

Demonstration to Use Regular Expression, Python and LaTeX Template to Automatically Create Documents

DR. HOWARD CHEUNG

Learning objectives

To understand the procedure to automatically create many documents from data and template automatically

To know what tools are available for the task

- Regular expression
- Typesetting software

Background

You may need to create a lot of documents based on the same template and a lot of data

- Scorecards of student performance
- Utility bills

The template and the blanks to fill in are the same. Only the data are different.

It's really inefficient to create each document on you own, right?

- Can we make it faster?

Demonstration

Three author indexes for three categories of conference papers

- Include all authors of 200 papers
- Alphabetically ordered by the last name of the authors
- Each paper number should contain a link to the pdf file of the paper

Creating the indexes manually?

- Time-consuming
- Error-prone in the ordering of authors and links

So what do we want?

Paper information in all categories in one text file

Author indexes of 3 categories of papers with paper numbers and links

R-01: Evaluating Alternative Refrigerants & Technologies
Time: Monday, 14/Jul/2014: 1:00pm - 3:00pm • Location: 214 A&B
ID: 2175
R-404A Alternative Refrigerant with Low Compressor Discharge Temperature
Barbara Minor¹, Vladimir Sulc², Jeff Berge³, Michal Koloda⁴, Michal Hegar⁵
¹DuPont Fluoroproducts; ²Thermo King, Corp, Ingersoll Rand; ³Thermo King, Corp,
Equipment Manufacturing; barbara.h.minor@dupont.com

ID: 2250
AHRI Low Global Warming Potential Alternative Refrigerants Evaluation Program (I
Xudong Wang, Karim Amrane
Air-Conditioning, Heating, and Refrigeration Institute, United States of America

ID: 2256
Evaluation and Soft-Optimization for R410A Low-GWP Replacement Candidates through
Yunho Hwang¹, Abdullah Alabdulkarem^{1,2}, Radia Eldeeb¹, Vikrant Autel, Reinhard I
¹University of Maryland, College Park, United States of America; ²King Saud Uni

ID: 2278
An Evaluation of the Environmental Impact of Different Commercial Supermarket Re
Mohamed Beshr¹, Vikrant Autel, Omar Abdelaziz², Brian Fricke², Reinhard Raderma
¹University of Maryland, United States of America; ²Oak Ridge National Laborato

ID: 2328
Assessment Of Life Cycle Climate Performance (LCCP) Tools For HVAC&R Application
Stephen Anthony Kujak, Panayu Robert Srichai, Kenneth J. Schultz
Ingersoll Rand, United States of America; skujak@trane.com

ID: 2505
Novel Reduced GWP Refrigerant Compositions for Stationary Air Conditioning

Author Index

Aaraj, Youssef	2169	Becerra, Elizabet Vera	2670
Abdelaziz, Omar	2277, 2278	Beck, Stephen B. M.	2126

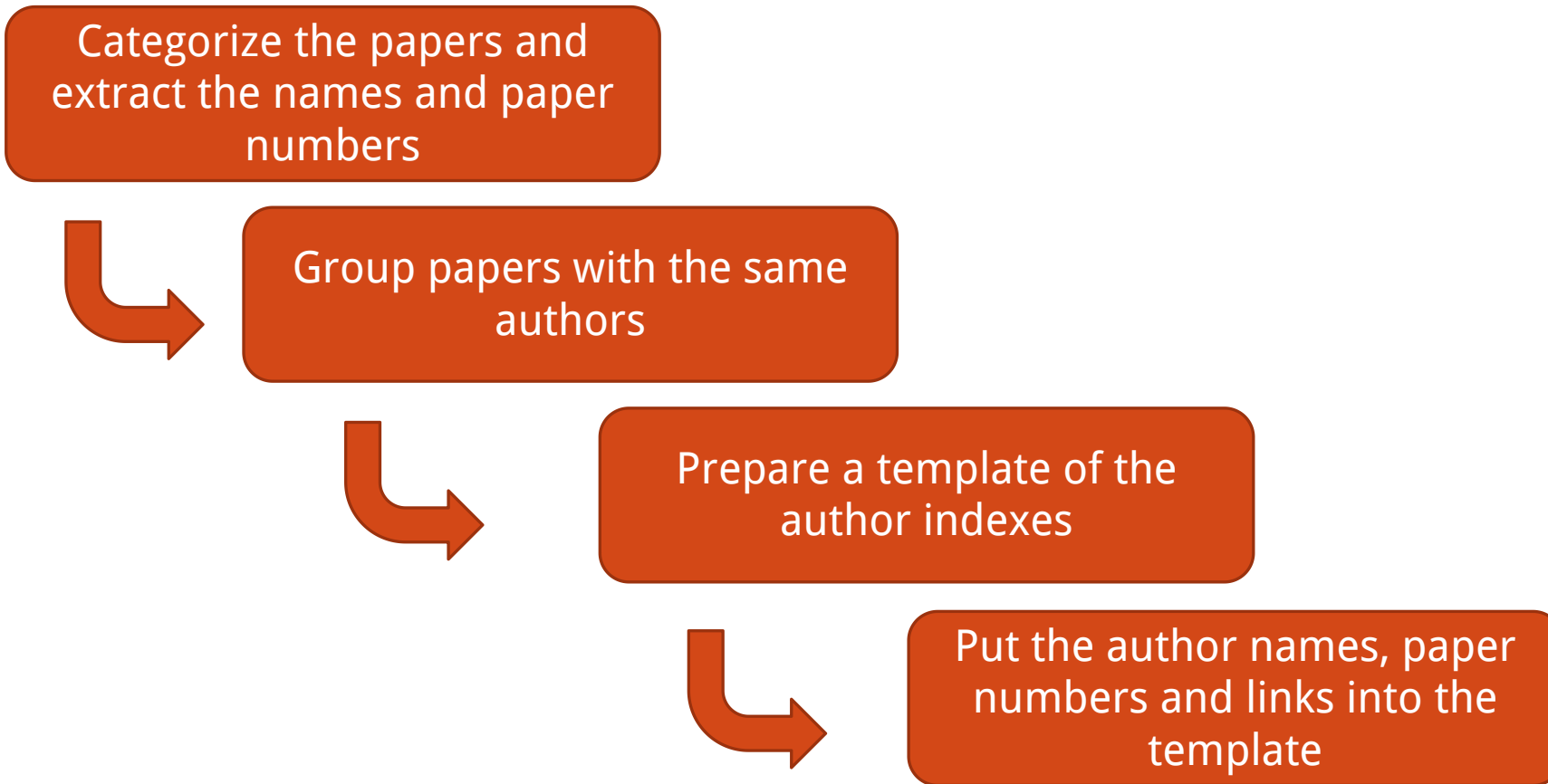
Author Index

Afshari, Afshin	1197	Chakraborty, Anutosh	1129, 1134
Akasaka, Ryo	1260	Chen, Michael	1200

Author Index

Alabdulkarem, A	Aleksandr, Drozdov	1197	Chakraborty, Anutosh	1129, 1134
Alleyne, Andrew	Almbauer, Raimund	1260	Chen, Michael	1200
Allgood, Charles	Anami, Keiko			
Almbauer, Raimund	Andreotti, Thiago			
Alsaleem, Fadi M	Angadi, Shruti			
Altan, Hasim	Arantes, Danilo Ma			
	Aboulmagd, Ahmed	3559	Duan, Yifei	3107
	Arceno, Eduardo		Fan, Cheng	3135
	Adams, Aaron L.	3377	Fong, Nai-Kong	3139, 3522
	Arjenah, Mohammad	3616	Foreman, J. Christopher	3377, 3572
	Aw, Kuan Thai	3321, 3503	Fugate, David	3494, 3623
	Albee, Kristen	3314	Gasparella, Andrea	3303, 3480,
	Balide, Venkateshar	3437		3525,
	Beinert, Michael	3547		3523, 3525,
	Bell, Ian	3486		3547, 3552
	Benanti, Travis L.	3303, 3525		
		3365	Georges, Emeline	3314
		3616, 3617	Georgescu, Michael Vincent	3619
		3259	Giguere, Daniel	3656

Framework



Software tools

LaTeX

- A typesetting software
- To define the format and font in the author indexes
- To add the links to the paper numbers in the author indexes
- To create the final author index files

Python

- Data processing
- To use regular expression to extract data in regular patterns
- To create the input files for LaTeX

Regular expression

A sequence of characters that are used to identify patterns

Normal pattern matching can only work with exact characters and numbers

- You can match '1a' with '1a' but not '1b'
- You need to match '1a' and '1b' for both terms

Regular expressions allow you to match strings by types of characters

- '1[a-z]' to match all terms from '1a' to '1z'
- '1[a-z]+' to mach terms like '1azeewfewg' or '1aydspvwov'

Typesetting

A system to define font and format of documents

- LibreOffice Writer, MS Word, OpenOffice Writer, Markdow and LaTeX

LaTeX aims at separating the preparation of visual appearance and content

Preamble for major
formatting issues

Document content

LaTeX file

```
2 % document class definition
3 \documentclass[letter, 10pt, oneside]{article}
4 \setlength{\parskip}{0in}
5 \setlength{\parindent}{0in}
6 \textwidth = 470pt
7 \linespread{1.1}
8 \textheight = 630pt
9 \oddsidemargin = 15pt
10 \marginparsep = 4pt
11 \marginparwidth = 0pt
12 \hoffset = -0.5in
13 \footskip = 10pt
14 \pagestyle{empty}
15 \setlength{\columnsep}{1in}
16 \columnwidth = 3in
17
18 % include packages
19 \usepackage[pdftex, hidelinks]{hyperref} % for hyperlink
20 \usepackage[utf8]{inputenc}
21 \usepackage{supertabular} % for long table
22 \usepackage{array} % for left-aligned cells
23 \newcolumntype{P}[1]{>{\raggedright\arraybackslash}p{#1}}
24
25 \begin{document}
26
27 \twocolumn[
28 \begin{@twocolumnfalse}
29 \centering{\Large{\bf{Author Index}}} \}
```

Compile



Author Index			
Aleksandr, Dronov	1197	Chakraborty, Ananta	1129, 1134
Alubauer, Raimund	1260	Chen, Michael	1290
Anami, Keiko	1562, 1565	Chen, Jindang	1485
Androsi, Thiago	1217	Chen, Xia	1130, 1297
Angadi, Shruti	1146	Chen, Zhenhua	1289
Arantes, Danilo Martins	1217	Cheng, Junning	1150, 1163
Arreno, Eduardo	1224	Chien, Toshiro	1434
Arpuch, Mohammad	1253	Choi, Yang-Hee	1346
Aw, Kuan Thai	1216	Choi, Gyung Min	1464
Balke, Venkatesham	1146	Choo, Wei Chong	1290
Beisert, Michael	1667	Chu, Chengguo	1479
Bell, Ian	1506, 1604	Chukanova, Ekatarina	1476
Benanti, Travis L.	1491, 1507	Chung, Jin-Ah	1350
Berger, Erwin	1260	Cipollone, Roberto	1651, 1652
Bernagolo, Stephanie	1627	Climaco, Douglas	1367
Bhaktia, Aditya	1144, 1151,	Cohen, Sophie	1168
	1152, 1153	Conde, Ricardo	1644
Bisachi, Giuseppe	1651	Costabile, Giulio	1651
Biselman, Tim R.	1470	Cook, Graeme	1358
Bilal, Nasir	1701, 1702, 1703	Couto, Paulo Rogério Carrara	1248
Bjerre, Preben	1315	Devlage, Sébastien	1604
Bolajo, Siva Ramu Krishna	1543	Degroot, Joris	1478
Bortoli, Marcos Giovanni Drupa de	1248, 1416	Deshamps, Cesar J.	1302, 1367,
Bowers, Chad D.	1507		1372, 1374, 1376
Bowman, Anthony J.	1549	Deshmukh, Satish Kondrerao	1199
Brashers, Craig R.	1177, 1178,	Desideri, Adriano	1451
	1378, 1446	Dewitte, Philippe	1595
Branch, Scott	1241	Dey, Subhrajit	1146
Brancher, Ricardo D.	1302	Dhan, Sandeep	1144
Brasz, Joost J.	1329, 1435,	Dickes, Rémi	1604
	1470, 1498, 1609	Ding, Hui	1548
Brann, James E.	1272, 1649	Diniz, Marco C.	1372, 1374
Brock, Martijn van den	1451, 1506	Doi, Masahito	1418
Brumner, Andreas	1449	Dose, Vamsidhar	1146
Brune, Fabian	1315	Dong, Ying	1348
Buckley, David	1390, 1358	Drasman, Nikil	1226
Bukac, Hubert	1589	Du, Bin	1482
Caballero, Piero	1435	Du, Chunxu	1297
Cambio, Matt	1409	Dumont, Olivier	1604
Chai, Guocai	1408	Dutra, Thiago	1367

Actual demo using Jupyter notebook

Summary

Demonstration

- Creating documents using the same template and different data

Introducing a procedure to create the documents

- Using Python and regular expression
 - To extract and organize the text data
- Using a typesetting software LaTeX
 - For the template of the document
 - For compilation of the actual document

Post-event evaluation

Please let us know what you think

<https://tinyurl.com/y7qu22zu>

Evaluation of Second PyData Hong Kong Meetup on 2018/06/25

* Required

Is the starting time of the event appropriate? *

☐ Yes

☐ Too late

☐ Too early

