

INEX Book Track – Results evaluation method description

Introduction

Measuring TOC extraction quality is a fuzzy task. TOC extraction is rendering several entities and the evaluation of different submissions will be performed by comparing each of the rendered entities to the entities in the ground truth files.

The atomic unit of each TOC is a *TOC Entry*. TOC entries are defined by their *Title* and contain additional properties – *Depth Level* and *Target Page*.

Measurement concepts

A TOC entry is fully recognized if the Title and its properties have a match in the ground truth file.

A TOC entry can also be partially recognized if:

- Its Title matches a Title from the ground truth file
- Its Title and some of its properties match to a Title and its properties in the ground truth file

Measurement entities

For fully and partially recognized entries five entities will be exposed in the results:

- Titles** – A title is correctly recognized if there is a matching title in the ground truth file
- Depth levels** – A depth level is correct if there is an entry with matching title of the same depth level in the ground truth file
- Target pages (links)** – A link is correctly recognized if there is an entry with matching title linking to the same physical page in the ground truth file
- Complete entries** – An entry is entirely correct if there is an entry with matching title and same depth level linking to the same physical page in the ground truth file
- Entries disregarding depth** – Equivalent to *Links*

Matching concepts

Titles – Titles match if the distance¹ between two title strings is less than 20% of the shorter string and if the distance of first and last five characters (or less if the string is shorter) is less than 60%. This method of measurement has been introduced to reduce the impact of the OCR errors. This method has been empirically verified to produce best results against OCR errors minimizing false positives.

¹ Distance is calculated using a modified Levenshtein algorithm. The modification consists of the following: The cost of alphanumeric substitution, deletion and insertion is 10. The cost of non-alphanumeric substitution, deletion and insertion remains 1. Distance between two strings A and B in percent is calculated as D=

$$\frac{\text{LevenshteinDist} * 10}{\text{Min}(A.\text{len}, B.\text{len})} [\%]$$

Depth level and Links – These match if there is an identical Link (or depth level) in the compared file

Entity accuracy measures

For each entity three measures will be calculated for each book:

- **Precision** – $\frac{\sum \text{Correctly recognized entities } X}{\sum \text{Entities recognized as } X}$
- **Recall** – $\frac{\sum \text{Correctly recognized entities } X}{\sum \text{Entities } X \text{ in the ground truth file}}$
- **F-Measure** – $\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$

Each of the measures will be averaged on the entire book set to produce the final report.

Score sheet

A score sheet will be produced for each team. An example follows:

General Info

Participant Id	Some ID
Run Id	Some ID
Task	book-toc
Toc Creation	automatic
Toc Source	full
XML	Yes
Source files	JPG
	PDF
Description	Description here...

Results

	Precision	Recall	F-Measure
Titles	x%	x%	x%
Levels	x%	x%	x%
Links	x%	x%	x%
Complete entries	x%	x%	x%
Entries	x%	x%	x%

disregarding depth			
-----------------------	--	--	--