

Document as a Data

Tidyverse Korea

이광춘

Table of Contents

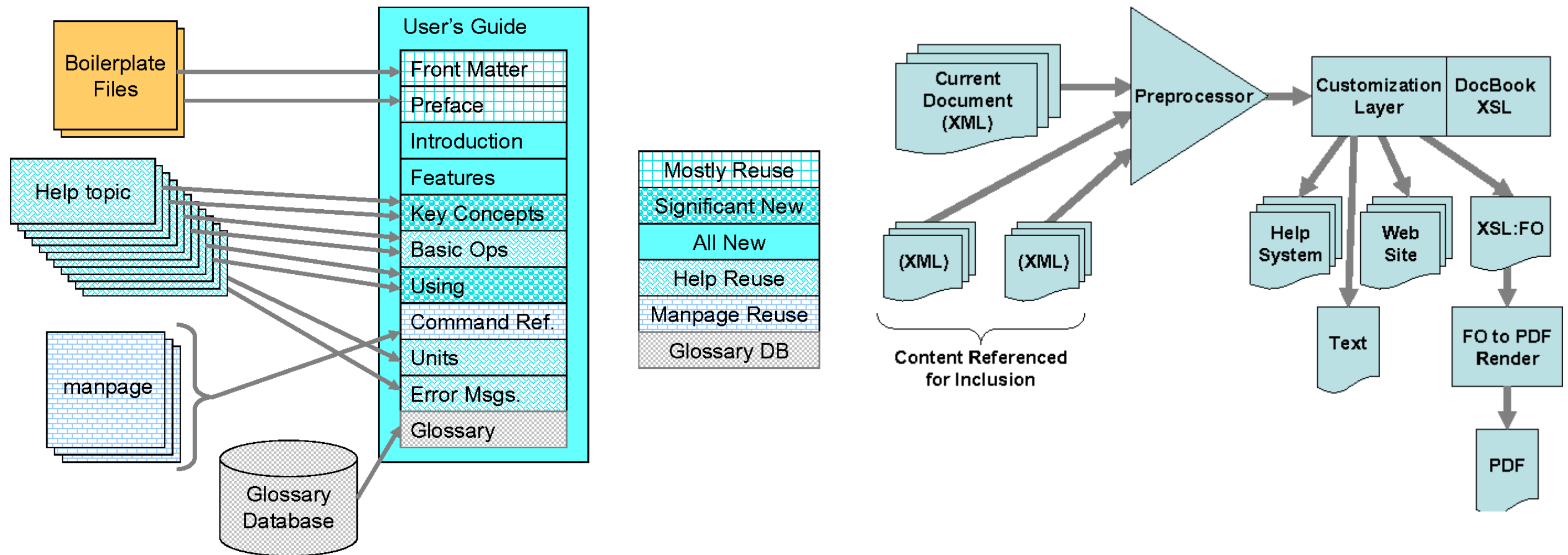
I. Document Anatomy

II. Compendium

III. OCR

Document Anatomy – Help Manual

Anatomy of Hewlett-Packard User Manual Documents

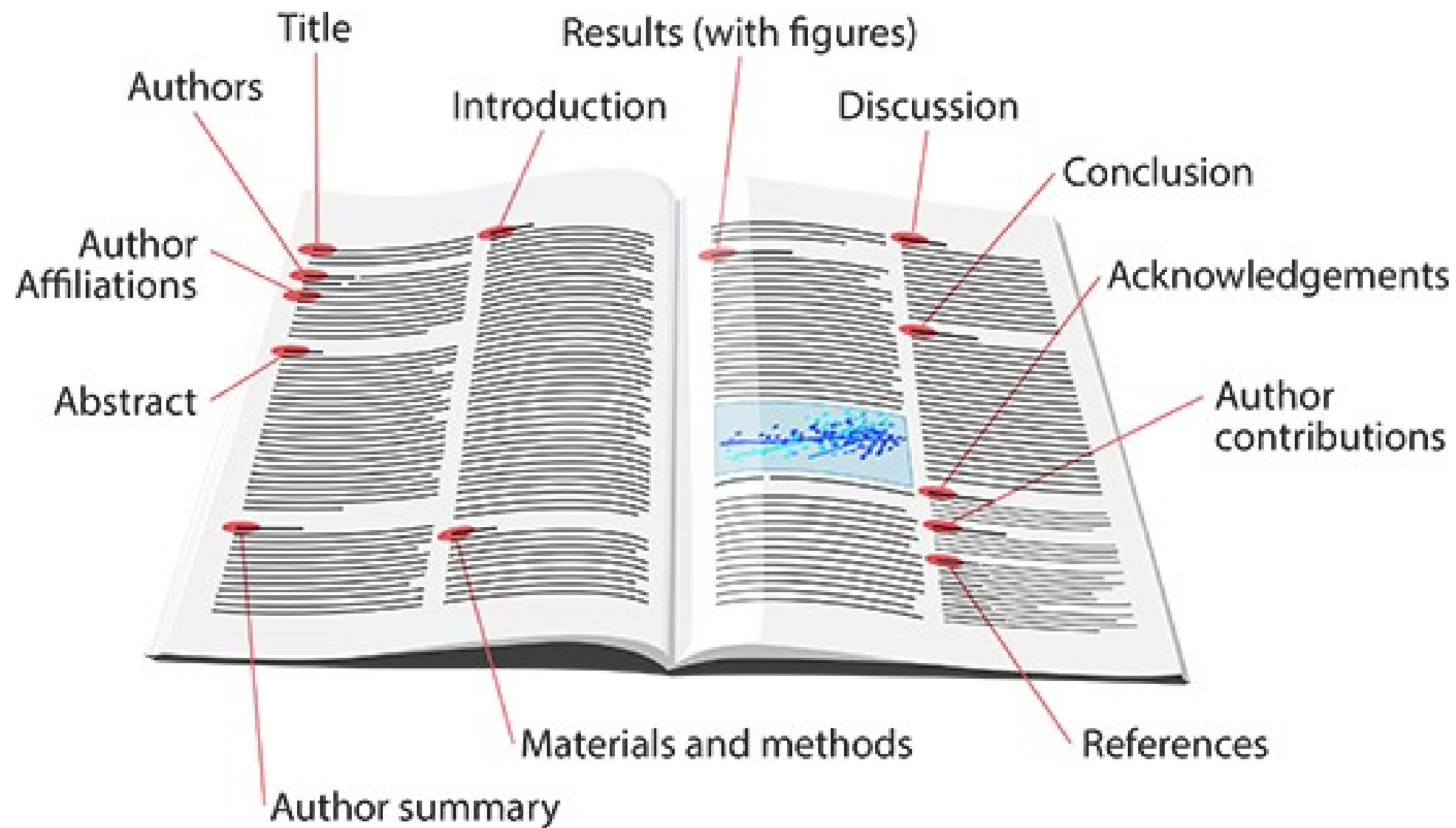


[Kathy Haramundanis](#)

[and Larry Rowland \(Hewlett-Packard Company\), "Experience paper: a content reuse documentation design experience", SIGDOC '07 Proceedings of the 25th annual ACM international conference on Design of communication Pages 229-233, El Paso, Texas, USA — October 22 - 24, 2007](#)

Document Anatomy

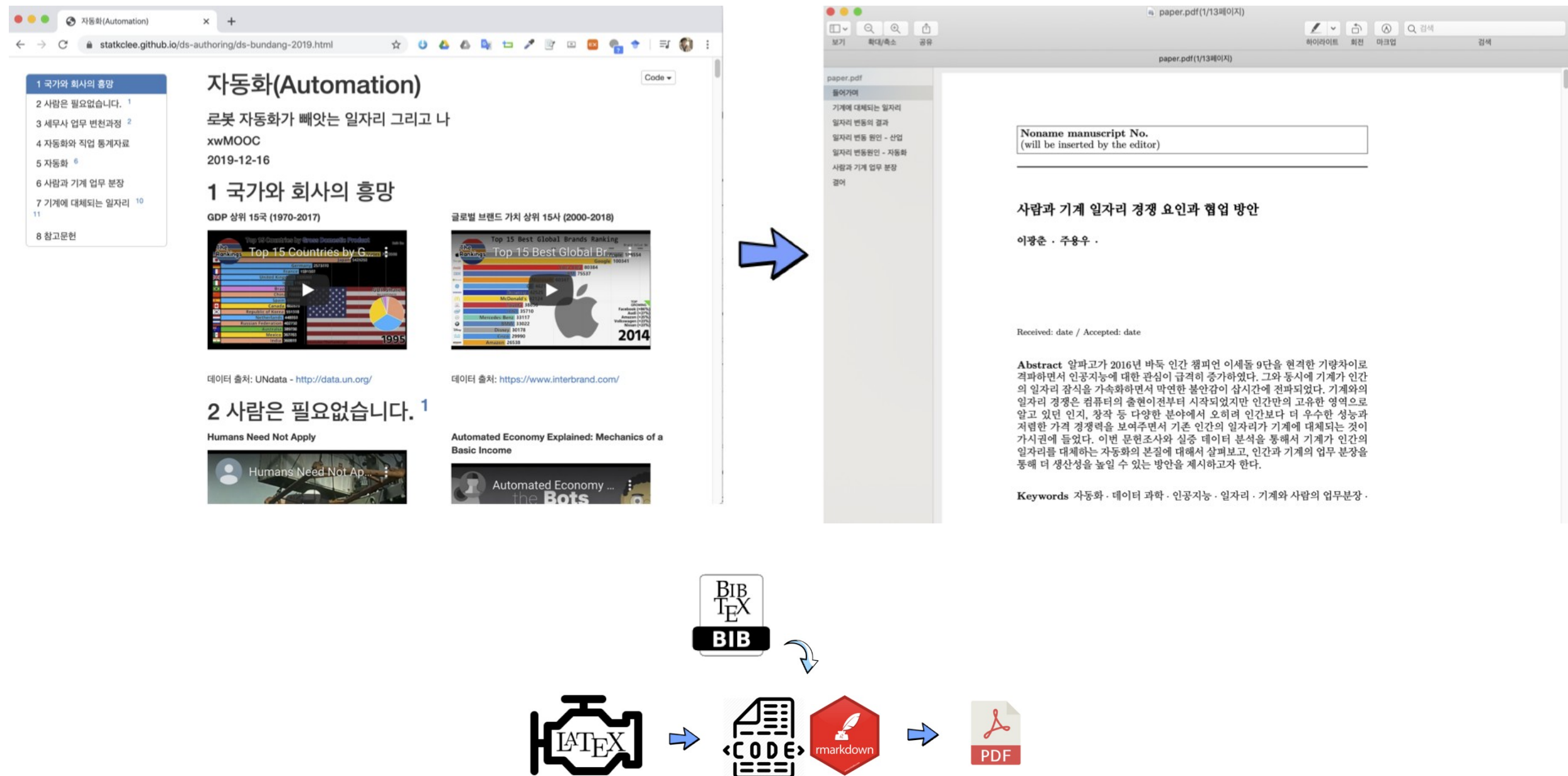
Anatomy of a research paper



출처: <https://csevilla.unepportfolio.org/2017/03/26/what-is-a-scientific-paper/>

Document Anatomy - Papers

Creating a research paper through Coding



출처: https://statklee.github.io/comp_document/automation-kasdba.html

Compendium

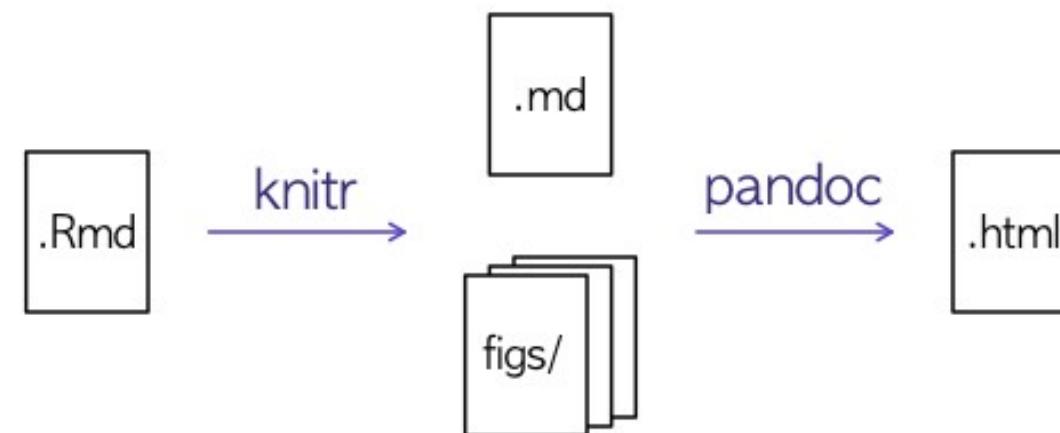
Document as a Code

```
.
├── .Rprofile
├── .gitignore
├── README.md
├── analysis
│   ├── archive
│   └── markdown
├── data
│   ├── documentation
│   ├── handmade
│   ├── html_reports
│   ├── processed
│   ├── public
│   └── source
├── etl
├── publish
├── scratch
├── viz
└── {{cookiecutter}}.Rproj
```

Version Control

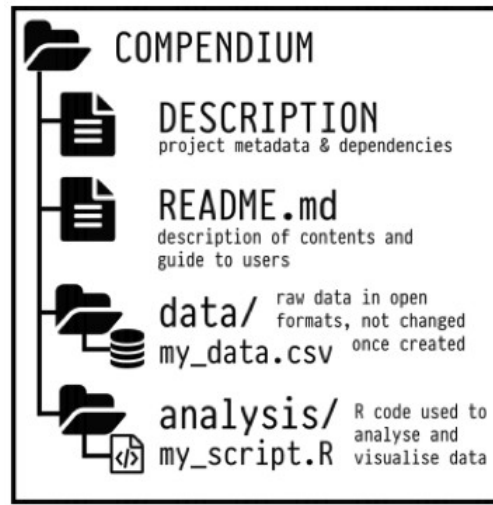


Literate Programming

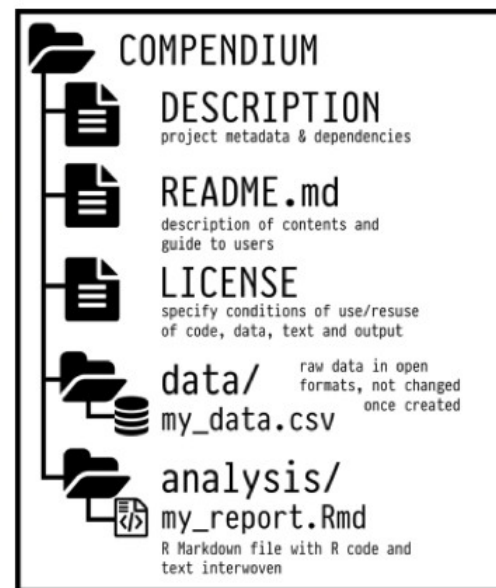


Compendium – Reproducible Research

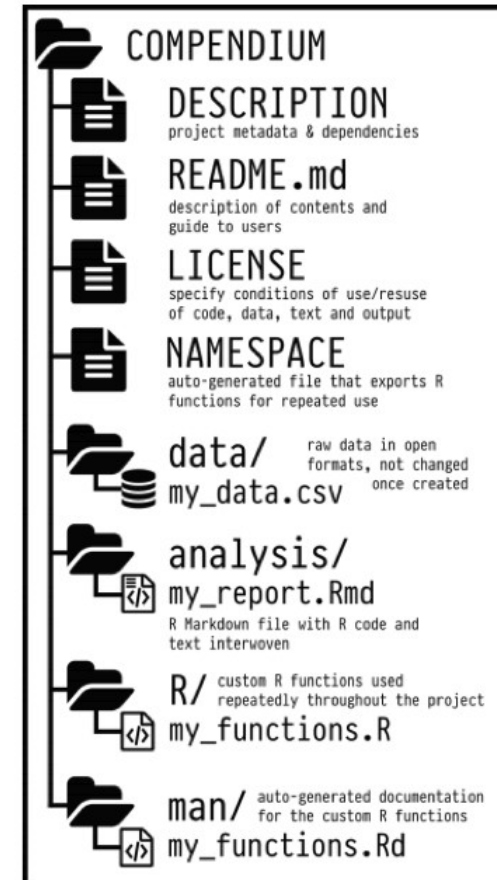
데이터 분석



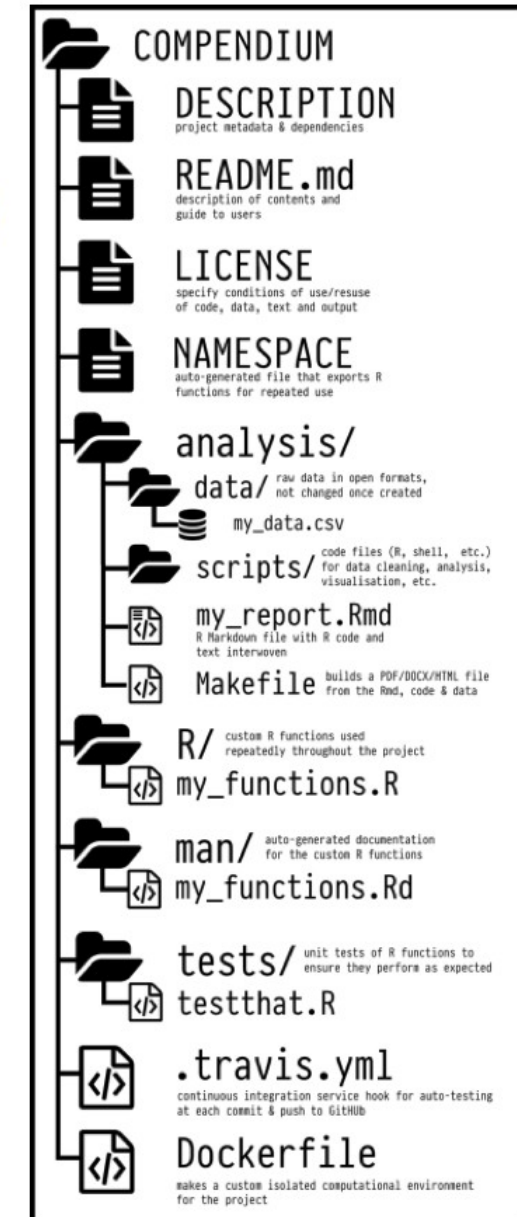
분석 보고서



분석 패키지



재현가능한 패키지



출처: https://statklee.github.io/comp_document/cd_compendium.html

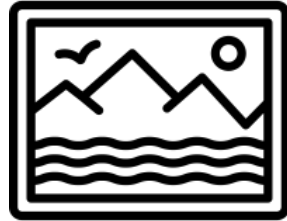
OCR

OCR (Optical Character Recognition)

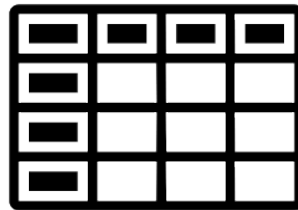
이미지 형태의 문서를 텍스트 형태의 문서로 변환해주는 기술

Lorem Ipsum

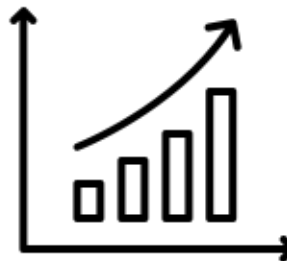
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque vitae fermentum ipsum, sodales consequat leo. Praesent rhoncus nunc est, eget iaculis eros malesuada porttitor. Aenean lectus diam, mollis ac aliquam laoreet, ornare vestibulum enim. Duis hendrerit felis ac elementum commodo. Proin vitae scelerisque velit, eu elementum tortor. Praesent dignissim nulla dui, vel malesuada quam pellentesque in. Sed porttitor tincidunt pharetra. Cras rutrum leo at mi lobortis, sed suscipit nulla porttitor. Pellentesque vitae lacus malesuada, aliquet tellus at, aliquam dolor.



Vestibulum pellentesque orci a dolor lacinia, fringilla scelerisque nisi luctus. Pellentesque eget volutpat arcu. Aenean rutrum neque metus, sit amet ultricies arcu imperdiet nec. Donec ultricies volutpat ligula, non suscipit eros pretium at. Maecenas condimentum convallis volutpat. Quisque et sodales nisi. Fusce semper tincidunt enim et ultricies.

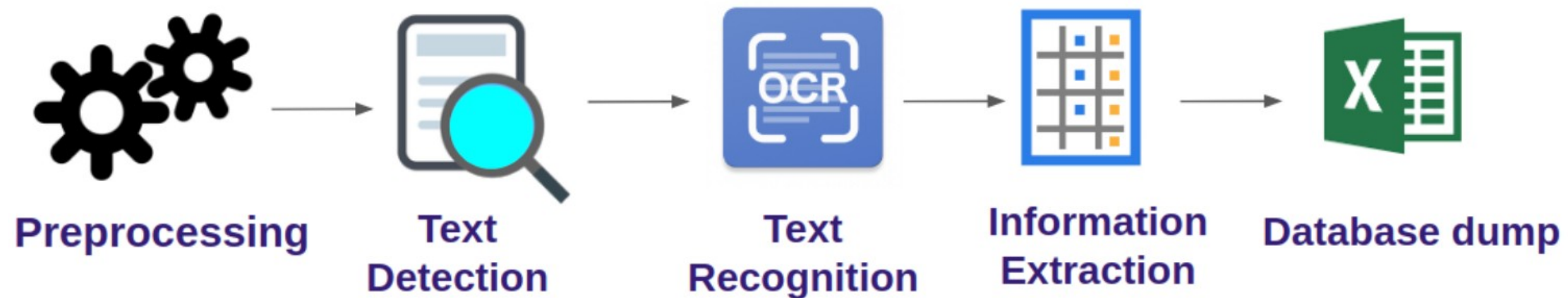


Vestibulum pellentesque orci a dolor lacinia, fringilla scelerisque nisi luctus. Pellentesque eget volutpat arcu. Aenean rutrum neque metus, sit amet ultricies arcu imperdiet nec. Donec ultricies volutpat ligula, non suscipit eros pretium at. Maecenas condimentum convallis volutpat. Quisque et sodales nisi. Fusce semper tincidunt enim et ultricies.



- 텍스트
- 이미지
- 표
- 그래프
- 문서구조
- 참고문헌
- ...

OCR 프로세스



<원본 이미지>

문서확인번호 1458- -9 - 1/2

이 초본은 개인별 주민등록표의 원본 내용과 동일
없음을 증명합니다.
담당자: 전화 02-351-5188
신청인:안 (19 * * *)
용도 및 목적: 2016년 월 일

주민등록표
(초본)

서울특별시 은평구청장

성명(한자)	인	(安)	주민등록번호	80	-2				
번호	인	거	사	탈	변	경	내	용	
*** 공 단 ***									
"주민등록번호 생략내역 없음"									
번호	주	소	전	입	일	/	변	동	
번	동	지	출	생	년	월	일	세	
1	서울특별시 은평구	15	-	출생등록				의	
2	서울특별시 은평구	1982-10-30	1982-10-30	전입				의	
3	서울특별시 은평구	1984-08-24	1984-08-24	전입				의	
4	서울특별시 은평구	1989-10-28	1989-10-28	전입				의	
5	서울특별시 은평구	1991-05-31	1991-05-30	전입				의	
6	서울특별시 은평구	1995-11-17	1995-11-17	전입				의	
7	서울특별시 은평구	2000-09-26	2000-09-26	전입				의	
8	[은평구통번령지초제3조제4항제거등변경]								
9	서울특별시 은평구	2002-09-01	2002-09-01	통번변경				의	
10	서울특별시 은평구	2004-11-06	2004-11-06	전입				의	

서울특별시 은평구청장

<전처리 이미지>

문서확인번호 1458- -9 - 1/2

이 초본은 개인별 주민등록표의 원본 내용과 동일
없음을 증명합니다.
담당자: 전화 02-351-5188
신청인:안 (19 * * *)
용도 및 목적: 2016년 월 일

주민등록표
(초본)

서울특별시 은평구청장

성명(한자)	인	(安)	주민등록번호	80	-2				
번호	인	거	사	탈	변	경	내	용	
*** 공 단 ***									
"주민등록번호 생략내역 없음"									
번호	주	소	전	입	일	/	변	동	
번	동	지	출	생	년	월	일	세	
1	서울특별시 은평구	15	-	출생등록				의	
2	서울특별시 은평구	1982-10-30	1982-10-30	전입				의	
3	서울특별시 은평구	1984-08-24	1984-08-24	전입				의	
4	서울특별시 은평구	1989-10-28	1989-10-28	전입				의	
5	서울특별시 은평구	1991-05-31	1991-05-30	전입				의	
6	서울특별시 은평구	1995-11-17	1995-11-17	전입				의	
7	서울특별시 은평구	2000-09-26	2000-09-26	전입				의	
8	[은평구통번령지초제3조제4항제거등변경]								
9	서울특별시 은평구	2002-09-01	2002-09-01	통번변경				의	
10	서울특별시 은평구	2004-11-06	2004-11-06	전입				의	

서울특별시 은평구청장

<잘라낸 영역>

문서 종류

문서확인번호 1458- -9 - 1/2

이 초본은 개인별 주민등록표의 원본 내용과 동일
없음을 증명합니다.
담당자: 전화 02-351-5188
신청인:안 (19 * * *)
용도 및 목적: 2016년 월 일

주민등록표
(초본)

서울특별시 은평구청장

전화번호

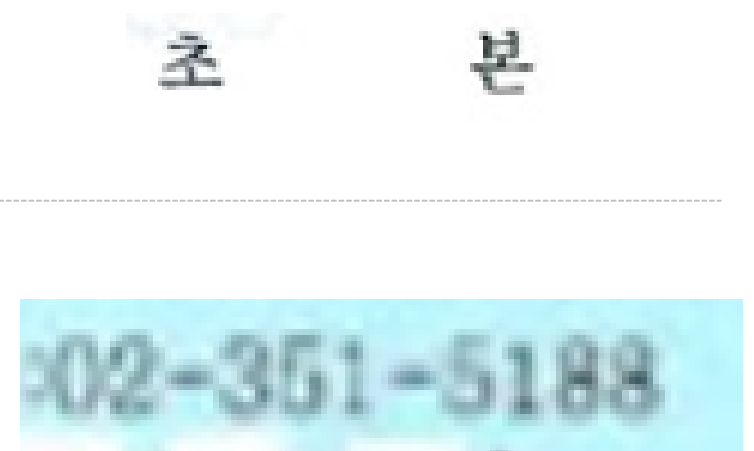
문서확인번호 1458- -9 - 1/2

이 초본은 개인별 주민등록표의 원본 내용과 동일
없음을 증명합니다.
담당자: 전화 02-351-5188
신청인:안 (19 * * *)
용도 및 목적: 2016년 월 일

주민등록표
(초본)

서울특별시 은평구청장

<글자 외곽선을 통해 추출한 이미지>



감사합니다.