

The ISCA Conference Archiving System

Martin Cooke (m.cooke@ikerbasque.org)

September 11, 2016

The ISCA Archive

ISCA provides an archiving service to ISCA-supported events. By providing some simple information in advance of your workshop, you can ensure that the archive is available at the start of the meeting, avoiding the need to produce your own conference proceedings. Page numbers and stamped titles are generated during the archiving process. We have recently added the option of assigning DOIs for full papers too. Check out the latest Interspeech (2016) in the archive to see how your archive will appear.

Overview of the archiving process

You will need to provide the following

1. a text file called `conference.txt`
2. a folder called `pdfs` of article pdfs, with names `<ID>.pdf`.
3. a folder called `abstracts` of textual abstracts, with matching names `<ID>.txt`.
4. a delimited table called `papers.txt` which links IDs to paper titles and authors.
5. a delimited table called `sessions.txt` which identifies which papers are in which sessions.
6. (optional) a proceedings introduction file in pdf format

Preliminaries

There is no administrative support in ISCA for producing the archive, so I *implore* you to follow the simple set of requirements to the letter. In particular

- all text files should be in **plain text format**
- all text files should be **saved as UTF-8**. This is particularly crucial for text containing accents and the like. UTF-8 has been around for a decade, is the default on most sensible text editors.
- follow the naming and delimiter conventions precisely

The conference.txt file

The `conference.txt` is a list of attribute-value pairs which describe the archive. An example is shown in Table 1. Most of the fields should be self-explanatory. The `title` attribute should be reasonably short as it will be stamped on each pdf, along with the date. The optional `intro` attribute identifies the name of a pdf file that contains additional conference info (e.g., conference booklet) that you wish to appear in the archive. The `series` attribute is used where the conference is one of a series (e.g., for Interspeech). Even if your conference is a one-off, it is worth providing a single word identifier to name the series. If your conference is part of a series then you may already have an ISSN. An ISSN is required if you wish your papers to have DOIs. However, we can temporarily assign an ISSN if you do not yet have one. Talk to me about this.

```
series = PQS
title = PQS 2016 Workshop on Perceptual Quality of Systems
location = Berlin, Germany
startDate = 29/8/2016
endDate = 31/8/2016
URL = http://pqs.qu.tu-berlin.de/
chair = Sebastian Möller and Sebastian Egger
intro = intro.pdf
ISSN = 2312-2846
```

Table 1: *Example conference.txt file.*

The pdfs folder

There should be one pdf per full paper. Papers should not contain page numbers, logos nor conference details. These will be added automatically by the system. The identifiers used to name papers are arbitrary, so might be `1.pdf`, `5.pdf` etc, or could be named after the first author, or something else (see Table 2).

The abstracts folder

Abstracts should not contain authors or titles or anything other than the abstract itself. It is acceptable to have abstracts that do not have papers associated with them (typically for keynotes), but all papers with a pdf must also have an abstract in the `abstracts` folder. The IDs used in the naming of abstracts must correspond to those used in naming the pdfs.

The papers.txt file

This is a 3 column text file with the ‘|’ symbol used to delimit columns (Table 2) and column headers `id`, `authors` and `title`.

id	authors	title
Haizou	Haizhou Li	Voice conversion and spoofing countermeasures
Narayanan	Shrikanth S. Narayanan	Understanding ...
Najim	Najim Dehak	I-Vector Representation Based on GMM and DNN
3	Pierre-Michel Bousquet, Jean-Francois Bonastre	Constrained ...
6	Yosef Solewicz, Hagai Aronowitz, Timo Becker	Reducing Noise ...

Table 2: *Example truncated papers.txt file.*

If you produce this file by exporting from a spreadsheet, please ensure that the correct delimiter has been used and that no unnecessary quotes are included. IDs must correspond to those used in the abstracts and pdfs folders.

The authors column is quite flexible (e.g., use of commas, ‘and’ or ‘&’). All of the forms shown in Table 3 are acceptable. It is assumed that everything except for the final token are given names. If multiple surnames exist (as in Spain), these should be grouped like this: Enrique {Lopez Garcia}.

John Smith
J Smith
John K. Smith
John Karl Smith
J Smith and Joe Bloggs
J Smith & Joe Bloggs
J Smith, J Bloggs and R Brown

Table 3: *Acceptable author formats*

The sessions.txt file

This table relates papers to sessions and as such enables the archive to reflect the logical grouping and sequence of papers. The title column can contain any string, while the papers column is a comma-separated list of paper IDs in the order in which they occur in the session (or, at least, in the order in which you want them to appear in the archive).

title	papers
Keynote:	Haizhou Li Haizou
Text-Dependent Speaker Verification	68,67,24,63
Speaker Recognition: i-vector approaches	30,22,17,3,60

Table 4: *Example sessions.txt file*

That’s it!