Document as a Data

Tidyverse Korea 이광춘, 2020-11-12

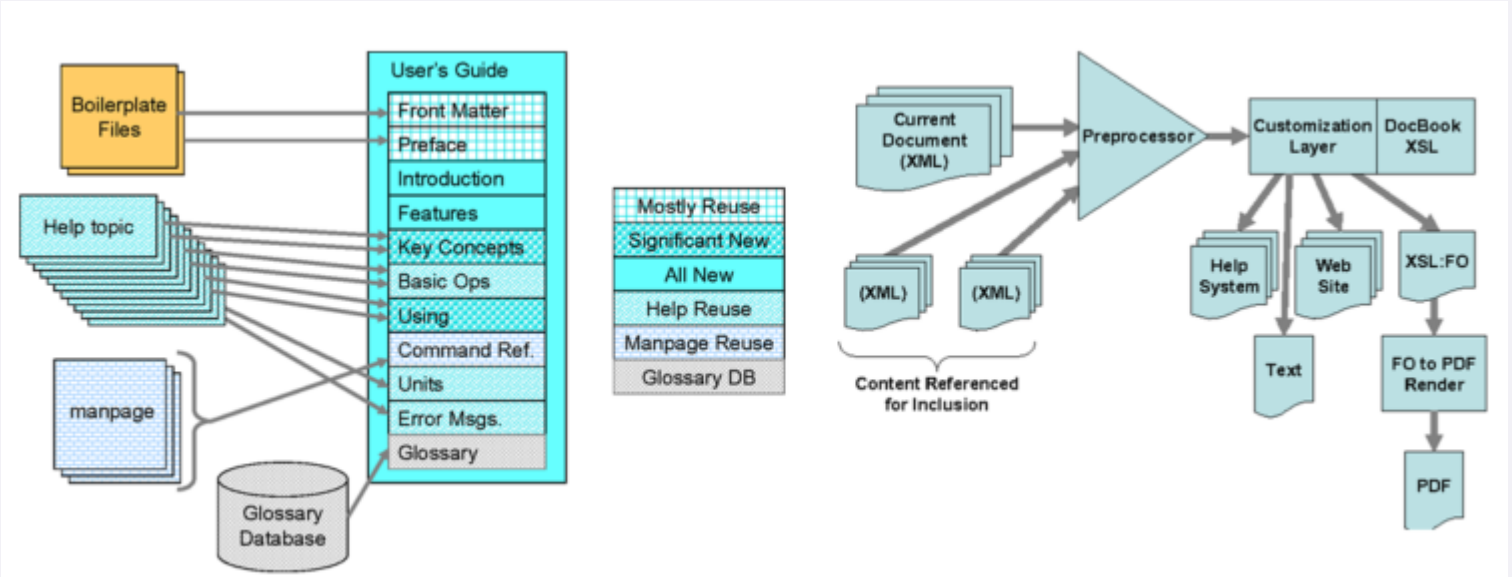
발표 형식 aringan

1. Document Anatomy
2. Compendium
3. OCR

image credit: 산림청, https://www.forest.go.kr/

Anatomy of HP User Manual Documents

HP Manual

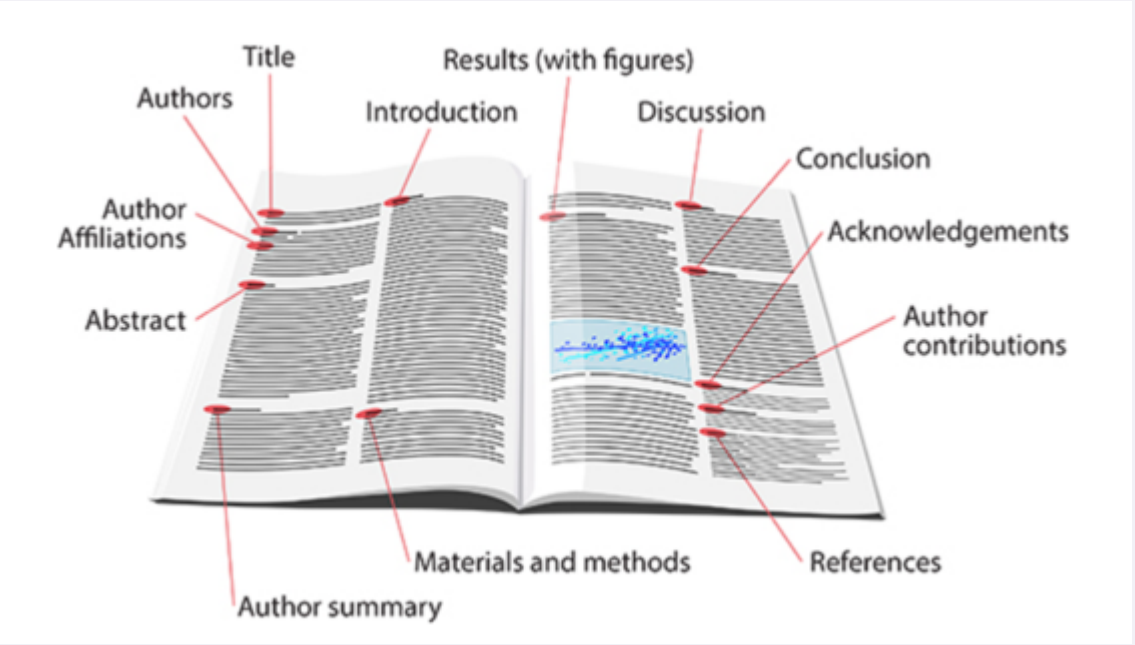


Kathy Haramundanis and Larry Rowland (Hewlett-Packard Company), "Experience paper: a content reuse documentation design experience", SIGDOC '07 Proceedings of the 25th annual ACM international conference on Design of communication Pages 229-233, El Paso, Texas, USA — October 22 - 24, 2007

Anatomy of a research paper

HP Manual

Paper



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

Creating a research paper through Coding

HP Manual

Paper

Journal

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

Creating a research paper through Coding

HP Manual

Paper

Journal

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

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

Compendium – Reproducible Research

HP Manual

Paper

Journal

Compendium

R.R.

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

OCR (Optical Character Recognition)

HP Manual

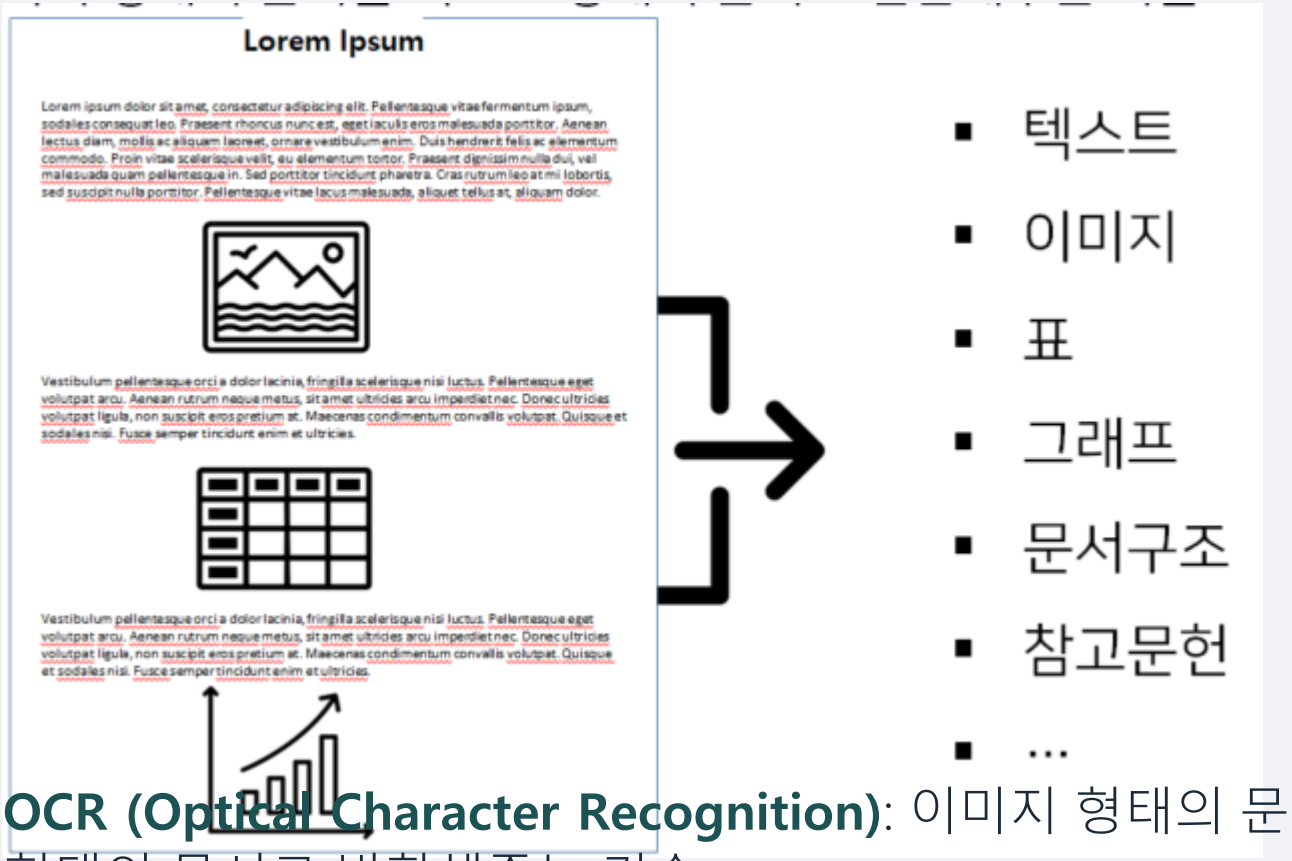
Paper

Journal

Compendium

R.R.

OCR



OCR (Optical Character Recognition): 이미지 형태의 문서를 텍스트 형태의 문서로 변환해주는 기술

OCR 프로세스

HP Manual

Paper

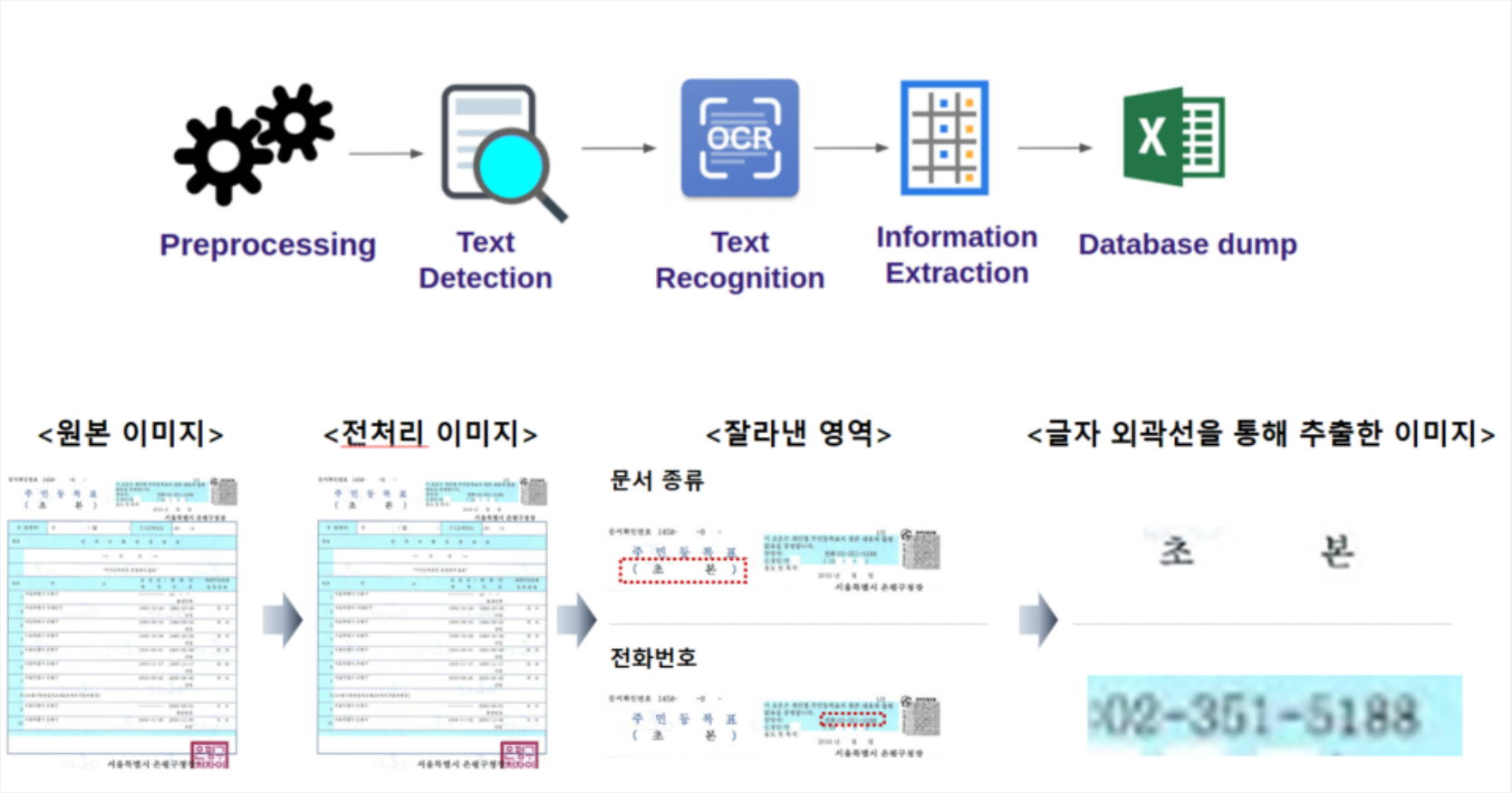
Journal

Compendium

R.R.

OCR

Process



경청해 주셔서 감사합니다.

질문?

끝