

Open and collaborative classification of RTSS papers

A problem about research, not research about a problem

Enrico Bini

University of Turin

Real-Time Systems Open Problems Seminar

(special edition @RTSS19 PC meeting)

July 24th, 2019

Outline

1 Introduction

2 Classification of RTSS papers

3 Future

Motivation

- Researchers are aging. So the scientific memory does
 - ▶ Thomas Nolte: “This problem is addressed by a PhD thesis every 7 years”
- Forgetting research is going to worsen (I think)
 - ▶ Growing number of papers per year
 - ▶ “Publish or perish” pressure reduces the care in paper writing, including review/study of existing research
 - ▶ Tendency to delegate to AI-something (most notably Google scholar) the job of finding related works
- (personal motivation) I decided to illustrate at my new department (in Turin) what real-time research is about
- Result: a classification of all RTSS papers from 1990 to 2015

Outline

1 Introduction

2 Classification of RTSS papers

3 Future

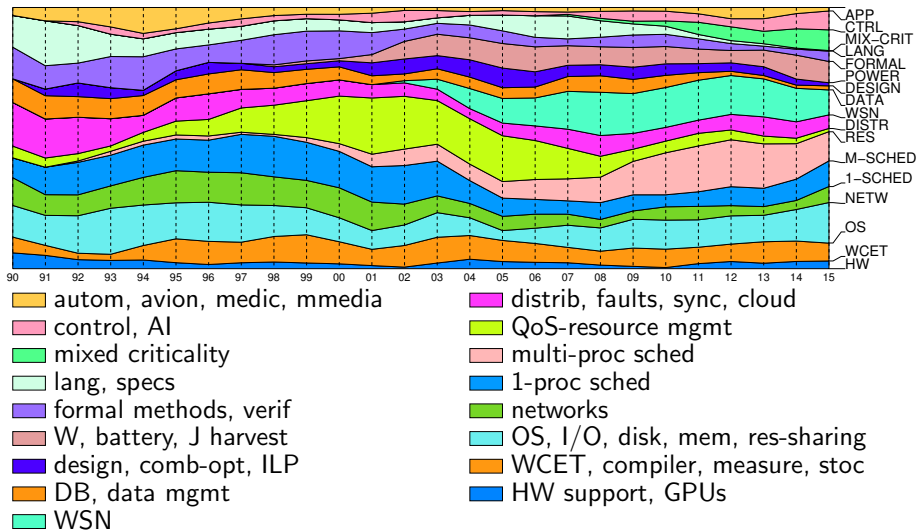
Classification of RTSS papers: methodology

- Exhaustive classification of all RTSS papers available on IEEExplore (from 1990 to 2015): 994 papers.
- Took three full days of work
 - 1 a number n ($= 17$) of topics was identified (personal choice)
 - 2 each (of the 994) paper j was assigned a weight $w_{ij} \geq 0$ representing how much such paper j is related to topic i

$$\forall \text{ paper } j, \quad \sum_i w_{ij} = 1$$

- 3 Weights w_{ij} were computed by
 - ★ reading the title
 - ★ reading the abstract
 - ★ scanning the paper
- Example: R.J. Bril, S. Altmeyer, M.M.H.P. van den Heuvel, R.I. Davis, and M. Behnam, "Integrating CRPD into FP ...", RTSS 2014
 - ▶ $w_{ij} = 0.8$, $i = 1$ -SCHED (single proc. sched. analysis)
 - ▶ $w_{ij} = 0.2$, $i =$ WCET (Compiler-WCET/multithreaded-proc sched/cache/stochastic/monitoring/measurements)

Classification of RTSS papers: results



The status

- Data is available on github:

`https://github.com/ebni/classify-rt`

which you can get by

```
git clone https://github.com/ebni/classify-rt.git
```

- The git repository contains:
 - ▶ the raw data w_{ij} paper data (.csv format) of 2014, 2015
 - ▶ the aggregate $\sum_{j \in \text{paperPerYear}(y)} w_{ij}$ per topic i data of 1990–2013 (source w_{ij} are probably lost)
 - ▶ this presentation (which you can compile via Makefile)
- Adding a new year requires:
 - ▶ to assign all w_{ij} (about 2 hours depending on level of details)
 - ▶ to copy/paste data over an .xlsx file
 - ▶ some LibreOffice/Excel work

Published at IEEE TCRTS

- After chatting with Jim Anderson about this effort, he proposed me to publish the figure on the IEEE TCRTS



The screenshot shows a web browser window with the URL `sites.ieee.org/tcrts/classification-of-rtss-papers/`. The page header features the IEEE TCRTS logo and the text "Technical Committee on Real-Time Systems". A navigation menu includes "Home", "News", "Committees", "Conferences", "Awards", "Education", "Research", "Software", "Member recognition", and "Contact". The "Research" menu is open, showing "Research groups", "Projects", and "Classification of RTSS papers". The main content area has the heading "Classification of RTSS papers" and a disclaimer: "Disclaimer: The classification presented on this page does not represent the views of the IEEE TCRTS executive committee, but rather those of the researcher, Prof. Enrico Bini, who has performed the classification." To the right, there is a calendar for July 2019.

| July 2019 | | | | | | |
|-----------|----|----|----|----|----|----|
| M | T | W | T | F | S | S |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |

<http://sites.ieee.org/tcrts/classification-of-rtss-papers/>

Outline

1 Introduction

2 Classification of RTSS papers

3 Future

Possible contributions (welcome)

- 1 Extending to years not covered
- 2 Revise the weights assigned by me
- 3 Extracting keywords/topics in an automated way (example: Easychair can do something)
- 4 Make data processing more automated (by scripts, python)
- 5 Adding other conferences
- 6 Refine the classification into more topics (17 was picked just to ease the plot)
- 7 Write a literature review of all RTSS papers: once papers are classified, adding a line to a survey would not cost much
- 8 Adding a DoI to each document
- 9 Having well in mind that: Writing a to-do item is much simpler than doing it. . .

Possible contributions (welcome)

- 1 Extending to years not covered
- 2 Revise the weights assigned by me
- 3 Extracting keywords/topics in an automated way (example: Easychair can do something)
- 4 Make data processing more automated (by scripts, python)
- 5 Adding other conferences
- 6 Refine the classification into more topics (17 was picked just to ease the plot)
- 7 Write a literature review of all RTSS papers: once papers are classified, adding a line to a survey would not cost much
- 8 Adding a DoI to each document
- 9 Having well in mind that: Writing a to-do item is much simpler than doing it. . .

Volunteers are welcome!