

모 금융회사



오픈소스 및 머신러닝 도입 이야기



김형준

- 0 발표자 소개
- 1 인터넷 폐쇄망에서 분석 시스템 구축 (feat. 엔지니어가 없을 때)
- 2 분석 보고서 자동화
- 3 Machine Learning 삽질기 (분석 & 개발)

- 0 발표자 소개
- 1 인터넷 폐쇄망에서 분석 시스템 구축 (feat. 엔지니어가 없을 때)
- 2 분석 보고서 자동화 하기
- 3 Machine Learning 삽질기 (분석 & 개발)

삼질의 시작

(2005.3~2012.8) 심리학 / 인류학



형준 7은 E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 (적극성10 + 면역4)
형준 8은 형준 7 + PI 추가
E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 + PI9 + PI10 + PI11 + PI12 + PI13 + PI31 + P
I32 + PI33 + PI34 + PI35 + PI36
형준 9는 추가된 PI 수정
E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 + PI9 + PI10 + PI11 + PI12 + PI13 + PI31 + P
I32 + PI33
형준 10은 형준 9에 개방성 추가
E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 + PI9 + PI10 + PI11 + PI12 + PI13 + PI31 + P
I32 + PI33 + ROP1 + OP5 + OP6 + OP9 + OP12 + ROP21

형준 11은 (적극성 + 면역 + 개방성 + 상은이가 만든 OA + OF 중 몇 문항)

E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 + PI9 + PI10 + PI11 + PI12 + PI13 + PI31 + P
I32 + PI33 + ROP1 + OP5 + OP6 + OP9 + OP12 + ROP21 + O21 + O22 + O23 + O29 + O30 + O34 + O36 + O42 + O43 + O44

김형준 한 줄 요약 : 10.12.06 06:53
E1 + E5 + E7 + E12 + SM1 + SM2 + NRA1 + NRA2 + NRA3 + NRA4 + PI5 + PI6 + PI7 + PI8 + PI9
+ PI10 + PI11 + PI12 + PI13 + PI31 + PI32 + PI33 + ROP1 + OP5 + OP6 + OP9 + OP12 + ROP21
+ O21 + O22 + O23 + O29 + O30 + O34 + O36 + O42 + O43 + O44

이게 최종 문항이요!!!!!!!!!!!!!!!!!!!!!! 넣어주세요!!

이거 알파값 .915 , 요인 분석은 안됨 ㅋㅋㅋㅋ

이미지 출처: <http://news.donga.com/more29/3/all/20160107/75770659/1>

나는 왜?

1999.2

형준이에게

‘컴퓨터 프로그래머 김형준.’ 선생가는 형준이의 이름이 이렇게
사람들에게 알려질 날이 있었는지?
익숙하게 컴퓨터를 다루는 형준이의 모습을 보면서, 그리고 방학때

...

1999. 2. 18. 목.

형준이를 참 좋아하는 선생님.



그러나... 취미생활



2009 ~ 2010



사단법인 한국동남아연구소
KISEAS Korean Institute of Southeast Asian Studies

English

연구소 소개 | 회원소개 | 활동 | 출판물 | 협력기관

연구소 소개
(사)한국동남아연구소는 어문학, 역사학, 인류학, 정치학, 경제학, 사회학, 언론학, 지리학 등을 전공한 다양한 학자들의 학제적 연대를 바탕으로 설립된 자립적 민간연구소써 동남아에 관한 학술적 연구, 정책적 제안, 사회적 확산, 국제적 교류 등을 중심적 목표로 추구하고 있습니다.

공지사항

- ▶ 2017-09-05 [환경재단] 제5회 그린아시아포럼 ...
- ▶ 2017-08-18 [한국동남아학회] 2017 연례학술대회...
- ▶ 2017-08-16 한-아세안 관계조망 국제회의 개최
- ▶ 2017-08-16 제2회 한-아세안 학술에세이 공모전 ...

서울대학교관악학생활관
SNU Gwanak Residence Halls

관악학생활관 소개 | 시설안내 | 입주/퇴거 생활안내 | 문화/자치 활동 | 가족생활관 | 커뮤니티

2014 ~ 2015

생각이 꽃피는 공간
서울대학교 관악학생활관

관악학생활관 입직원들은 미래의 인재인 여러분을 돕고 있다는 큰 자부심을 느끼고 있으며, 여러분의 생활이 건강하고 행복할 수 있도록 모든 면에서 최선을 다 할 것입니다.

2017학년도 2학기
온라인 오리엔테이션

Go! ? ?

· 기간: 2017. 8. 24(목) ~ 9. 3(일) · 대상: 생활관 거주자 전원

관악학생활관 학생상담센터
STUDENT COUNSELLING CENTER
02) 880-7979
919동 식당과 매점 사이

온라인 입주신청
ONLINE APPLICATION

온라인 오리엔테이션
ONLINE ORIENTATION

919 체력단련실
919 FITNESS CENTER
02) 881-9081
919 B동 1층

시설이용신청
FACILITY RESERVATION

관악학생활관자치회
SNU DORMITORY STUDENT COUNCIL

바로가기 메뉴 SERVICE GUIDE

고장/수리 의뢰 | 오늘의 식단 | 주요시설 연락처

비상상황시 연락처

통합관제실 02) 881-9119 (900동-906동)
합동근무실 02) 881-9019 (918동-926동)

일반공지 | 상벌공지 | MORE+ | 입주/퇴거 공지 | MORE+ | 자료실 | 안전교육 | 뉴스레터 | MORE+

[09-18] [전체] SNU 무선랜 일시 중단 안내 NEW
[09-18] [재정] 919동 세대 내 갑압밸브 설치공사 NEW
[09-15] [919 Fitness] 2017년 10월 등록 안내 - October registr...
[09-12] 2017 한울재 글로벌 푸드부스 모집 공고 NEW

[09-18] 2017학년도 2학기 추가입주대상자 8차 발표(학... NEW
[09-11] (수정)2017학년도 2학기 추가입주대상자 7차 발...
[09-04] (수정)2017학년도 2학기 추가입주대상자 6차 발...
[08-28] 2017학년도 2학기 추가입주대상자 5차 발표(학...

[02-01] [900-906동] 대학원생활관 상업용 세탁...
[02-09] [918,919,921-926] IP주소 설정 방법
[03-09] 학부생활관 주요안내사항(2016학년도 1...
[03-09] 대학원생활관 주요안내사항(2016학년도 1...

삼질의 진화

(2012.9~2015.2) 인지과학



We use empirical methods and formal tools to uncover the mechanisms of human learning and inference.

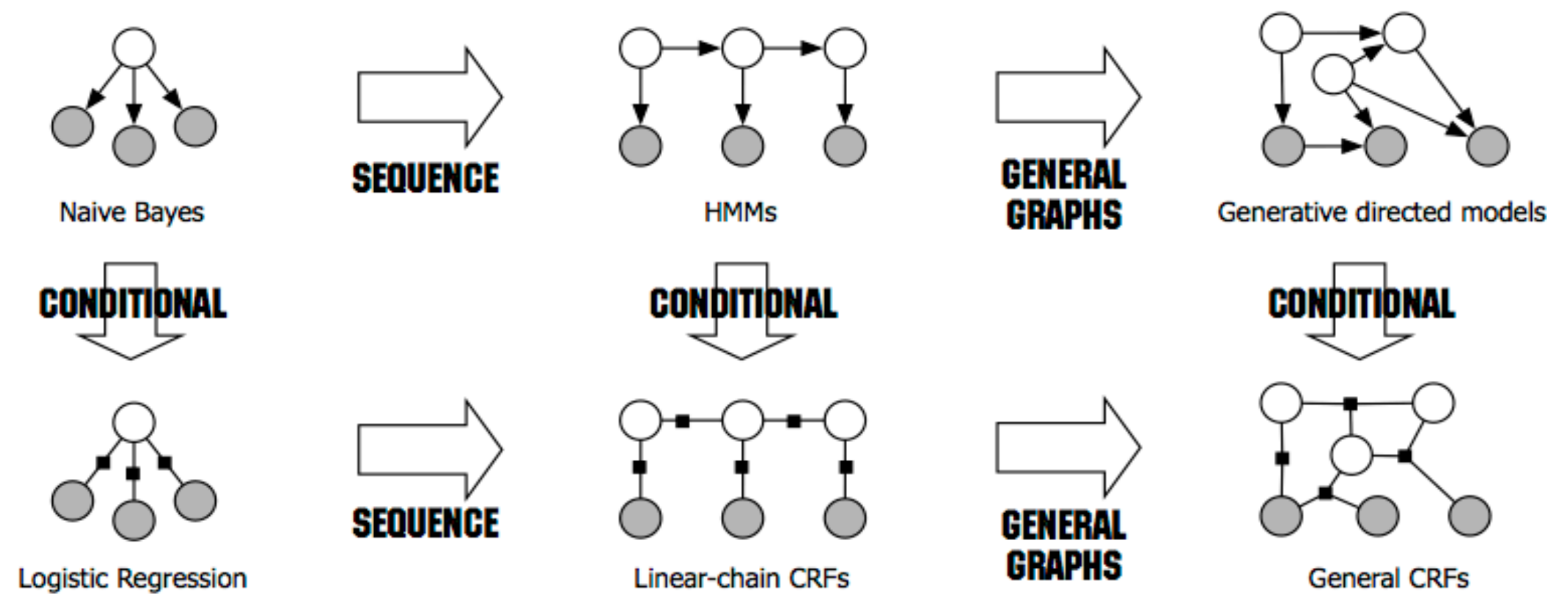
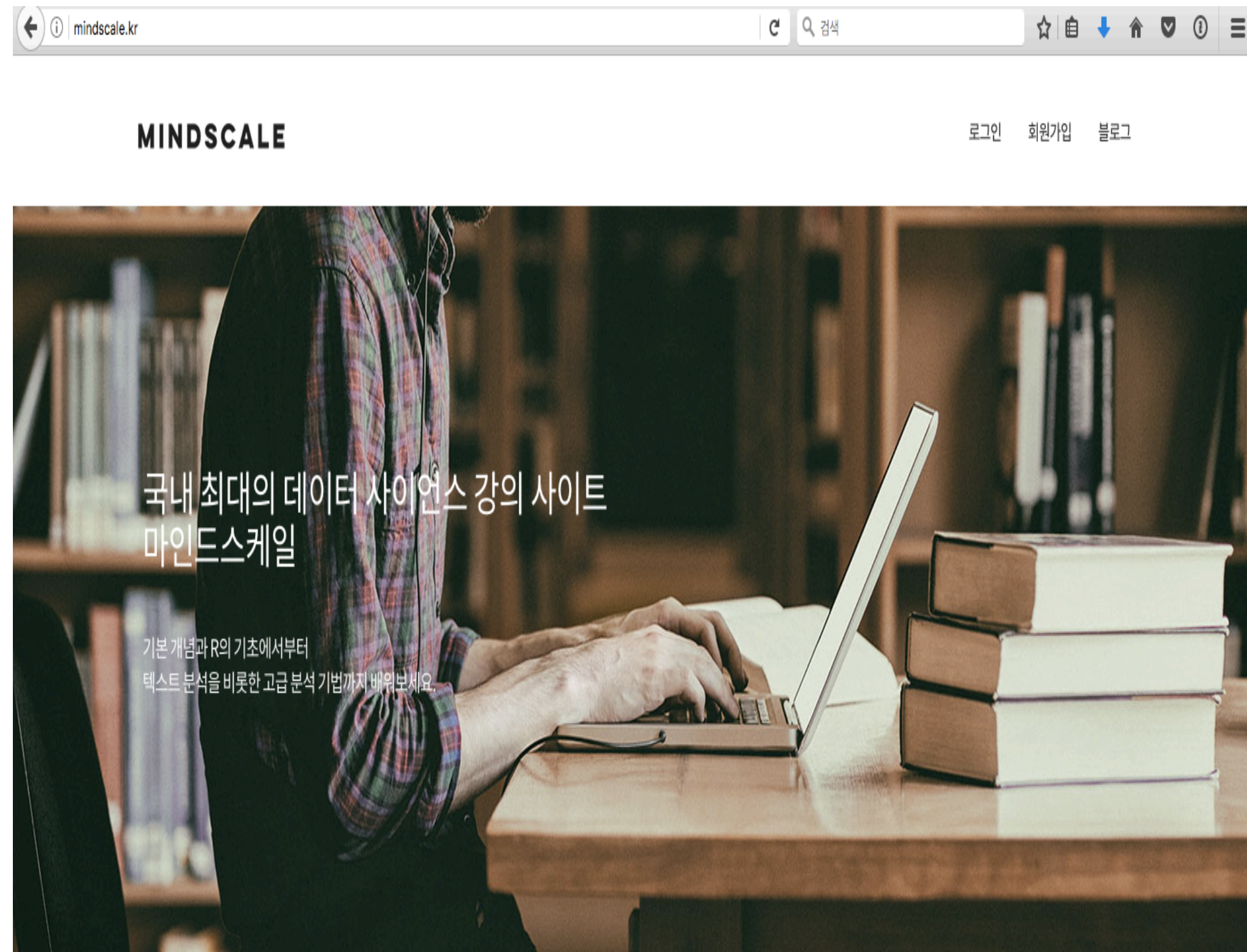


Figure 1.2 Diagram of the relationship between naive Bayes, logistic regression, HMMs, linear-chain CRFs, generative models, and general CRFs.

이미지 출처: <http://cocosci.mit.edu/> (좌), http://www.research.ed.ac.uk/portal/files/10482724/crftut_fnt.pdf (우)

삽질의 진화

(2015) MINDSCALE
: 데이터분석 온라인 교육 사이트 콘텐츠 제작
<http://mindscale.kr/>



8

(2015) R User Conference in Korea 2015
: R을 이용한 텍스트 감정분석
- 여론과 감성 발견하기

SLDA

Supervised LDA

Document response

Regression parameters

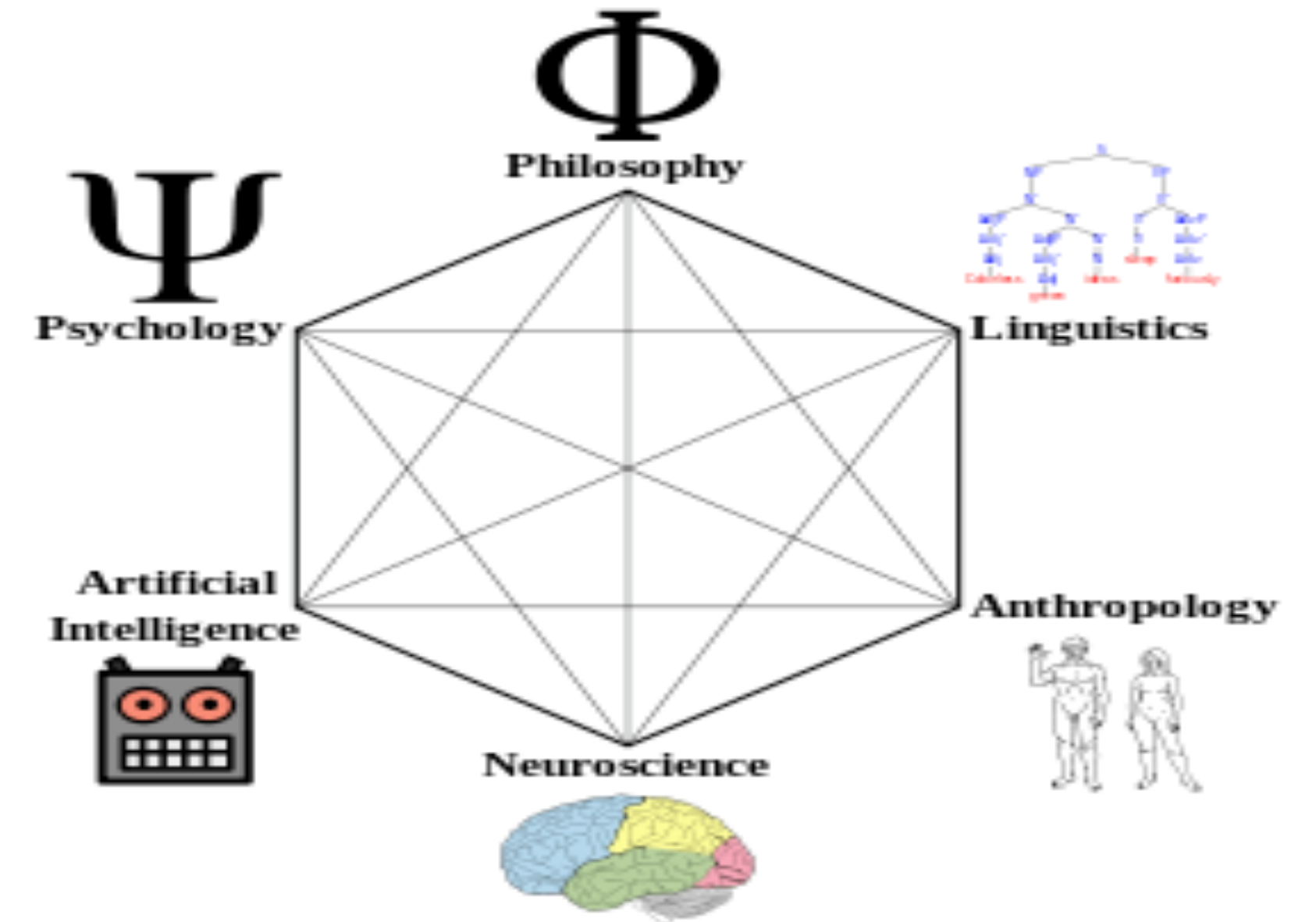
- 1 Draw topic proportions $\theta | \alpha \sim \text{Dir}(\alpha)$.
- 2 For each word
 - Draw topic assignment $z_n | \theta \sim \text{Mult}(\theta)$.
 - Draw word $w_n | z_n, \beta_{1:K} \sim \text{Mult}(\beta_{z_n})$.
- 3 Draw response variable $y | z_{1:N}, \eta, \sigma^2 \sim \text{N}(\eta^T \bar{z}, \sigma^2)$, where $\bar{z} = (1/N) \sum_{n=1}^N z_n$.

29/45

삽질의 진화



출처 : https://m.blog.naver.com/PostView.nhn?blogId=nile_parents&logNo=220510088317&categoryNo=46&proxyReferer=&proxyReferer=https%3A%2F%2Fwww.google.co.kr%2F

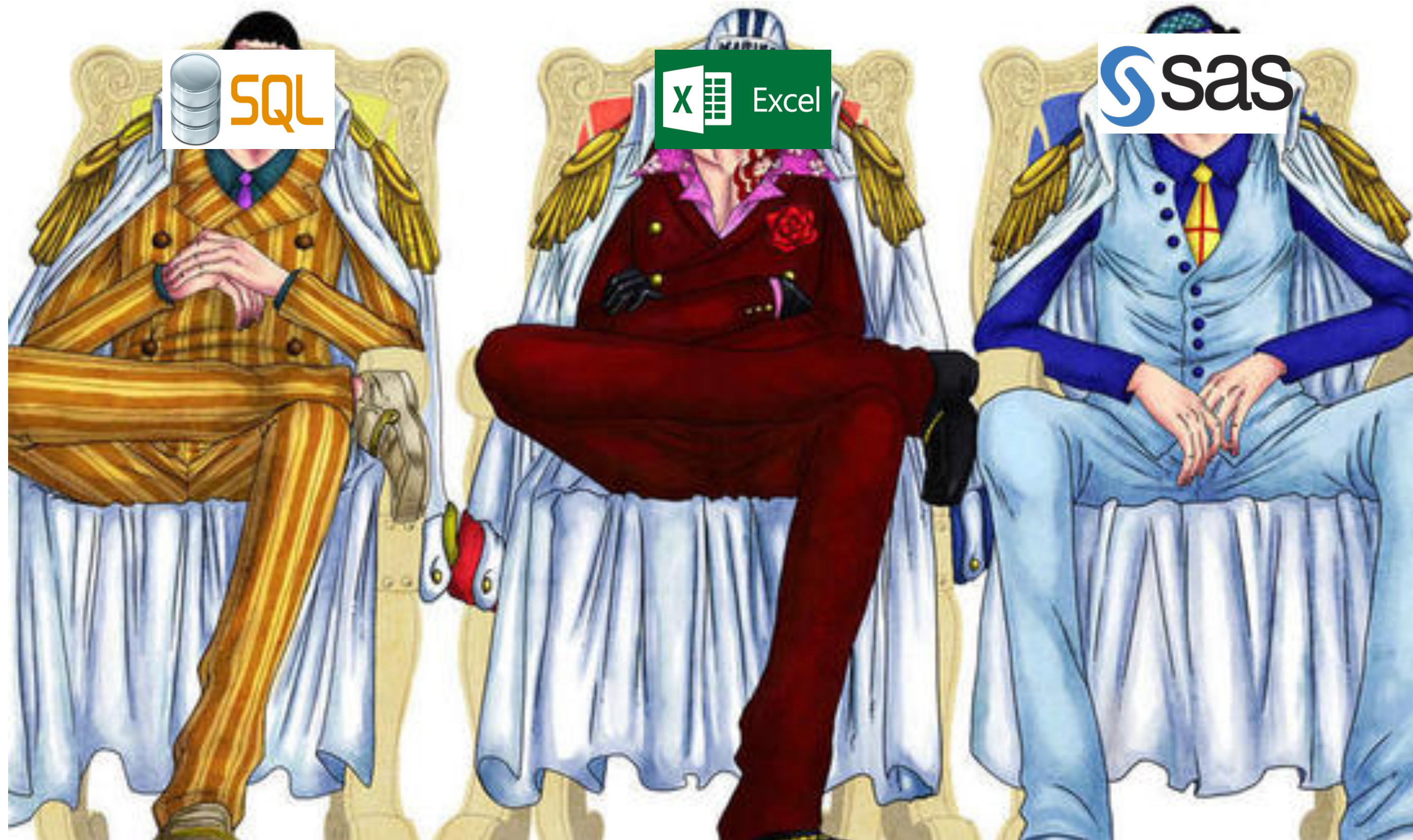


출처 : http://goos.wiki/index.php?title=%EB%82%98%EB%8A%94_%ED%96%89%EB%B3%B5%ED%95%A9%EB%8B%88%EB%8B%A4 (상)
https://en.wikipedia.org/wiki/Cognitive_science (하)

- 0 발표자 소개
- 1 인터넷 폐쇄망에서 분석 시스템 구축 (feat. 엔지니어가 없을 때)
- 2 분석 보고서 자동화
- 3 Machine Learning 삽질기 (분석 & 개발)

SQL, EXCEL, SAS

11



이미지 출처: <https://namu.wiki/w/%EC%82%BC%EB%8C%80%EC%9E%A5>

R과 Python을 하고 싶어요..

12



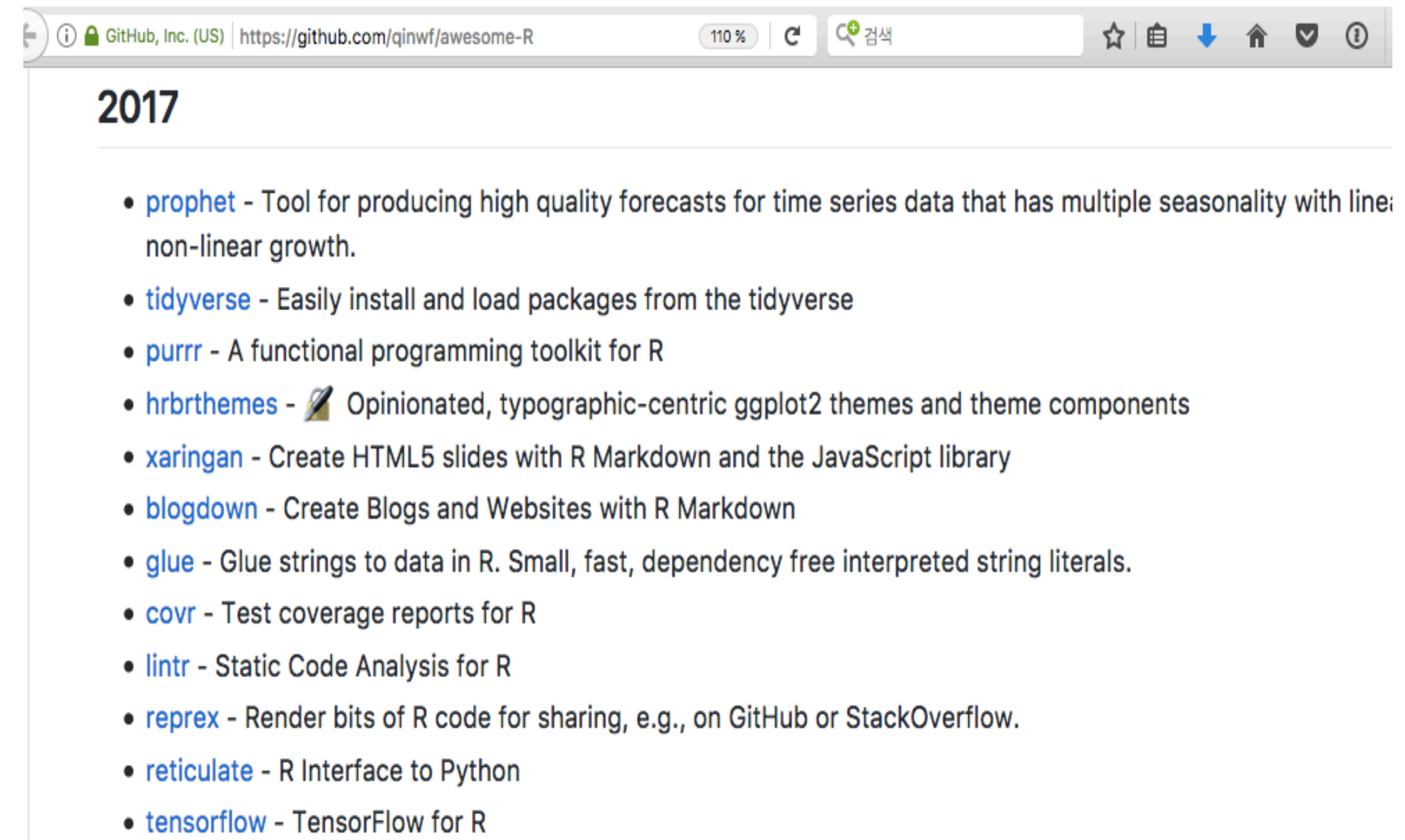
R과 Python을 하고 싶어요..

13

R

miniCRAN 이용

- 모든 패키지 다운로드
- <https://github.com/qinwf/awesome-R>
 - 크롤링
 - miniCRAN 이용 의존성 패키지 목록 추출
 - 관련 패키지 다운로드



2017

- [prophet](#) - Tool for producing high quality forecasts for time series data that has multiple seasonality with linear and non-linear growth.
- [tidyverse](#) - Easily install and load packages from the tidyverse
- [purrr](#) - A functional programming toolkit for R
- [hrbrthemes](#) - Opinionated, typographic-centric ggplot2 themes and theme components
- [xaringan](#) - Create HTML5 slides with R Markdown and the JavaScript library
- [blogdown](#) - Create Blogs and Websites with R Markdown
- [glue](#) - Glue strings to data in R. Small, fast, dependency free interpreted string literals.
- [covr](#) - Test coverage reports for R
- [lintr](#) - Static Code Analysis for R
- [reprex](#) - Render bits of R code for sharing, e.g., on GitHub or StackOverflow.
- [reticulate](#) - R Interface to Python
- [tensorflow](#) - TensorFlow for R

R과 Python을 하고 싶어요..

14

Python

conda 이용
pip 이용

- pip show 이용 (의존성 패키지 추출)
- pip download 이용 (패키지 다운로드)
- <http://www.lfd.uci.edu/~gohlke/pythonlibs/>
 - windows binary 다운로드
- conda의 경우 R 연동 문제로..



AD3, Alternating Directions Dual Decomposition.

[ad3-2.1-cp27-cp27m-win32.whl](#)
[ad3-2.1-cp27-cp27m-win_amd64.whl](#)
[ad3-2.1-cp34-cp34m-win32.whl](#)
[ad3-2.1-cp34-cp34m-win_amd64.whl](#)
[ad3-2.1-cp35-cp35m-win32.whl](#)
[ad3-2.1-cp35-cp35m-win_amd64.whl](#)
[ad3-2.1-cp36-cp36m-win32.whl](#)
[ad3-2.1-cp36-cp36m-win_amd64.whl](#)

Aiohttp, a http client/server for asyncio.

[aiohttp-2.2.5-cp34-cp34m-win32.whl](#)
[aiohttp-2.2.5-cp34-cp34m-win_amd64.whl](#)
[aiohttp-2.2.5-cp35-cp35m-win32.whl](#)
[aiohttp-2.2.5-cp35-cp35m-win_amd64.whl](#)
[aiohttp-2.2.5-cp36-cp36m-win32.whl](#)
[aiohttp-2.2.5-cp36-cp36m-win_amd64.whl](#)

Akima, interpolation of data points in a plane based on Akima's method.

[akima-2015.1.29-cp27-none-win32.whl](#)
[akima-2015.1.29-cp27-none-win_amd64.whl](#)
[akima-2015.1.29-cp34-none-win32.whl](#)
[akima-2015.1.29-cp34-none-win_amd64.whl](#)
[akima-2015.1.29-cp35-none-win32.whl](#)
[akima-2015.1.29-cp35-none-win_amd64.whl](#)
[akima-2015.1.29-cp36-cp36m-win32.whl](#)
[akima-2015.1.29-cp36-cp36m-win_amd64.whl](#)

버전 업데이트는요???

서버를 주신다면..
동기화를 해서...
관리는 누가해요??



Nobody knows...

R은 3.2.x

Python은 3.5.x

ODBC / JDBC



- 0 발표자 소개
- 1 인터넷 폐쇄망에서 분석 시스템 구축 (feat. 엔지니어가 없을 때)
- 2 분석 보고서 자동화**
- 3 Machine Learning 삽질기 (분석 & 개발)

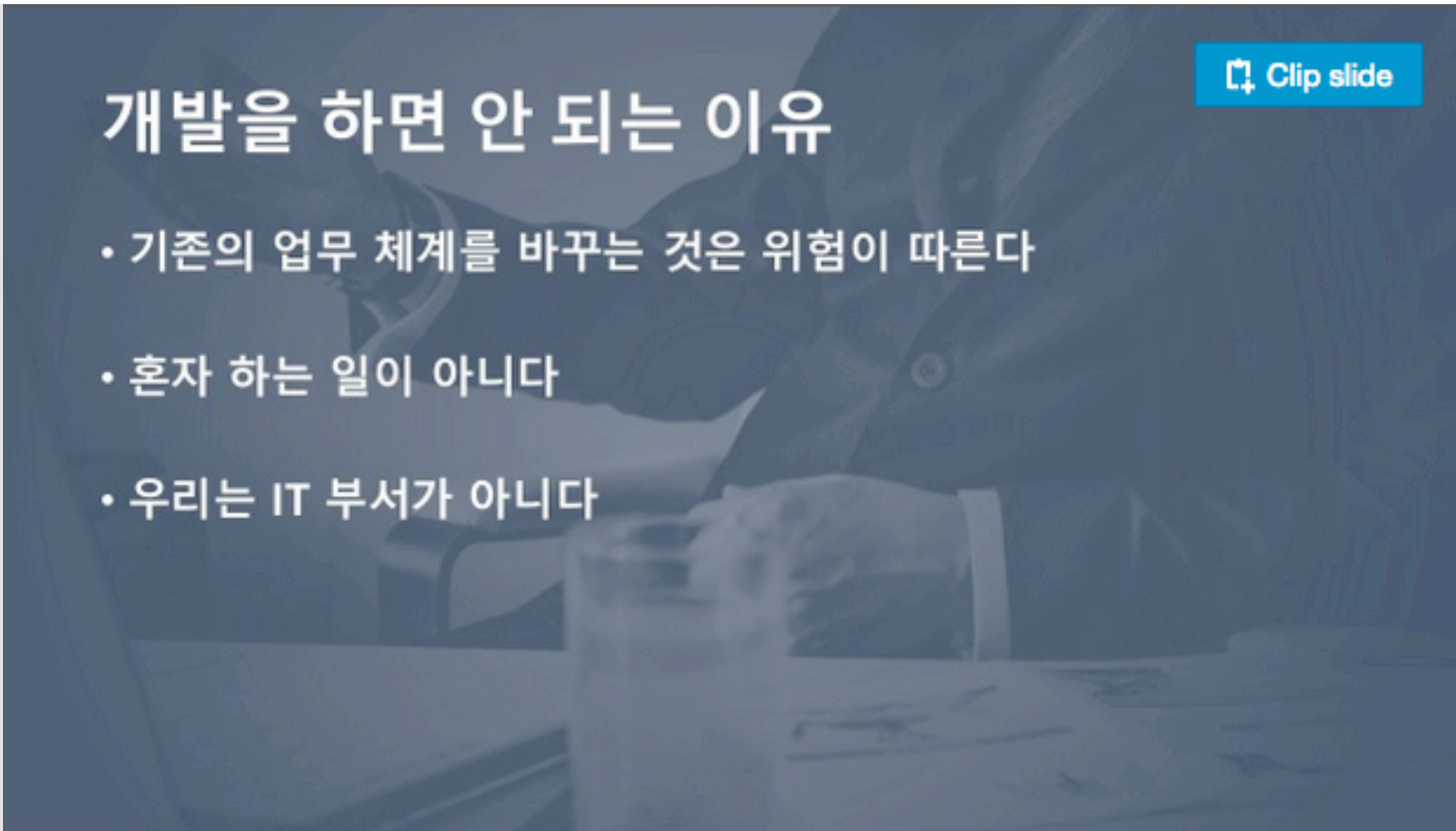


Dances with the Last Samurai

개발자 아닌 사람이
개발부서가 아닌 통계부서에서
하지 말라는 개발하는 이야기



이성용
saintdragon2@gmail.com



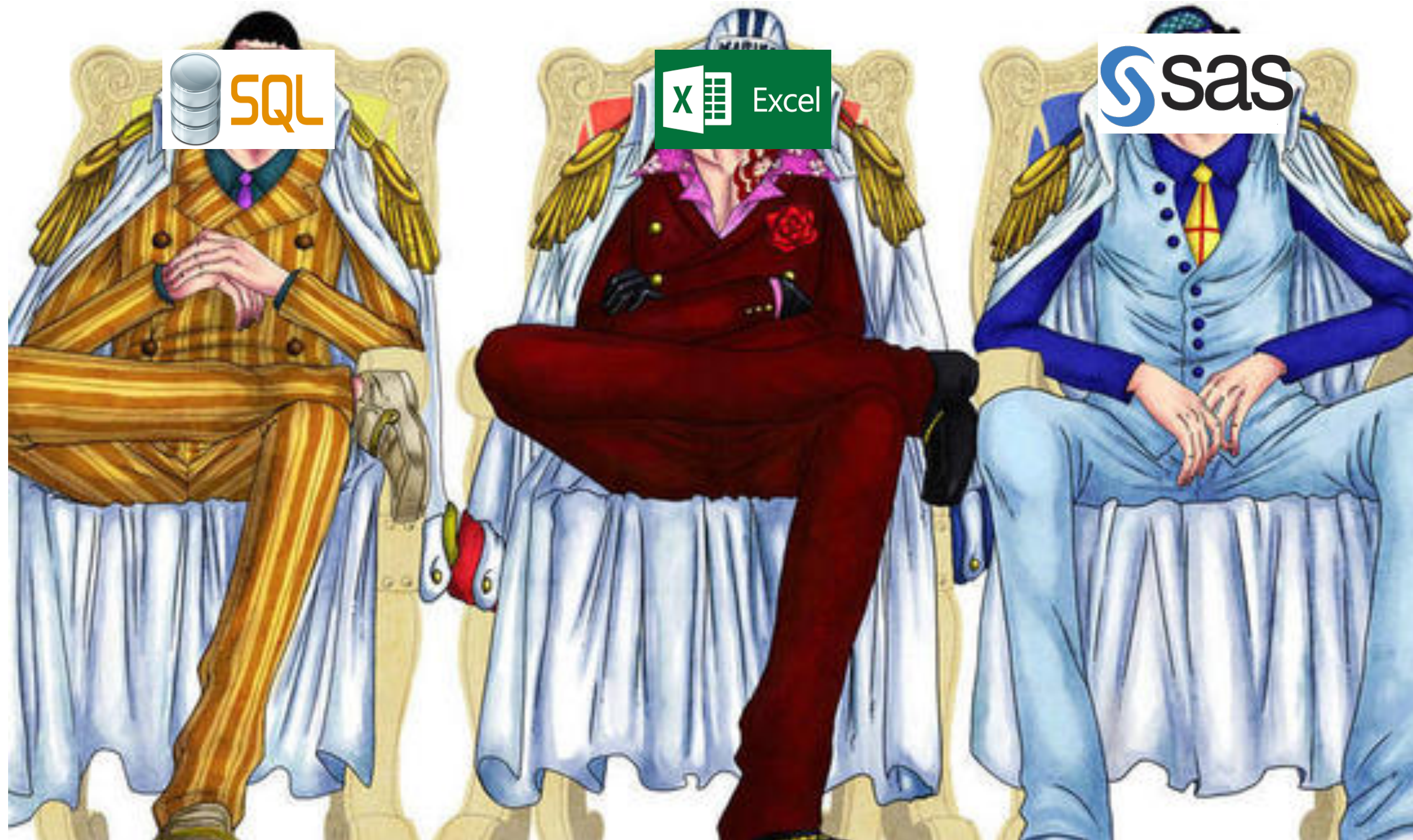
Clip slide

개발을 하면 안 되는 이유

- 기존의 업무 체계를 바꾸는 것은 위험이 따른다
- 혼자 하는 일이 아니다
- 우리는 IT 부서가 아니다

SQL, EXCEL, SAS

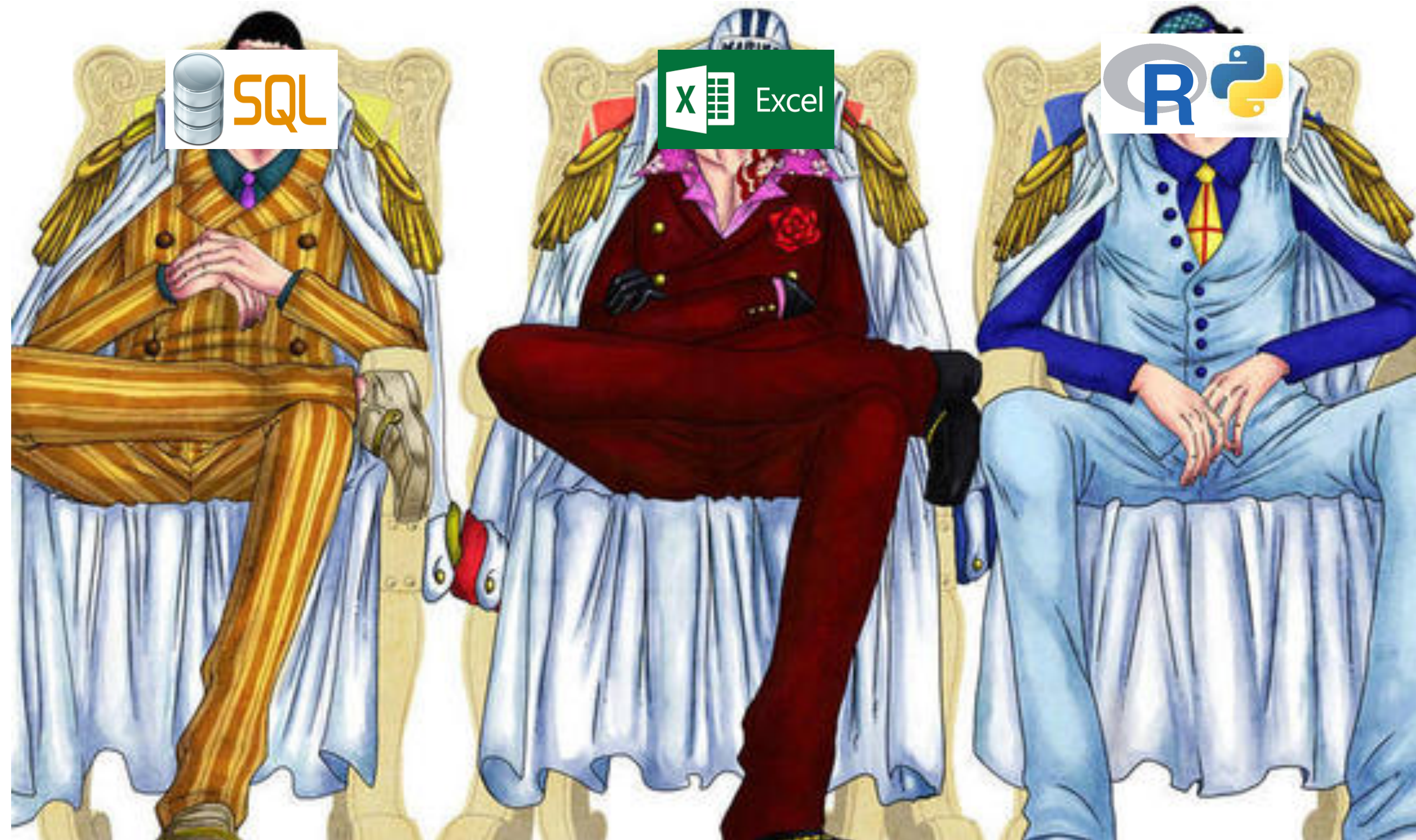
18



이미지 출처: <https://namu.wiki/w/%EC%82%BC%EB%8C%80%EC%9E%A5>

SQL, EXCEL, (R or Python)

19



이미지 출처: <https://namu.wiki/w/%EC%82%BC%EB%8C%80%EC%9E%A5>

보고서 자동화

20



이미지 출처:

<http://blog.naver.com/PostView.nhn?blogId=djfdma8&logNo=220124054628&parentCategoryNo=&categoryNo=&viewDate=&isShowPopularPosts=false&from=postView>

Datayanolja 2017

보고서 자동화

21





- 내가 엑셀 파일을 열면 문서보안에 걸려요
- 상사가 문서보안 걸린 파일을 줘요



“An eye for an eye and a tooth for a tooth”



- COM Object를 활용하자
- 엑셀을 열어서 정보를 얻자
-  : excel.link  : xlwings
- 엑셀 작업 시 보안이 안걸리게 할 수도 있음 (비밀)

보고서 자동화



한글이..

```
<!-- Y --> <!-- // Y --> <!-- TV --> <!-- // TV -->
id="articleBodyContents"> <!-- -->
N V' ϕ . LG G5 `z`
j(ATAP) i â å (Head of Creative) db ` I/C
. 2017 h t Nw . <br><br><span class="e
type=w540"><em class="img_desc">
G5 θ ü , PCó L ī
```

이미지 출처: <http://hashcode.co.kr/questions/1975/%EC%9B%B9%ED%81%AC%EB%A1%A4%EB%A7%81-%ED%95%9C%EA%B8%80-%EA%B9%A8%EC%A7%90-%EC%A7%88%EB%AC%B8%EC%9D%B4%EC%9A%94>

보고서 자동화

23





보고서 자동화

25



gdemin / excel.link

Watch 6 Unstar 19 Fork 6

Code Issues 3 Pull requests 0 Projects 0 Wiki Insights

Merge pull request #5 from soeque1/master [Browse files](#)

Applying MultiByte Character Sets

master (#5) v0.9.7 v0.9.5

gdemin committed on GitHub on 21 Sep 2016 2 parents a9a1c70 + 3a917cd commit ca87b72a4499f0091a24d5c64e62962786fbed49

Showing 1 changed file with 2 additions and 5 deletions. [Unified](#) [Split](#)

7 src/converters.cpp [View](#)

```
@@ -53,22 +53,19 @@ char *
53 53 FromBstr(BSTR str)
54 54 {
55 55     char *ptr = NULL;
56 - DWORD len;
57 56
58 57     if(!str)
59 58         return(NULL);
60 59
61 - len = wcslen(str);
62 + int len = WideCharToMultiByte(CP_ACP, 0, str, -1, NULL, 0, NULL, NULL);
63 61
64 62     if(len < 1)
65 63         len = 0;
```

보고서 자동화

26



PYTHON FOR EXCEL. FREE & OPEN SOURCE.

Latest Release: v0.11.4 (Jul 23, 2017): Add-in & SQL Extension

The screenshot shows two windows side-by-side. On the left is an Excel spreadsheet with columns A-E and rows 1-6. Cells A1-B2 contain the values 1, 2, 3, 4. On the right is the Jupyter QtConsole window showing the following Python code and output:

```
In [2]: import xlwings as xw
In [3]: xw.Range('A1').value = [[1, 2],[3, 4]]
In [4]: xw.Range('A1').expand().value
Out[4]: [[1.0, 2.0], [3.0, 4.0]]
In [5]: xw.books.active
Out[5]: <Book [Workbook1]>
```

보고서 자동화

27

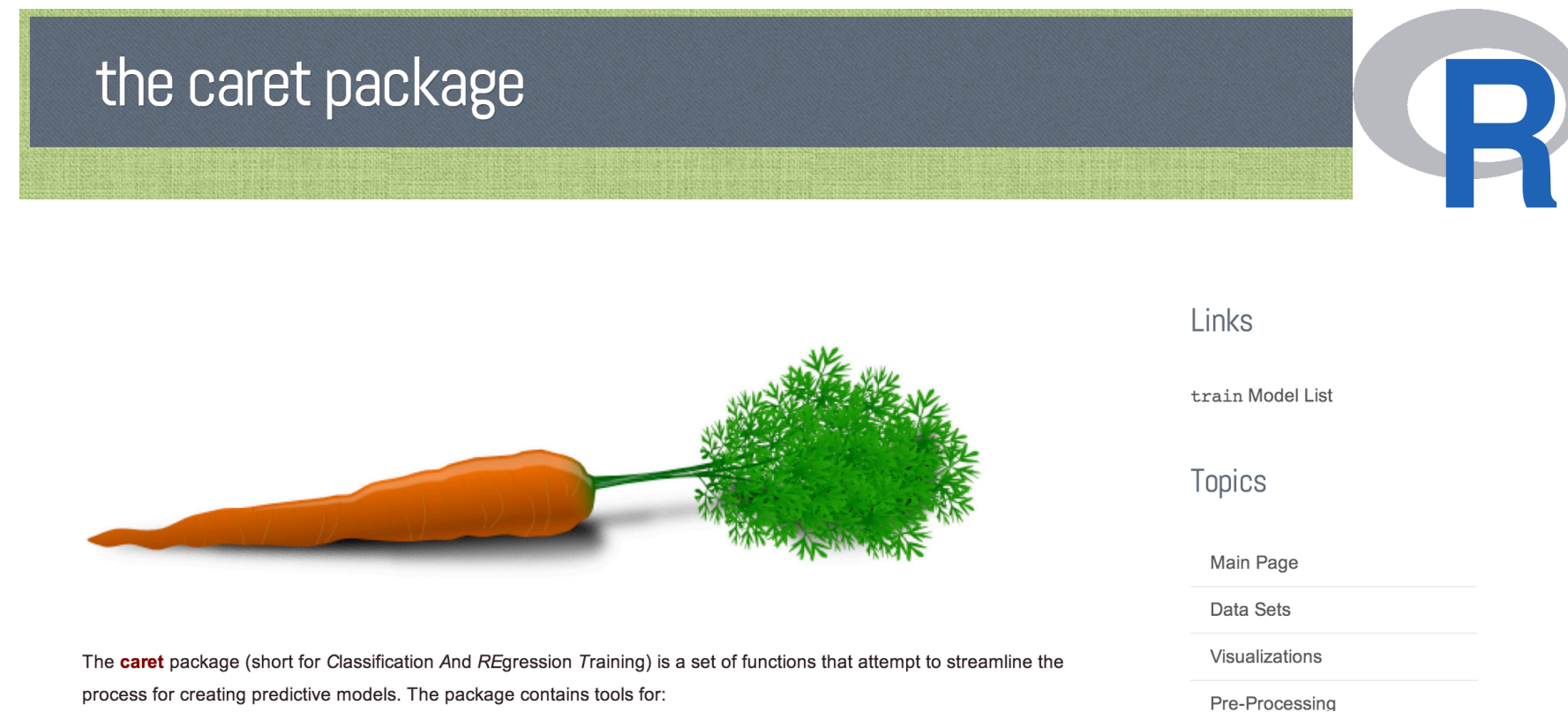


- 0 발표자 소개
- 1 인터넷 폐쇄망에서 분석 시스템 구축 (feat. 엔지니어가 없을 때)
- 2 분석 보고서 자동화
- 3 **Machine Learning 삽질기 (연구 & 개발)**

Machine Learning 연구

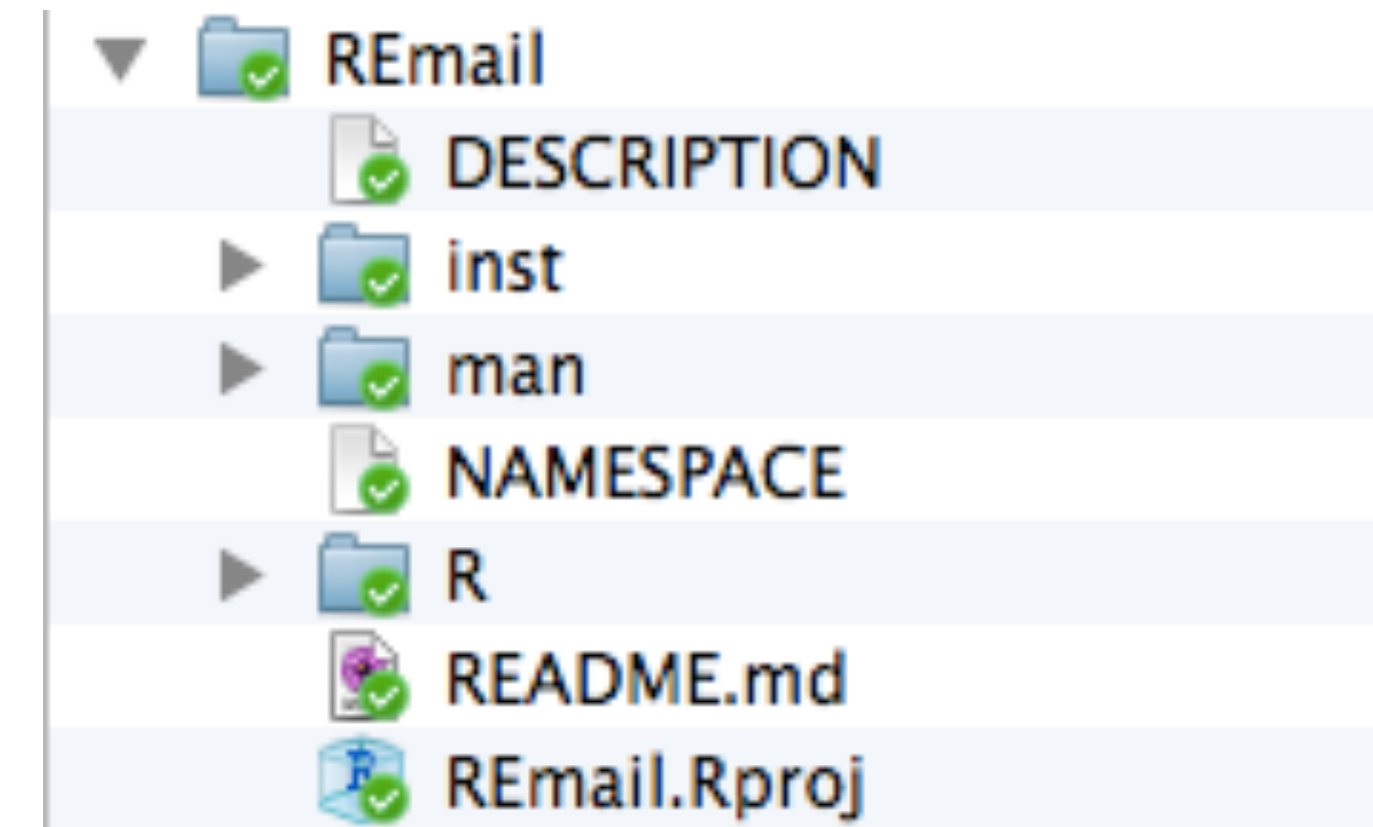
29

많은 기능이 지원되는 caret에
새로운 알고리즘 추가



VS

별도 라이브러리 생성



장점:

- 타 알고리즘 비교 가능
- 병렬처리 등 다양한 기능 구현 쉬움

단점:

- 기능 구현 시간이 느려
- 라이선스 이슈 (GPL)

장점:

- Tech 덜 신경
- 특정 알고리즘에 맞는 아키텍처

단점:

- 알고리즘 일반화가 어려움

데이터 이슈 및 고려점들

- 머신러닝에 적합한 데이터 형태(Input)
: 6개월내, 3개월내, 1개월내 → 거래별, 월별
- 그럼에도 불구하고 기존 방법론 대비 변별력 향상
- 머신러닝 시 경계할 점들
: 상관 높은 변수들이 많이 선택될 위험성
시간이 지남에 따라 변별력이 낮아짐 (과거와 미래의 패턴이 달라짐)
계산 시간을 어떻게 줄일 것인가(하드웨어 VS 소프트웨어)

Machine Learning 개발

31



Machine Learning 개발 및 배포

32



- 알고리즘 확장성
- 오류처리
- 라이선스

- 호환성
- 세션 처리
- 웹프레임워크



docker

- 가상 Windows 미지원

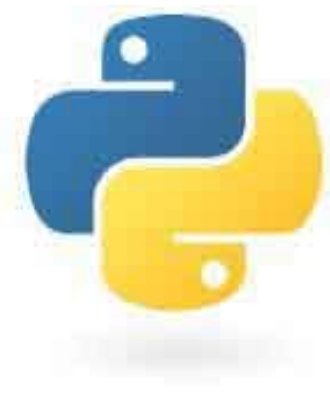


GitHub

- 서버 좀 ...

Machine Learning 개발

33



장점:

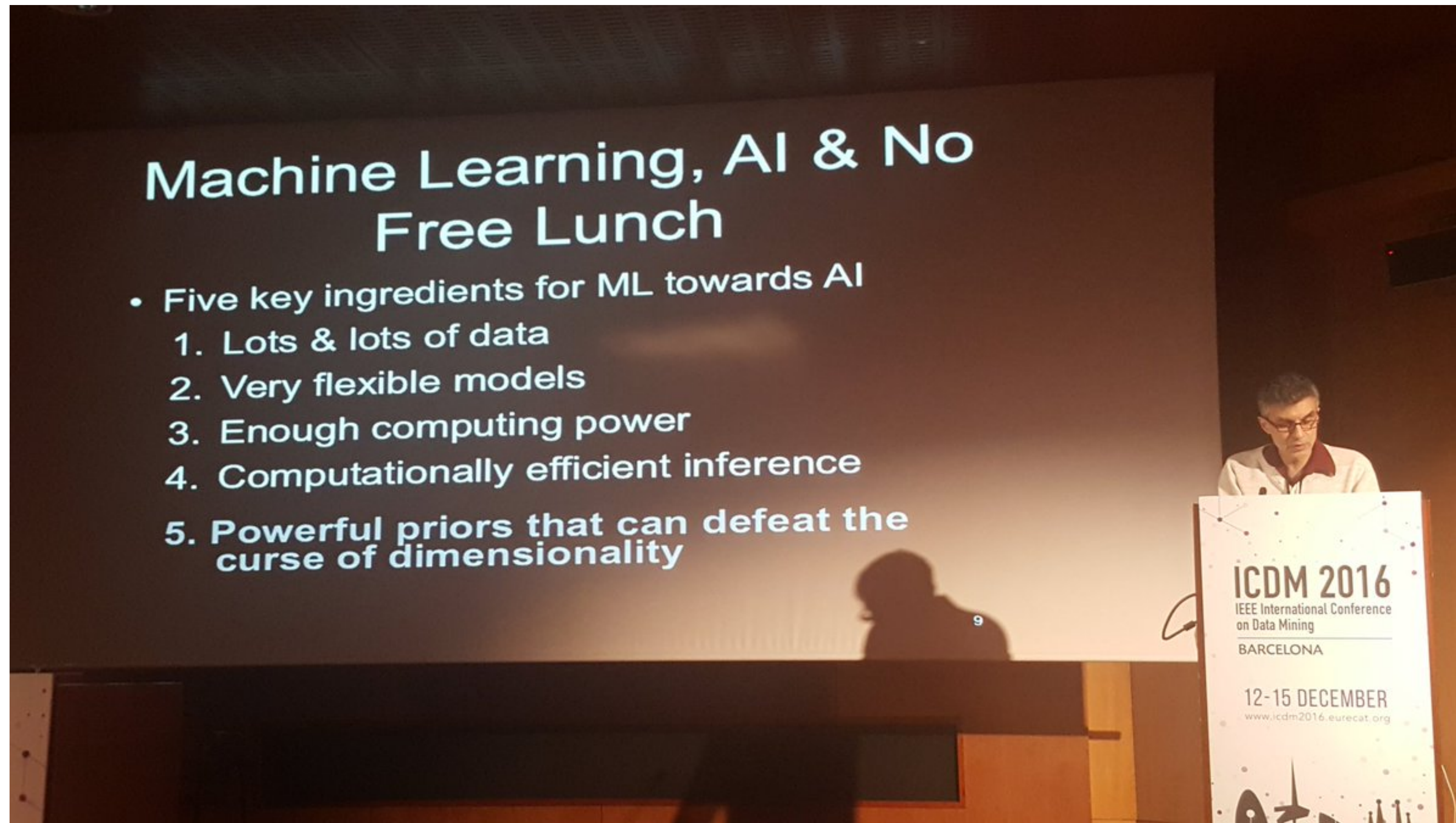
- 라이선스에 비교적 자유로움
- 개발 관련 지원 라이브러리들

단점:

- 언어 진입장벽
- Only Python만의 한계
- 혼자서 개발이 가능한가?

Machine Learning

34



이미지 출처: https://twitter.com/ibeis_org/status/808591224873558017

더 생각해볼 문제들

- 알고리즘 연구도 중요하지만 데이터 형태가 더 중요할 수도
- 설명 가능한 모형 VS 예측에만 중점인 모형
- 서로 장·단이 존재하는 모형을 어떻게 활용할 것인가?
- 변별력이 높은 모형이 과연 좋은 모형인가?
- Machine Learning 솔루션 개발은 바람직한가?
- 모형 연구와 솔루션 개발은 꼭 분리되어야 하는가?

THANK YOU FOR WATCHING