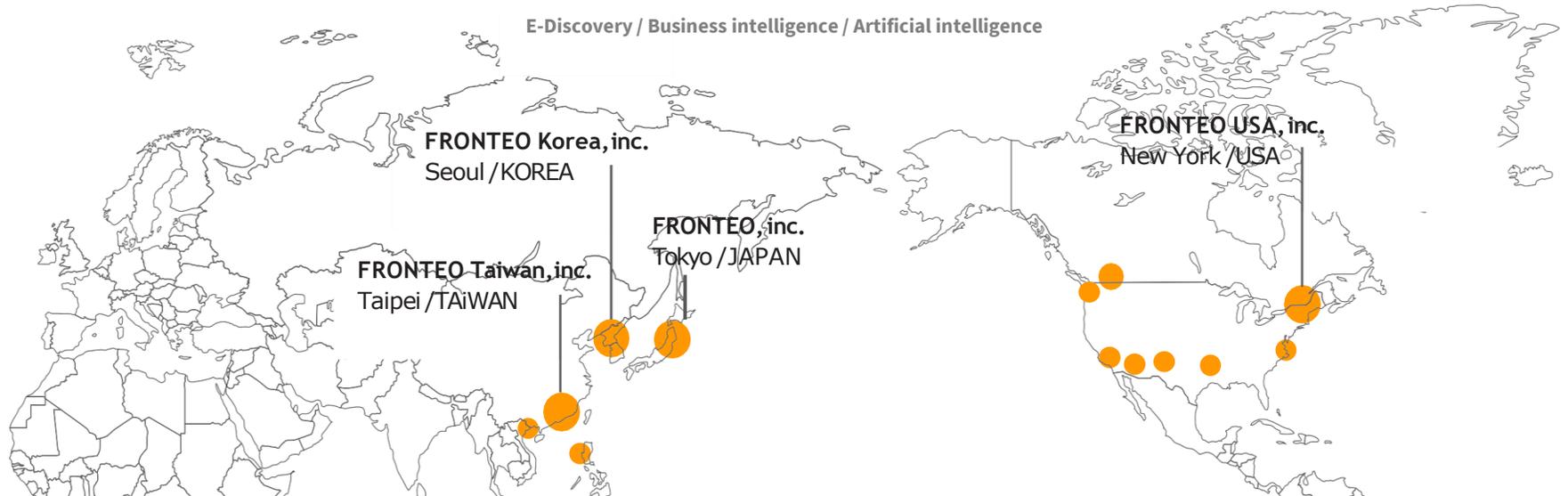


# Lit i View EXAMINER



# FRONTEO

E-Discovery / Business intelligence / Artificial intelligence



국제 소송 대응 / 사기 조사 :

총 **9,200** 건 이상



리걸테크	<p><b>국제 소송</b> (eDiscovery)</p>	<p>미국 · 유럽 소송 Anti Trust소송 /iP소송 /Class Action소송 /PL소송 등</p> <p>지원 건수 <b>7,800건 이상</b></p>
	<p><b>부정 조사</b> (디지털 포렌식)</p>	<p>기술 정보의 유출 · 영업 정보 유출 · 회계 부정 · 횡령 리베이트 · 불법취업 · 카르텔 · 뇌물 증거 은닉 사안 · 증거 제출 사안 etc.</p> <p>수주 건수 <b>1,400건 이상</b></p>
포이화	<p><b>교육</b> (디지털 포렌식)</p>	<p>경찰청, 방위성, 전국 경찰 본부 기타 관공서, 기업 (제조업 · IT관련업 · 전력 · 통신업 · 금융 기관 등)</p> <p>수강자 <b>990명 이상</b></p>

# FRONTEO의 인공 지능

---

# Kibit

Kibi : 인간의 미묘한 마음의 움직임      Bit : 정보량의 최소 단위

**Landscaping 과 행동정보과학을 결합한 FRONTEO의 인공지능  
사람보다 약 4,000배의 빠른 속도로 리뷰 가능**

Landscaping

(특허 제5567049호, 특허 제5526209호)

학습 · 평가를 담당하는 인공지능 관련 기술

행동정보과학

대규모 지식을 구성하는 접근 방식

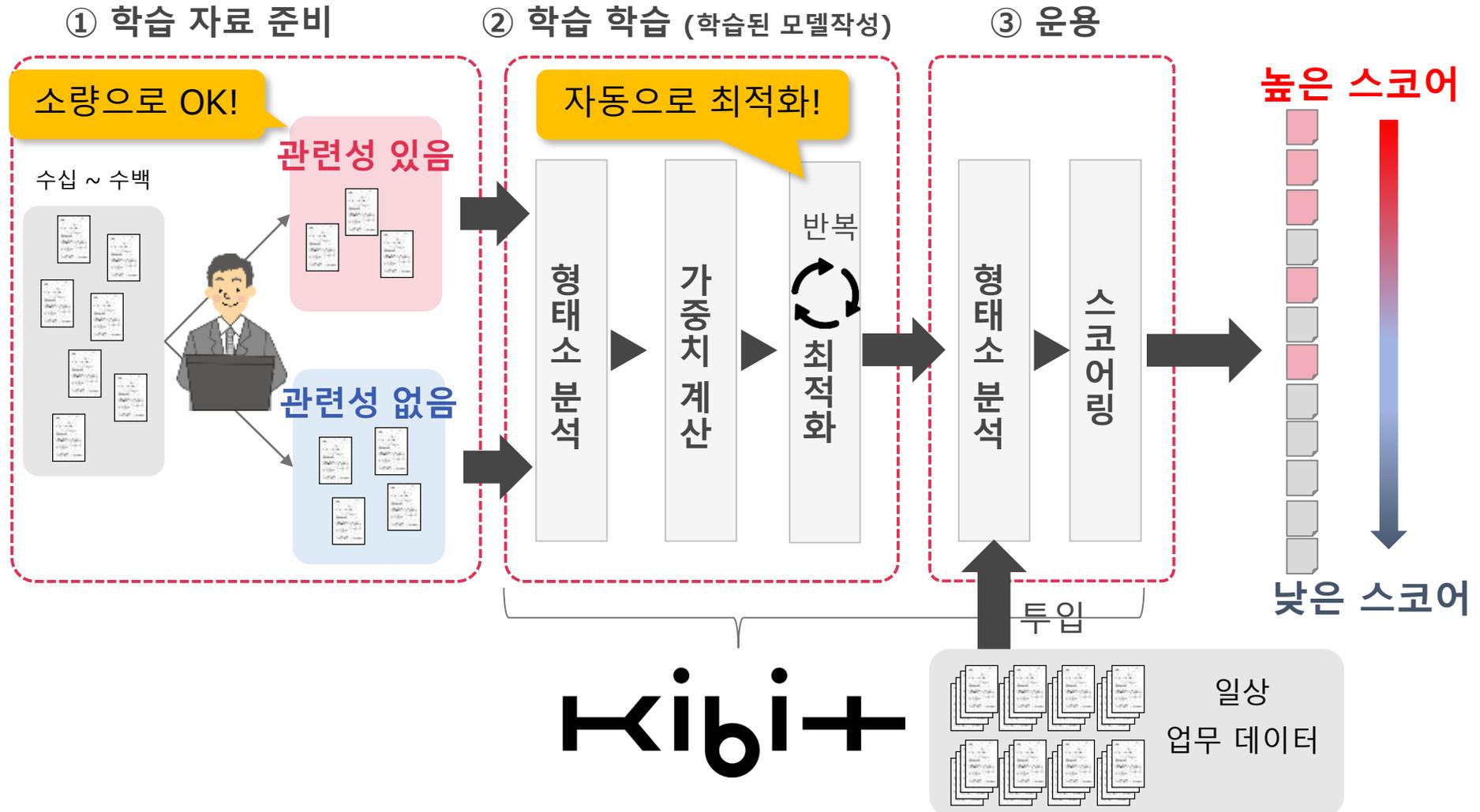
자연어에 특화

소량의 학습 데이터 OK!

# 인공지능 「Kibit」 개요

인간의 판단기준을 소량의 데이터에서 학습한 인공 지능 Kibit이 대량의 데이터에 스코어링을 실시하여 스코어가 높은 문서부터 검토함으로써 시간과 비용을 크게 줄일 수 있습니다.

인공지능 「Kibit」 학습에서 운용까지의 프로세스



# FRONTEO Solution Family

Legal-Tech를 비롯, 다양한 분야에서 활용되는 FRONTEO Solution



**Lit i View  
E-DISCOVERY**

**Lit i View  
EXAMINER**

**KiBi+  
Patent Explorer**

**KiBi+  
Knowledge Probe**

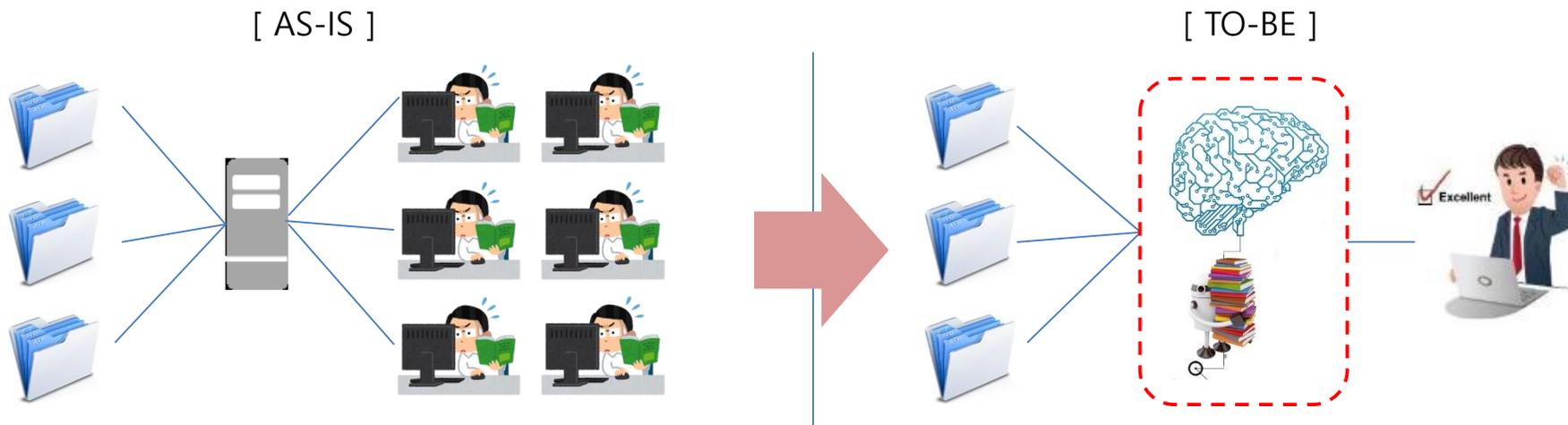
**KiBi+  
Email Auditor**

**KiBi+  
Find Answer**

Lit i View  
**EXAMINER**

# 현행 업무 현황 및 극복 과제

감사/수사/조사에 필요한 대용량의 문서를 처리 하기 위해서 현행 솔루션으로는 많은 리소스가 소요 됩니다.  
 또한 문서의 종류가 office, pdf, hwp, pst 등 다양하며 양도 점차 증가함에 따라 **업무환경의 변화와 새로운 패러다임에 부합하는 솔루션의 도입이 필요**합니다.



고용량 프로세싱 **속도 이슈** 발생

**외산 솔루션**으로는 해결이 어려운 **'한글'** 언어 분석

문서 분석에 투입되는 **인력 비용 증가**

**단순 키워드** 검색으로는 **찾기 어려운 문서** 다수

**불필요한 문서** 추출로 검토 대상 과대 발생

**조사 업무에 원활한 수행 어려움!**

# 인공지능을 탑재한 감사/수사/조사 솔루션의 필요성 대두

감사/수사/조사 업무에서 처리할 데이터의 양은 갈수록 증가하고 있으며 고용량의 비정형 데이터 내에서 사용자가 필요로 하는 정보를 얻기 위해서는 많은 시간과 비용이 소요 됩니다.

**최근에는 인공지능(AI) 기술을 활용하여 업무의 효율성을 증대 시키는 것이** 최신 트렌드로 자리잡고 있습니다.



## 비정형 데이터 분석 인공지능의 강점

- eDiscovery의 최신 문서 분석 방법인 '**Predictive Coding\***' 을 활용하여 대용량의 문서를 효과적으로 분석
- 머신러닝을 통해 인간이 직접 검색하는 방식을 최소화하여 시간 및 비용 단축 실현
- 빠르고 정확한 검색을 지원하여 산출물의 질(質)을 향상시키고 전반적인 업무 효율성 극대화

**\*[참고] Predictive Coding : Machine Learning 기법을 활용하여 Training 데이터를 학습하고, 그 학습 결과를 기반으로 다른 데이터를 분석 및 평가하여 의사결정 기준에 부합하는 데이터를 추출하는 원리**

## 추진 방향

프론테오는 아시아 최고의 eDiscovery 소송 경험에서 축적한 노하우와 인공지능 KIBIT을 탑재한 감사/수사/조사 솔루션 **Lit i View Examiner**를 제안합니다.

### 신속하고 안정적인 프로세싱 구현

- ✓ 고용량의 데이터를 빠른 시간 안에 처리하여 감사/수사/조사 시간 단축
- ✓ 한글 및 다국어(CJK)를 지원하여 국내 조사 시스템의 환경에 적합한 시스템 구현
- ✓ OFFICE, PDF, HWP 등의 다양한 문서를 프로세싱 지원하며 이미지/PDF 내TEXT를 추출해 문서의 누락 최소화

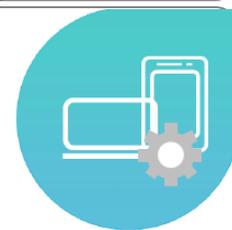


### 효과적인 검색 및 리뷰 환경 제공

- ✓ 키워드 검색은 기본! 키워드로 검색하기 어려운 문서까지도 인공지능 분석을 통해의 사결정에 부합하는 데이터를 정확히 추출
- ✓ 데이터 중복제거, 각종 필터링을 통해 원하는 문서들 중심의 신속한 리뷰 제공
- ✓ 다양한 키워드 조합 시뮬레이션으로 세부적 문서 분석 지원

### 업무 효율성 증대

- ✓ 사람대비 최대 4000배 빠른 리뷰 속도
- ✓ 인력 투입 최소화를 통해 인건비 감축하여 ROI 증대
- ✓ 단일 서버에 여러 조사인원이 분산 배정되어 다수의 인원으로 동시 리뷰 및 분석 지원



### 서비스 확장이 용이한 시스템 구현

- ✓ 프로세싱 용량 별 솔루션 산정, 향후 증설 시 간단한 설정으로 용량 증설 지원 (단, H/W 제외)
- ✓ 데이터 인덱싱과 호스팅을 동시에 제공하는 일체형 솔루션으로 손쉽게 구축
- ✓ 자사는 솔루션 외 eDiscovery/포렌식에 대한 전문 인력이 국내에 상주하며 서비스 상품도 동시 제공 가능

Lit i View Examiner는 다양한 영역에 활용 가능하며 특히 감사/수사/조사/포렌식 등의 조사 대상이 미확정적인 다량의 문서 내에서 증거 확보, 사건 경위 파악, 위험 요소 감지 등에 탁월한 효과를 얻을 수 있습니다.



- ◎ 문제점
  - 세무조사 기관으로 탈세 관련 대용량 데이터 분석 필요
  - 조사 대상 데이터 양이 너무 방대하여 적합한 검색 솔루션 찾기 어려움
- ◎ 해결 방안
  - Lit i View 를 도입하여 고용량 프로세싱 환경 구축
- ◎ 기대효과
  - 분산되어 있는 조사 대상 데이터 취합하여 하나의 리뷰 가능 환경 제공
  - 빠른 프로세싱 속도로 세무 조사 시간 최대 단축



- ◎ 문제점
  - 다량의 문서에서 주식 불공정 거래, 매입/매도 감시 중 부정 행위 탐지 어려움
  - 조사와 관련한 문서의 키워드 선정의 어려움, 키워드로도 추출이 어려움
- ◎ 해결 방안
  - Lit i View Examiner의 **predictive coding** 기능을 이용하여 부정 행위 관련 문서를 빠른 시간 안에 검색
- ◎ 기대효과
  - 다량의 문서에서 키워드 검색으로는 찾기 어려운 문서를 손쉽게 추출 가능
  - 빠른 검색으로 조사 시간 단축, 업무 효율성 증대시킴

# 기대 효과

Lit i View Examiner 는 신속하고 정확한 문서 취합, 분석, 리뷰 환경을 제공하여 각종 조사 업무를 진행하는데 안정적이고 효과적인 성과를 얻을 수 있도록 지원합니다. 인공지능을 활용한 분석을 통해 사용자가 처리해야 할 업무 부담을 크게 줄일 수 있으며, 조사 과정의 시간을 대폭 단축시킴으로써 업무 능력을 향상시키는 데 결정적인 역할을 담당합니다.

## 효율적인 조사 시스템 구축

### 조사/감사 효율성 극대화

- ✓ 인공지능을 통해 분석 시간 단축 & 정확한 분석을 통해 검색 최적 시스템 구현
- ✓ 분산된 데이터를 수집, 분류, 검색하여 관리 용이성 확보

### 질 높은 조사 성과물 도출

- ✓ 원활한 한글 문석 분석을 조사 내용의 누락 최소화
- ✓ 단순 키워드 분석으로 추출 불가능한 조사 내역까지 검색하여 수준 높은 결과물 제공

### 운영 비용 절감

- ✓ 공공기관 : 감사, 조사, 수사에 대한 인프라 구축, 인력 투입, 운영 비용 단축
- ✓ 일반기업 : 내부 감사 운영 비용 및 벌금, 과징금으로 인한 비용/손실 최소화



심각한 정보유출 사고 5건 중 1건은 **부주의한 직원**으로 인해 발생  
 - 2017, 카스퍼스키 랩 시큐리티 리포트



기밀정보 유출 42%는 **직원**에 의해 발생.  
 정보유출 처리 비용 : **중소기업 9,313만원, 대기업 9억5,934만원**  
 - 2017, Gartner, Security Report



**특허청**  
 Korean Intellectual Property Office

최근 **5년간 유출 피해의 81.4%**가 내부직원에 의한 사건  
 서류, 도면을 **이메일** 통해 전송하는 **비율이 44.2%**로 매우 높음  
 건당 21억 이상의 잠재적 피해  
 - 2017, 특허청, 국내 영업비밀 보유 616개 사 조사 결과

**내부감사  
요건발생**

- 각종 개인정보 유출사건 발생
- 내부직원 혹은 협력업체의 보안의식 부재로 인한 각종 보안 사건 발생
- 담합, 횡령, 카르텔 등 다양한 부정행위 개인 통해 발생

**대규모  
고속감사  
필요성**

- 전산화되는 업무 특징 상 대용량 데이터 감사 필요
- 다양한 유형의 악용 가능성에 대비
- 데이터 처리, 검수 인력 비용 ↑
- 다양한 파일 포맷에 대한 분석, 문서에 숨어있는 OCR 데이터에 대한 분석

**비정형  
데이터  
처리문제**

- 다양한 언어에 대한 Detection 필요
- 검증이 힘든 비정형 언어 데이터 분석 필요
- 첨부 파일, 혹은 파일 내 Embedded Object에 대한 분석 필요

## Lit i View EXAMINER

- 「Lit i View Examiner」 풍부한 경험을 바탕으로 독자 개발한 e-Discovery / 감사 전문 분석 Tool
- 한글 및 국내 파일, 이미지 지원 → Full Scope investigation 구현

인공지능 기반 "Predictive Coding" ⇒ 빠르고 정확한 증거자료 조사/분석



- ❑ 인공지능을 활용, 학습한 데이터 기반으로 대상 데이터에 대한 시나리오 적합성 여부 수치화
- ❑ 시간 당 문서 검토 수를 한산하면 수작업 대비 약 **4000배 ↑**
- ❑ 키워드 검색 맹점과 수작업 인간 오류 방지 지원
- ❑ 다양한 파일과 이미지, 임베디드 오브젝트를 통한 언어 추출 지원

**빠르고 정확한  
분석 기능 제공**

Lit i View Examiner 통해 다양한 시나리오별 / 노드 별 / 용의자 별 데이터 수집, 분석과 결과 확보 가능

## 제공 기능별

감사 데이터 배치 / 실시간 수집

비정형 데이터 분석

아시아권 언어 자연어 처리

OST, HWP 등 한국 특화 확장자 지원

다양한 판단 기준에 대한 처리 방안 지원

수치화 된 결과 제공

Central Linkage를 통한 의사결정 지원

## 수행 작업별

감사 데이터 수집 및 분석

- 감사 인원 / 감사 사건 별 / 시나리오 별 데이터 분석
- 시나리오 도출 및 피드백 서비스 제공
- 긴급 사항 발생 시 알림 서비스 제공

한국어 및 관련 파일 지원

- 한자 포함한 아시아권 언어 지원
- ALZ, GUL, HWP 등에 대한 지원
- 국/영문 혼합 문서에 대한 자동 분석 지원

다양한 범위의 데이터 추출

- PDF 등에 대한 OCR 기능 제공
- 첨부파일 내 첨부파일의 분석 제공
- 다양한 시나리오에 따른 분석 제공

다양한 형태의 분석 시나리오 및  
Central Linkage 분석 제공

- 단일 문서군에 복수 시나리오 적용 가능
- 발신-수신 도메인별 이메일 분석을 지원하는 Central Linkage 분석 제공

## 수행 결과별

감사 대상 인원 별 / 감사 사건 별 /  
감사 노드 별 활동 분석 및 데이터 점검

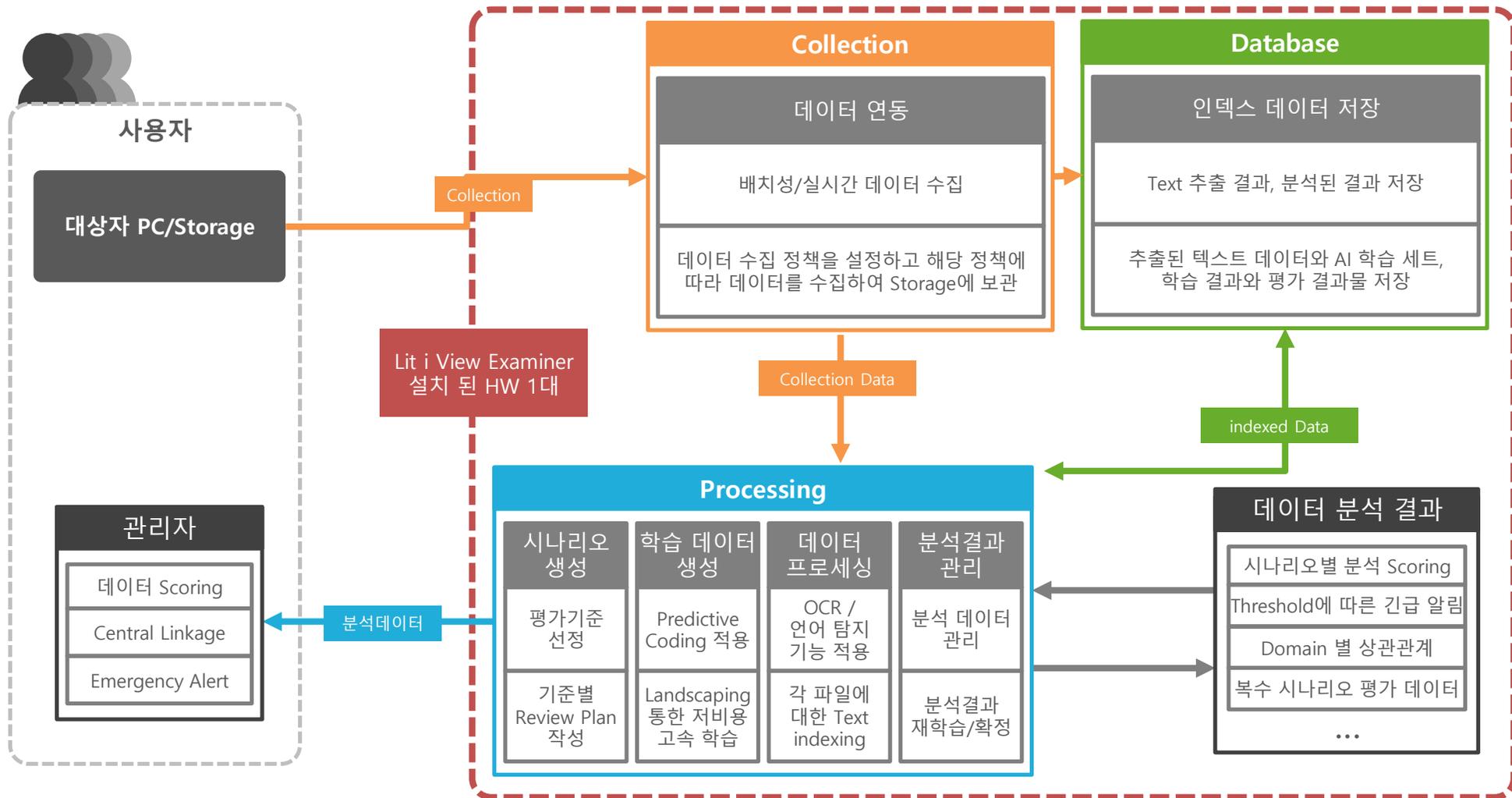
검증된 한국어 처리능력, 다양한 한국  
시장에서의 전용 확장자 지원을 통한  
높은 신뢰도 확충

단순 텍스트 분석이 아닌 문서 내 모든  
범위에 대한 데이터 추출 및 분석

적은 양의 학습을 통한 시나리오 반영  
기능과 분석 도구 제공

# Examiner 시스템 구성도

1대의 하드웨어에서 데이터 수집, 분석, 리뷰, 관리자 조회 웹 서비스까지 제공하는 All In One 시스템



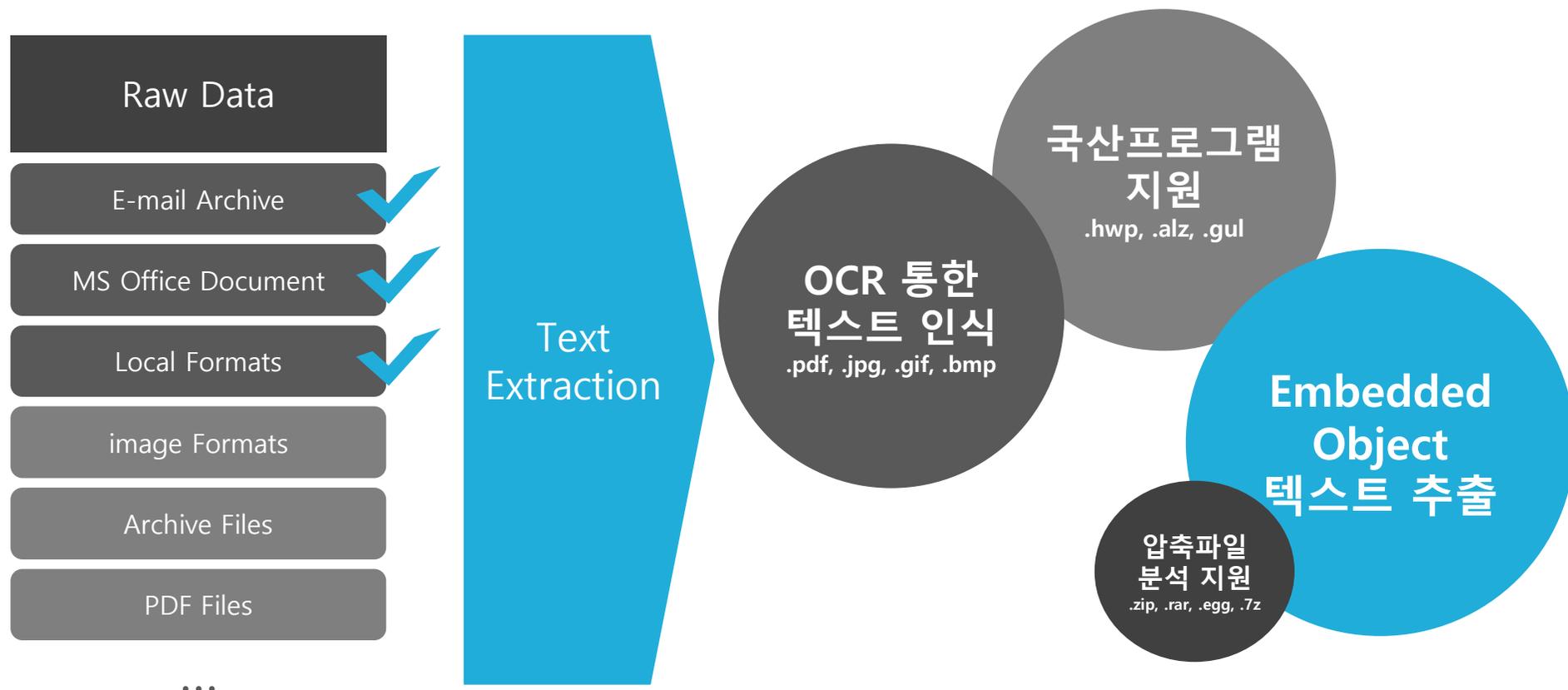
# Examiner Feature 1) Collection

Collection 과정에서 Text Extraction 수행, 언어 판독, 자연어 처리, OCR 통한 다양한 텍스트 추출 기능 지원



## Examiner Feature 2) Processing

Lit i View Examiner는 데이터 노드 별/개인 별/사건 별로 수집, 분석하여 텍스트 자료 구조로 변경하며 변경 과정에 OCR, Embedded Object, 언어 판독 등 다양한 기능 지원



## Examiner Feature 3) AI Scoring

LandScaping를 통한 소량 학습으로 작업 공수 **다운!** Predictive Coding을 통한 문서 평가 정확도 **업!**

감사 결과 개선, 성능 향상, 원가 절감 효과

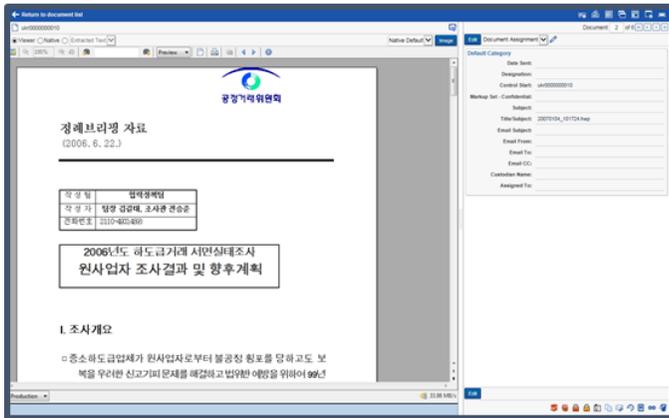


## Text Extraction

강력한 한글 추출 능력 지원

### 수행 능력 비교

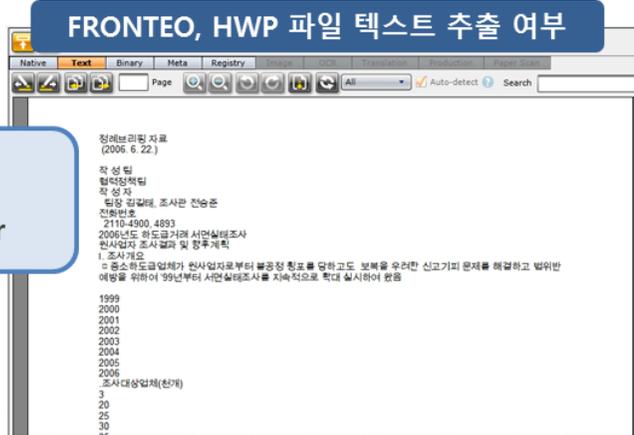
- ※ 타사 대비 **한글 및 특정 문서 확장자** 인식/검색기능 비교:
- 1) HWP 파일 텍스트 추출
  - 2) 한글 추출 기능 (이메일-MSG 파일)



FRONTEO  
LiV Examiner



타사  
Tool



### 타사, HWP 파일 텍스트 추출 여부



사람보다 4,000배 빠른 인공지능 통한 문서 분석

## Predictive Coding

### 문서리뷰 Tool 활용

검토속도 vs. 정확도

#### Review의 Speed

- 평균 1시간 당 50개의 문서 리뷰 가능
- 강한 동기 부여, 숙련된 경험 보유, 평균 36초 내 결정하는 간단한 구분 작업이라는 가정 하에 성립

#### Review의 정확도

- 숙련된 "speed reader" 는 1분당 1,000개의 단어들을 50%의 이해도 로 속독하게 된다.

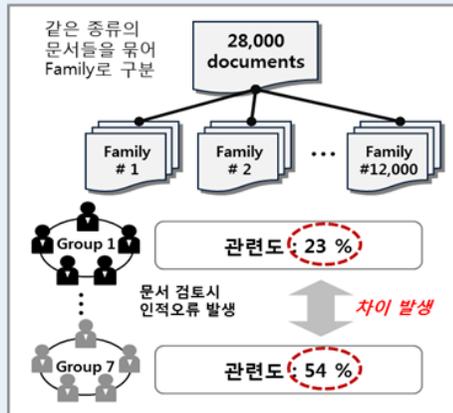


Lit i View  
**EXAMINER**

## 인공지능 기반 Predictive Coding

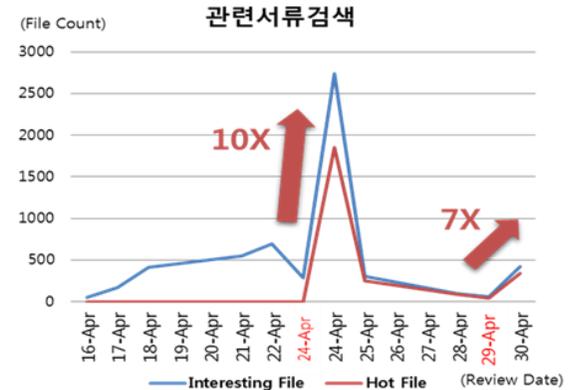
- 사람 문서검토 대비  
⇒ 4,000배 속도 향상
- 단순 문서리뷰 Tool 대비  
⇒ 정확도 7~10 배  
⇒ 속도 2배 이상 증가

### 일관성이 낮은 인위적 검토



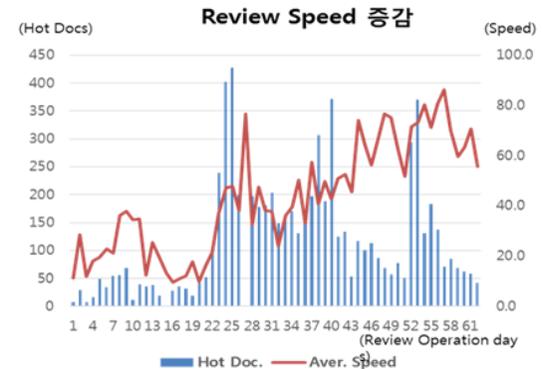
### K Case (Korean + Japanese Document)

- Review 시작일: 16-Apr, 2014
- 기술 도입일: 24&29-Apr, 2014
- 효과: 7~10배 이상의 관련서류 검색 가능



### C Case (Korean + English Document)

- Review 시작일: 13-Jan, 2014
- 기술 도입기간: after 26-Mar, 2014
- 효과: Review Speed, 2배 이상 증감



## 키워드 검색 기능

추출한 데이터 전체에 대한 키워드 기반 검색 기능 제공

The screenshot displays the FRONTEO email examiner interface. At the top, a table lists extracted email data with columns for subject, file name, full path, date, from, to, document ID, file name, page count, native status, text encoding, and text length. Below the table, a search bar is highlighted with a red box, containing the Korean text '출고가' (Chul-goga). The email preview below shows a message from '정민석 부장드림' (Director Jeong Min-seok) regarding a price reduction for '출고가' (Chul-goga) from 990,000 KRW. A red box highlights the word '출고가' in the text. A large red 'SAMPLE' watermark is overlaid on the right side of the preview. The interface includes a left sidebar with response status filters, a top toolbar with navigation and search icons, and a bottom 'Close' button.

MAIL.Subj	FS.Extensi	FS.FullPath	MAIL.Date	MAIL.From	MAIL.To	LIV.Doc ID	FS.File Nar	LIV.Mail Pa	LIV.Native	LIV.Text E	LIV.Te
RE:_RE:_미팅_Re msg		D:\itview\Uploa	2014/05/30 17:5	UBICOB <ubic_o	"yb" <ubic_yb@	41590	A-Telecom.pst::C	41590	3033	1	1
Re: RE: 미팅 Rep msg		D:\itview\Uploa	2014/05/30 17:5	TA UBIC <ubic12	UBICOB <ubic_o	41612	A-Telecom.pst::C	41612	3424	1	1
RE:_Re:RE:_RE: msg		D:\itview\Uploa	2014/06/03 15:0	UBICOB <ubic_o	"yb" <ubic_yb@	41641	A-Telecom.pst::C	41641	7433	1	1
RF: RE: RF: 미팅 msn		D:\itview\Uploa	2014/06/02 14:3	UBICOB <ubic_o	"vb" <ubic_vh@	41697	A-Telecom.pst::C	41697	5046	1	1

LIV.Doc ID : 41641 FS.File Name : A-Telecom.pst::018.msg

Search 출고가

안녕하십니까 임부장님 고부장님,

내부적으로 출고가에 대해 논의한 결과,  
저희는 출고가를 99만원으로 내부적으로 결론 내렸습니다.

정민석 부장드림.

**SAMPLE**

Close

# Examiner Detail

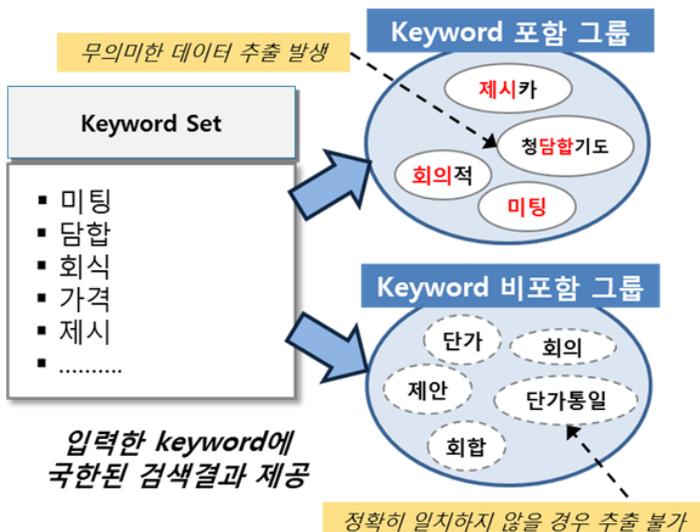
## 키워드 검색 기능

정밀한 키워드 추출 알고리즘 제공

### Keyword 검색 문제점

- 검색 키워드를 선정하는데 많은 시간이 소요됨
- 불필요한 문서까지 추출됨에 따라 검토 대상이 과대 발생함
- 중요 문서, 누락가능성 존재 (키워드에 따라 역으로 누락)
- 중요도 및 연관도에 관한 정보 식별 불가

### Keyword base search 원리



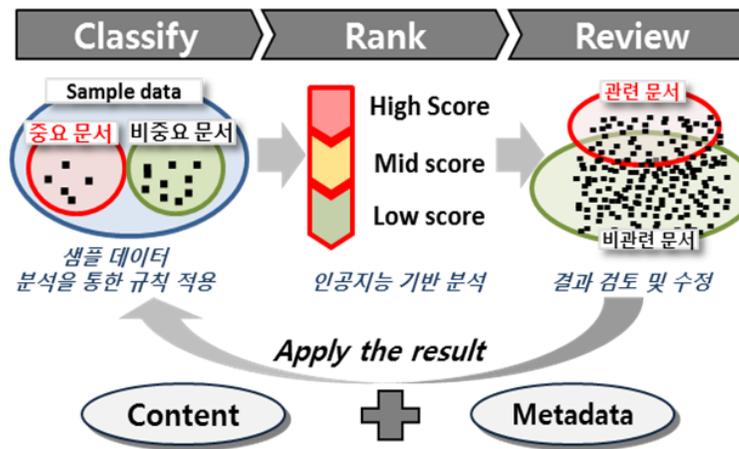
### Examiner 장점

Training set 설정

문서들간의 연관성  
인공지능으로 분석

Predictive Coding은 Sample training set에 기반하여 컴퓨터가 전체 문서에 대한 분류 방식을 습득하고 관련성에 따른 확률을 기반으로 자동적으로 문서를 분류하도록 함

### 인공지능 기술 활용 이미지





## 중요도 순 감사 리포트

학습된 패턴 기반 감사 리포트 제공 및 고위험군 적발

- FilteringRuleForDemo
- UKR\_Filtering
- UKR\_Filtering\_KW
- 카카오톡
- 리포트
- 정보검색
- 로그정보
- 행동추적

From: 2014/06/29 To: 2014/07/28 Apply

Term	2014/06/29 ~ 2014/07/28
Total Number of Audited Mails	864
Total Number of Mail Addresses	28
Suspicious Mail Count (Score 3000 ~ 9999)	65

**Overall Evaluation**

During the period of 2014/06/29 ~ 2014/07/28, suspicious mails which has 3000 ~ 9999 score has been detected in the 5 e-mail account(s).  
 Since the risk factor of [UKR\_Filtering\_KW] is low, the survey is not required.

**Suspicious Mail Addresses (5)**

#	Mail Address	# of Mails	Linkage
1	ubic_yb@nate.com	26	Central Linkage →
2	ubic12ta@gmail.com	18	Central Linkage →
3	ubic_ob@hotmail.com	17	Central Linkage →
4	ubic_bb@daum.net	3	Central Linkage →
5	sunny_yoon@ubic.co.kr	1	Central Linkage →

**Behavior Analysis of Suspicious Mail (Top 10)**

#	Keyword1	Keyword2	# of Mails
1	message	date	43
2	ubic	[mailto]	41
3	send	friday	40
4	ubicob	subject	40
5	ubic	subject	39
6	ceo	nasdaq	35
7	list	york	35
8	nasdaq	list	35
9	new	york	35
10	telecom	director	35

The chart displays the number of suspicious mails (yellow bars) and audited mails (green line) from 2014/6/29 to 2014/7/17. A significant spike in suspicious mails is observed on 2014/7/15, reaching approximately 45. This spike is highlighted with a red rectangular box.



# Thank You



**FRONTEO Korea, inc.**

| Tel : +82-2-350-3000 | Fax : +82-2-3288-3363

| <http://www.fronteo.com/kr/>

| 11th Floor ARC Place 142 Teheran-ro, Gangnam-gu, Seoul, South Korea 06236