Generieren eines 2D Datensatzes mit gleichen statistischen Eigenschaften

Algorithm Engineering 2022 Project Paper

Florian Zeidler
Friedrich Schiller University Jena
Germany
florian.zeidler@uni-jena.de

Niklas Rausch
Friedrich Schiller University Jena
Germany
niklas.rausch@uni-jena.de

ABSTRACT

The five-finger pattern [1]:

1. **Topic and background:** What topic does the paper deal with? What is the point of departure for your research? Why are you studying this now?
2. **Focus:** What is your research question? What are you studying precisely?
3. **Method:** What did you do?
4. **Key findings:** What did you discover?
5. **Conclusions or implications:** What do these findings mean? What broader issues do they speak to?

KEYWORDS

entity resolution, data cleansing, programming contest

1 INTRODUCTION

1.1 Background

1.2 Related Work

1.3 Our Contributions

1.4 Outline

2 THE ALGORITHM

2.1 XXX

In Figure 1 we convert the mock labels to sorted integer sets.

2.2 XXX XXX

The following findings are important to speed up preprocessing of the input data:

- Reading many small files concurrently, with multiple threads (compared to a single thread), takes advantage of the internal parallelism of SSDs and thus leads to higher throughput [2].
- C-string manipulation functions are often significantly faster than their C++ pendants. For example, locating substrings with `strstr` is around five times faster than using the C++ `std::string` function `find`.
- Hardcoding regular expressions with *while, for, switch* or *if-else* statements results in faster execution times than using standard RegEx libraries, where regular expressions are compiled at runtime into state machines.
- weiterer stichpunkt

`./graphics/integer_sets.pdf`

Figure 1: Beschreibung

3 XXX XXX XXX

Table 1 shows the running times of the resolution step of the five best placed teams.

Table 1: Beschreibung

Team	Language	F-measure	Running time (s)
PictureMe (**this paper**)	C++	0.99	**0.61**
DBGroup@UniMoRe	Python	0.99	10.65
DBGroup@SUSTech	C++	0.99	22.13
eats_shoots_and_leaves	Python	0.99	28.66
DBTHU	Python	0.99	63.21

4 CONCLUSIONS

REFERENCES

[1] Felicitas Macgilchrist. 2014. *Academic writing*. UTB.

[2] Zhenyun Zhuang, Sergiy Zhuk, Haricharan Ramachandra, and Badri Sridharan. 2016. Designing SSD-Friendly Applications for Better Application Performance and Higher IO Efficiency. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. IEEE. https://doi.org/10.1109/compsac.2016.94