

EXPERIMENTING WITH MAKEINDEX AND UNICODE, AND DERIVING KAMEINDEX

A. Bossard and K. Kaneko
GulT Meeting 2018
Rome, Italy, October 20th.

OUTLINE

Background: indexing and makeindex

Objective

makeindex basic features

makeindex and internationalization

Advanced features: deriving kameindex

Examples

Conclusions

BACKGROUND

Indexing is about gathering the keywords of a document and listing them together with the corresponding page numbers in a special section: **the index**.

It is especially important for paper materials as they do not provide the hyperlink feature of their electronic counterparts.

LaTeX provides a semi-automatic indexing system with `makeindex`: index entries are manually declared when editing the LaTeX document.

BACKGROUND (CONT.)

One important issue with the makeindex subsystem is its limited support for Unicode.

Developed to process ASCII files, the production of multi-lingual indices with makeindex is severely limited.

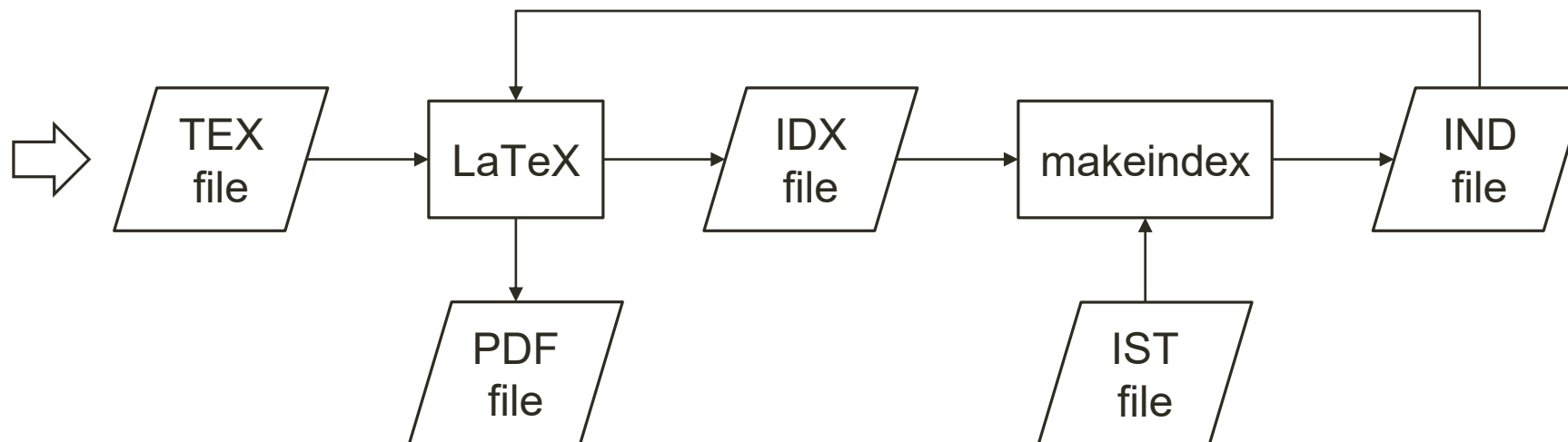
OBJECTIVE

1. To review (and demonstrate) the multi-lingual capabilities of makeindex.
2. To derive a Unicode-capable indexing solution, `kameindex`, which will seamlessly replace one component of the makeindex subsystem to expand the basic multi-lingual features of makeindex.

MAKEINDEX BASIC FEATURES

makeindex (L. Lamport, 1987) generates the index from a list of author-declared index entries.

makeindex functions according to the following flow.



MAKEINDEX BASIC FEATURES (CONT.)

makeindex enables the usage of subentries, page range, and various display customization features:

- Subentry: $\text{\index{entry!subentry}}$
- Page range: $\text{\index{range|() ... \index{range|)}}$
- Separate entry key and value: $\text{\index{key@value}}$
- and so on.

MAKEINDEX AND INTERNATIONALIZATION

makeindex does not support Unicode.

Yet, it is possible to generate multilingual indices thanks to several techniques.

For instance, because the letter ‘é’ is out of sequence in ASCII (or Unicode), correct sorting (i.e. entries starting with ‘é’ should appear in the E group) of index entries is possible by using a separate key and value: `¥index{ecole@école}` instead of `¥index{école}`.

This technique can also be applied to Japanese: `¥index{いす@椅子}` instead of `¥index{椅子}`.

Note that using Unicode when declaring entries is not problematic: their processing is.

MAKEINDEX AND INTERNATIONALIZATION

What can be achieved with makeindex:

Index

école, [2](#)

entropy, *see also* there

ersatz, *see* here

range, [1–2](#)

world, [1](#), [2](#)

 blue, [1](#)

water, [1](#)

 sphere, [1](#)

鳴き (kana: なき *naki*), [2](#)

匂い (kana: におい *nioi*), [2](#)

ADVANCED FEATURES: DERIVING KAMEINDEX

One important problem with this way of generating an index, that is, purposefully ignoring the fact that entry values are Unicode, is that `makeindex` will not be able to sort and group entries correctly.

Aside, `makeindex` output on the left, `kameindex` on the right.

Index

E

école, 2
entropy, *see also* there
ersatz, *see* here

R

range, 1–2

W

world, 1, 2
 blue, 1
 water, 1
 sphere, 1

鳴

鳴き (kana: なき *naki*), 2
匂い (kana: におい *nioi*), 2

Index

E

école, 2
entropy, *see also* there
ersatz, *see* here

R

range, 1–2

W

world, 1, 2
 blue, 1
 water, 1
 sphere, 1

な

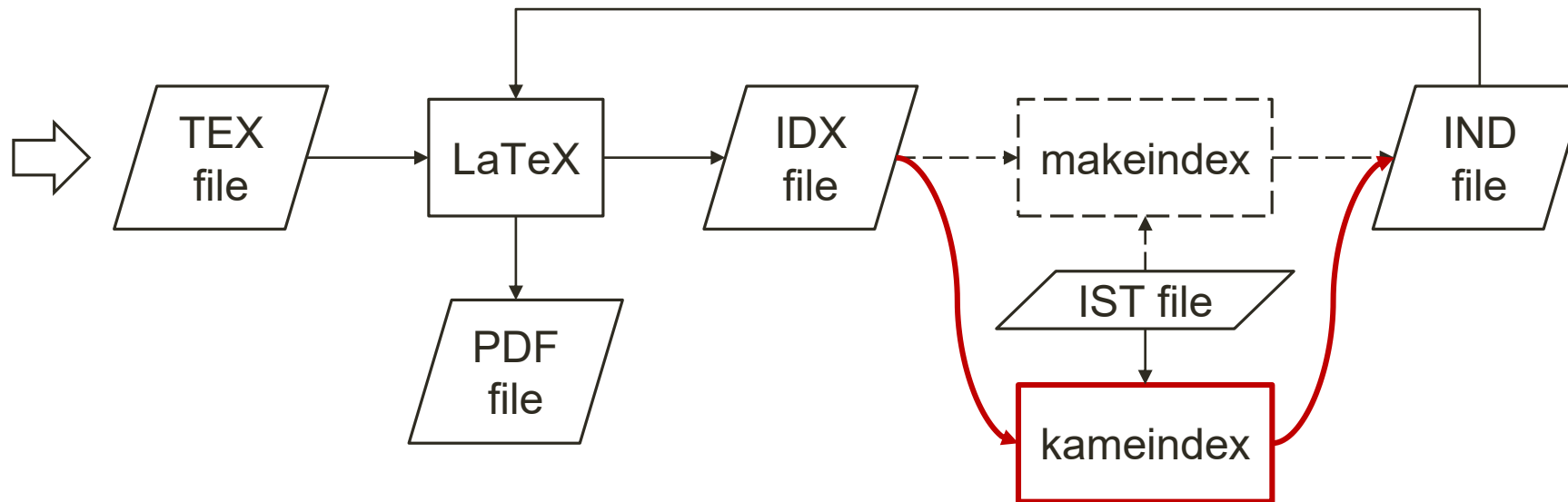
鳴き (kana: なき *naki*), 2

に

匂い (kana: におい *nioi*), 2

ADVANCED FEATURES: DERIVING KAMEINDEX

kameindex has been designed to independently process index entries that have been generated by the makeindex package. It thus replaces one single component of the makeindex subsystem, as detailed below.



ABOUT CHINESE AND KOREAN

Chinese can be supported by relying on pinyin to sort and group entries, romanization thus being used: easy!

The case of Korean is more complex since entry sorting and grouping are not based on the romanization of the index entry. (Details are abbreviated here.)

Aside, makeindex (left) and kameindex (right) results.

Index

D

电话 (*diànhuà*), 1
电力 (*diànlì*), 1

Y

英语 (*yīngyǔ*), 1

국문

국문 (*gugmun*), 1
암클 (*amkeul*), 1
언문 (*eonmun*), 1
한글 (*hangeul*), 1

Index

D

电话 (*diànhuà*), 1
电力 (*diànlì*), 1

Y

英语 (*yīngyǔ*), 1

국문

국문 (*gugmun*), 1

암클

암클 (*amkeul*), 1
언문 (*eonmun*), 1

한글

한글 (*hangeul*), 1

BRIEF COMPARISON WITH TEXINDY

1. texindy's usage is comparatively complex.
2. texindy is not compatible with hyperref. ⚠
3. texindy requires manual language selection. ⚠
4. texindy supports neither Japanese, Chinese nor Korean. ⚠

Of course, texindy has many other advantages...

CONCLUSIONS

Because not supporting Unicode, it is difficult (at best) to create multilingual indices with `makeindex`.

By replacing one component of the `makeindex` subsystem, `kameindex` enables easy multilingual index creation.

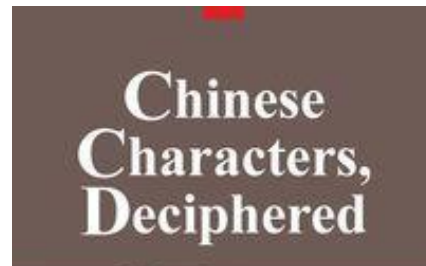
Future works include the completion of the style file support as well as the integration with related indexing systems such as `imakeidx`.

Fin

INFORMATION

kameindex (source and binaries) can be retrieved from the official homepage:

<http://www.sci.kanagawa-u.ac.jp/info/abossard/kameindex/>



Author: A. Bossard

Title: Chinese Characters, Deciphered

Publisher: Kanagawa University Press

ISBN: 978-4-906279-14-2