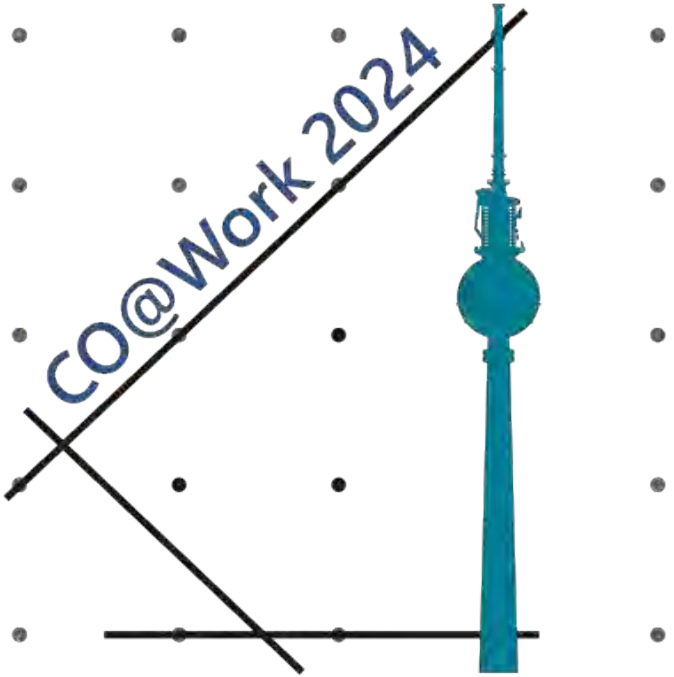
CO@Work2024: Data Experiment

Thorsten Koch



berlin
MODAL
FORSCHUNGSCAMPUS
htw.
ZIB
ZUSE INSTITUTE BERLIN

Documentation is important. Both for Programs and Data

```
v,i,j,k,l,s,a[99];
main()
{
for(scanf("%d",&s),*a-s;v=a[j*=v]-a[i],k=i<s,
j+=(v=j<s&&(!k&&!!printf(2+"\n\n%c"-(!l<<!j),
" #Q"[l^v?(l^j)&1:2])&&++l||a[i]<s&&v&&v-i+j&&
v+i-j))&&!(l%=s),
v||(i==j?a[i+=k]=0:++a[i])>=s*k&&++a[--i]);
}
```

Timo was right, this is not needed

What might it possibly do?

Is solves the n-queens problem for size 4 to 99. (Winner IOCCC 1990 Best small program)

Sooner or later, everyone does big data

- Optimization is a major topic of CO@Work.
- There is no optimization without data.
  To be able to optimize your experience we need some information from you.
- For example, to optimize the composition of the learning groups.
- Therefore, we ask you to provide some data as defined on the following pages.
- The process should be automatic.
  Therefore, it is very important, you follow the instructions carefully.
- Mandatory to submit by Wednesday, Sep 18, 8 pm for all participants
  (the data is required for the computational challenge)

Data needed

We ask everybody of you to provide the following information:

1. A JSON file with some information about you.
2. A picture of yourself (or an avatar if you don’t want your picture online).
3. A picture of the place where you stay.

We will assemble the pictures to a virtual group photo and a slide show and use the provided information to optimize the composition of the learning groups.

Data Formats

The file with the information about you should be
in JSON Format (ISO/IEC 21778:2017).
https://en.wikipedia.org/wiki/JSON

The file with your picture should be
in JPEG format (ISO 10918-1) and have a size of 512 × 512 pixels.
https://en.wikipedia.org/wiki/JPEG

The file with your place should be
in JPEG format (ISO 10918-1) and have a size of 1920 × 1080 pixels.

The JSON file should contain the following fields

Field Name	Type	Description
Name	String	Your full name in your native language
Email	String	email address you used for registration at CO@Work
Country	String	country of origin as an ISO 3166-1 Alpha-2
Languages	Array of Strings	List of all languages you speak as ISO 639-1 codes. Use capital case if you are fluent in the language and lower case if you only have limited knowledge.
Motto	String	motto/aphorism characterizing you to write under your picture
Clearance	String	I herewith grant the organizers the right to use and share the attached pictures for purposes related to CO@Work
Skill	Integer [0-100]	How would you rate your Skill in Computational Optimization
Level	Integer [1-5]	What is your current level of education: 1 = Undergraduate, 2 = Master's student, 3 = PhD student, 4 = Postdoc or professional, 5 = Prof.
Tools	Array of Strings	Which optimization software tools have you used: None, Xpress, Gurobi, SCIP, Copt, CPlex, HiGHS, GAMS, AMPL, …
CourseProject	Boolean	Have you worked on a real-world optimization problem (e.g., energy, logistics, finance) in a course project?
ResearchProject	Boolean	Have you worked on a real-world optimization problem (e.g., energy, logistics, finance) in an academic project?
IndustryProject	Boolean	Have you worked on a real-world optimization problem (e.g., energy, logistics, finance) in an industry project?
Experience	Integer [0-3]	0 = I have not implemented any, 1 = I used prebuilt implementations, 2 = I implemented optimization algorithms from scratch, 3 = I have developed advanced/custom optimization algorithms

JSON File Example

{ "Name": "Thorsten Koch“, "Email": "koch@zib.de", "Country": "DE",
"Languages": ["DE", "EN", "la"],
"Motto": "The code was hard to write, it should be hard to read",
"Clearance": "I herewith grant the organizers the right to use and share the attached pictures for purposes related to CO@Work",
"Skill": 92, "Level": 5,
"Tools": ["Cplex", "Xpress", "Gurobi", "Copt", "SCIP"],
"CourseProject": false, "ResearchProject": true,
"IndustryProject": true, "Experience": 3
}

Please submit the files as follows:

- *LastnameFirstname* should be the English transcription of your name
- The name of the JSON file should be *LastnameFirstname*.**json**
- The name of the file with your picture should be *LastnameFirstname*.**jpg**
- The name of the file with the picture of your place should be *LastnameFirstname*-**place**.**jpg**
- All 3 files should be **attached** to an email
- Send the email to coaw-data@zib.de
- The subject of the email should be CO@Work: Data for *LastnameFirstname*
- *Please, as soon as possible (e.g. today!)*

Wrong Subject

113 Emails + 12 Updates/corrections

With error: ~78, let's say 2/3

CO@Work: Data for PeterPeng
CO@Work: Data For PeterPeng
CO@Work: Data for Peter Peng
Data for PeterPeng
CO@Work: Data for LastnameFirstname

Deadline 18.09 at 20:00

18.09.24, 20:00
18.09.24, 20:17
18.09.24, 20:31
18.09.24, 21:00
18.09.24, 21:03
18.09.24, 21:05
18.09.24, 21:35
18.09.24, 22:36
18.09.24, 23:39
00:32
10:42

Picture sizes

200 x 200
428 x 512
500 x 500
512 x 513
519 x 519
524 x 289
623 x 623
1280 x 720
1310 x 494
1326 x 820
1620 x 1080
1697 x 1934
1920 x 1920

Some Errors

- *LastnameFirstname* not **lastnamefirstname** or **Lastname Firstname**
- The extension of the image files should have been .jpg, not **.jpeg**, or **.JPG**.
- *LastnameFirstname*-**place.jpg** not **_place** or **-Place**
- In 13 cases the email address was not the one used for the registration, including: [your_email@example.com]
- 2 times the Motto was empty
- The Clearance did not have a “.” or "\x93,\r\n" at the end, no line break in the middle, and **CO&Work** is also wrong, as is **hereby, here with**.
- Many cases were (probably) the editor changed " to “ or ” or inserted a line break making the JSON invalid.
- Languages wrong type [DE, EN, fr], CourseProject wrong type, ResearchProject wrong type, IndustryProject wrong type.
- Field Motto missing “Moto”, field Country wrong type, field Languages missing “Language”, field IndustryProject missing “Industryroject”.
- …

Non existing ISO-639 language codes:

PO 1

SE 1

CZ 1

GR 1

CN 1

VN 2

cn 2

AL 1

KG 1

ua 1

JSON parsing errors

Expecting property name enclosed in double quotes: line 1 column 2 (char 1)
Expecting property name enclosed in double quotes: line 1 column 3 (char 2)
Expecting property name enclosed in double quotes: line 1 column 2 (char 1)
Expecting property name enclosed in double quotes: line 2 column 5 (char 6)
Invalid control character at: line 7 column 129 (char 299)
Invalid control character at: line 7 column 99 (char 272)
Invalid control character at: line 4 column 99 (char 250)
Invalid control character at: line 4 column 99 (char 223)
Invalid control character at: line 6 column 99 (char 261)
Invalid control character at: line 4 column 99 (char 263)
Expecting value: line 3 column 10 (char 144)
Expecting value: line 10 column 5 (char 184)
Expecting value: line 6 column 10 (char 122)
Expecting value: line 8 column 5 (char 141)
Expecting value: line 3 column 10 (char 119)
Expecting ',' delimiter: line 1 column 32 (char 31)
Expecting ',' delimiter: line 7 column 5 (char 168)
Expecting ',' delimiter: line 1 column 28 (char 27)
Expecting ',' delimiter: line 1 column 32 (char 31)
Expecting ',' delimiter: line 1 column 32 (char 31)
Expecting ',' delimiter: line 1 column 26 (char 25)

Random mottos

How difficult can it be?

Man, I feel like these courses are giving me OCD (optimal combinatorial decisions)

The spice must flow.

We need more coffee

I don’t need luck, I have version control.

I have a bad feeling about this.

Why follow the rules when you can just solve the problem?

Life is a party and I'm the piñata

At least you could have provided a link to a JSON validator

There are 3 tough problems in OR: formulating the right constraints and off-by-one errors

MIP, MIP, hoooray!

Still waiting for a SnackOverflow

404 - Motto Not Found

Engineer at heart, mathematician by training.

The fish buys the umbrella, but the moon stays dry

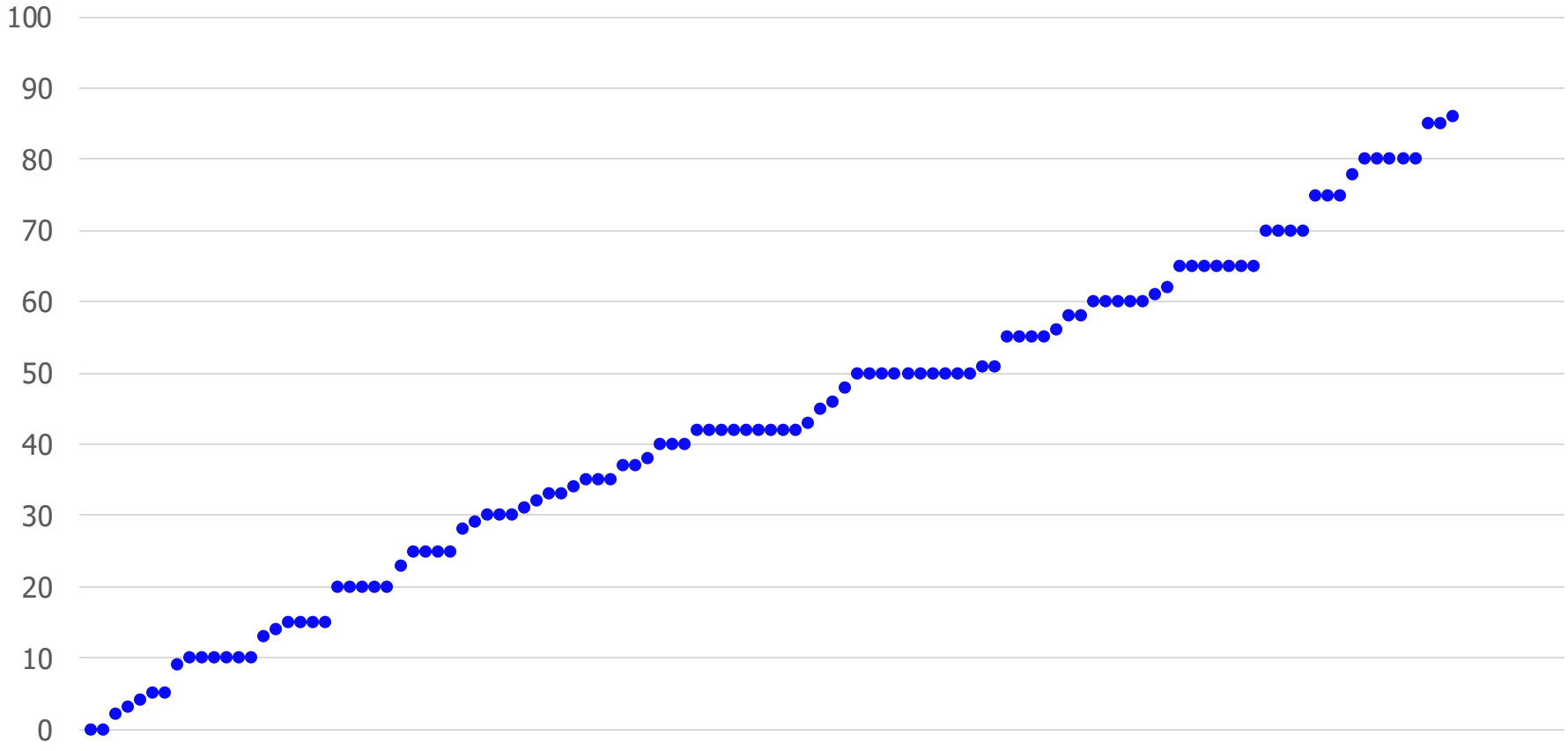
MY TURTLE IS SMARTER THAN ME

Submissions where I did not look enough to find an error

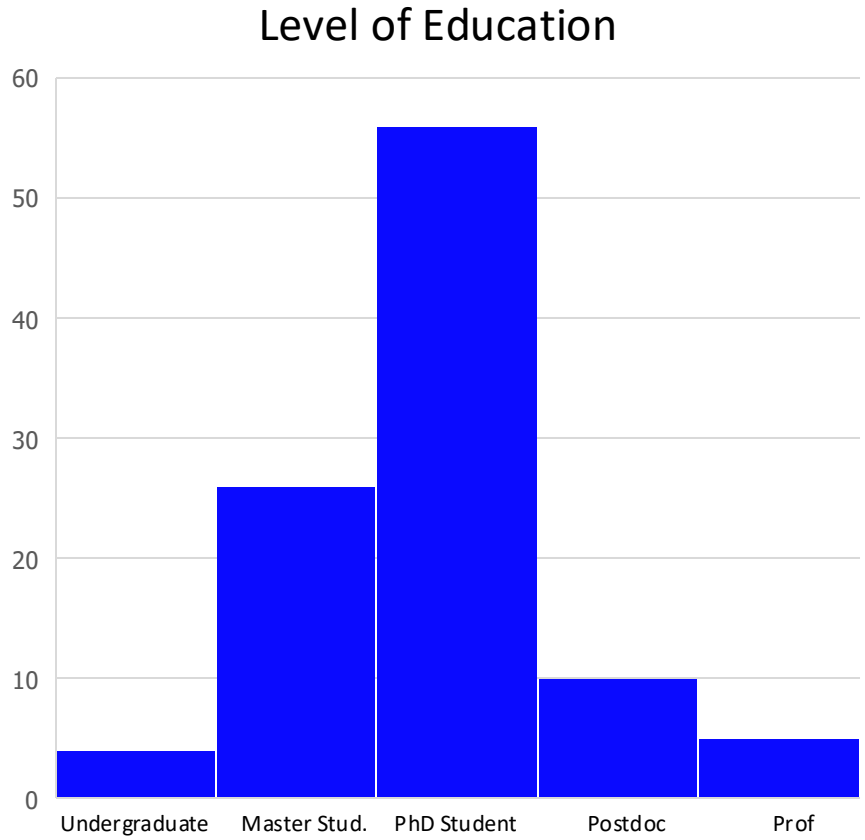
Not shown due to protection of personal data.

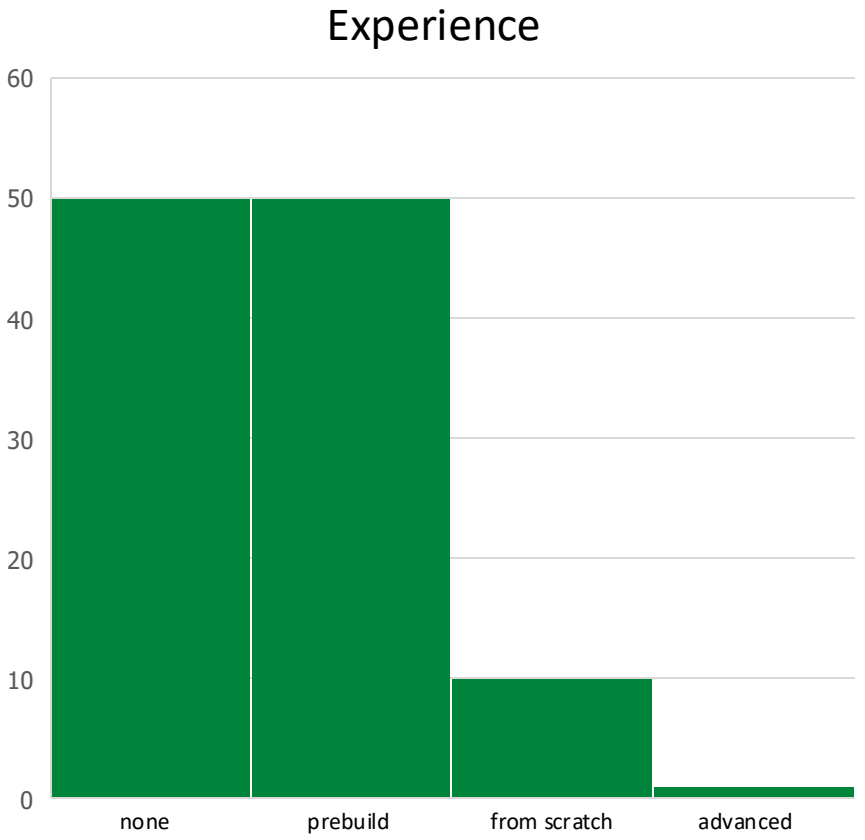
Congratualations anyway!

Skill Level



Level and Experience





Data

You would think a …

- … cellular network operator knows where its base stations are located?
- … fixed network operator can tell where the parts of its network are connected?
- … chemical company knows how many plants they have?
- … 5 m long pipeline cannot have a height difference from end-to-end of 100 m?

- Many companies have their data in Excel.
  There is no formal validation or referential integrality check.
- If they did formal validation, usually they found there was information they needed which they could not input and they started to “reuse” some data fields.
- If there is not at least 1 error per 100 data sets you are not looking hard enough.
- Usually, the data changes all the time.
- They might not want to give it to you.
- The data might just not exist.

The first result of an optimization project is usually to improve the quality of planning data available at the company.

The improvement potential is always 15%

- 5% ⇨ “So much we save by simply pushing the employees.”
- 10% ⇨ “Sounds poor. We could do similar ourselves if we would get as much money as you ask for.”
- 20% ⇨ “this sounds very ambitious. You must remember: if we give you the money, we have to promise 20% to our boss. We dare not to do this.“
- 30% ⇨ “Braggart! Get out!”

From this it follows that you have to say 15%.

- I said 15% and immediately got a signature
- I said 13 % ⇨ “Why such a crooked number? How could you be so precise?”
- I said 14%, same result.

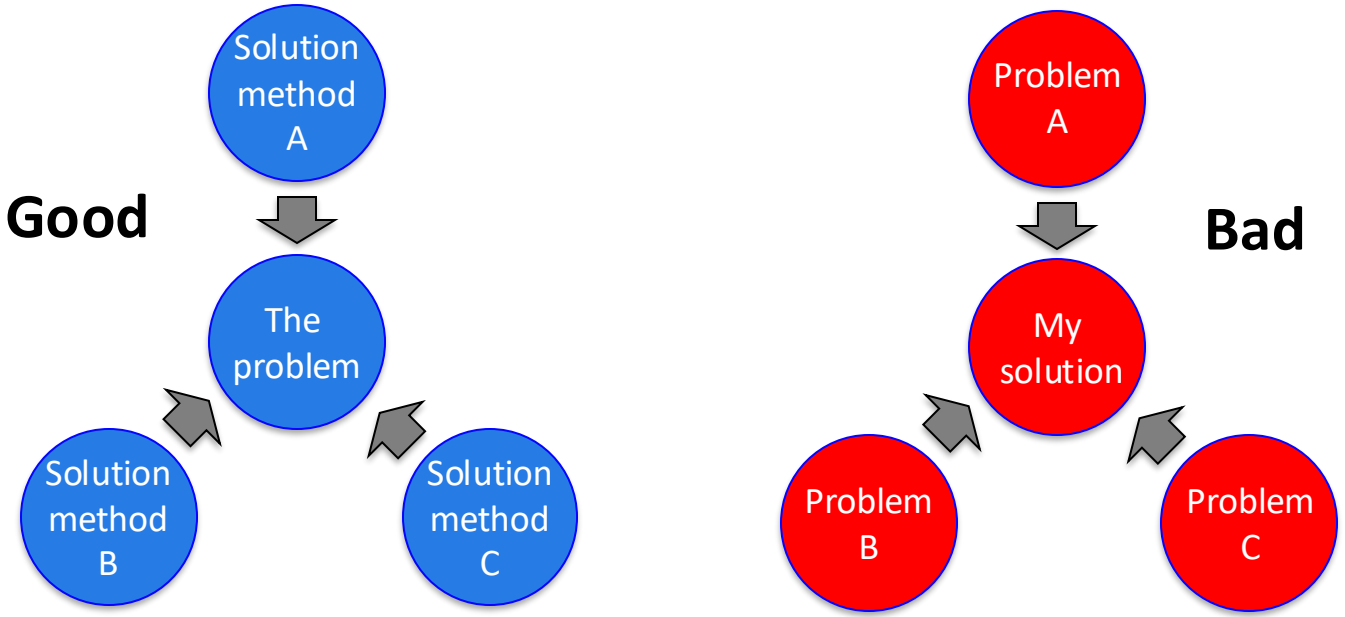
I stayed at 15 percent. Always 15 percent. Only 15 percent. All nodded, everybody satisfied. I had discovered an absolute Natural constant!

Mathematics always saves 15%. Completely regardless of the Problem!

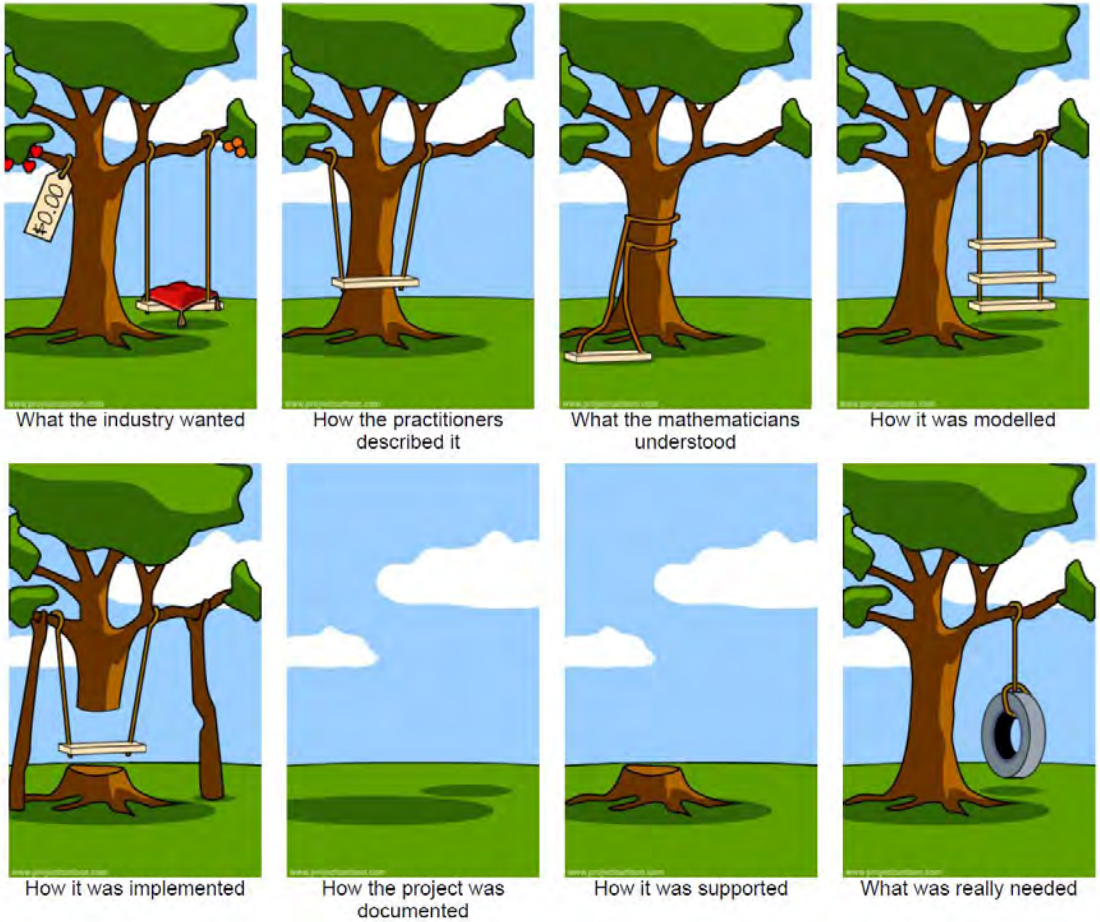
15 Prozent. QED. Oder gibt es schon falsche Fünfzehner?! Gunter Dueck. Das Sintflutprinzip. Springer. 2000

Think and act problem centric

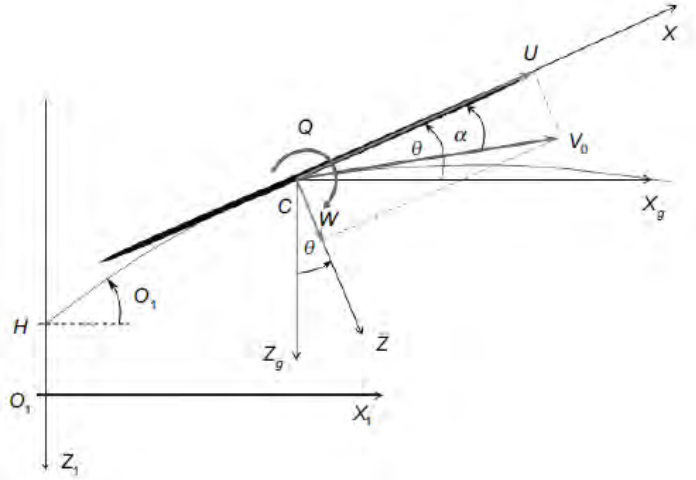
- It is more important to solve the right problem than to solve the problem right.
- Identifying the problem is half of the way to the solution.



Describing the problem



Javlin throwing: Theory

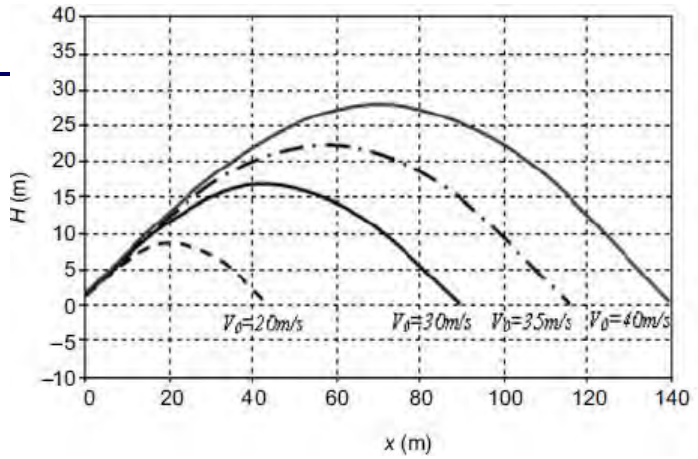


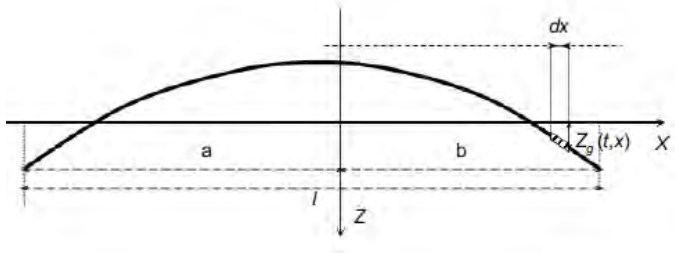
$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial U}\right)+\frac{\partial T^*}{\partial W}Q=Q_U^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial W}\right)+\frac{\partial T^*}{\partial U}Q=Q_w^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial Q}\right)+\frac{\partial T^*}{\partial U}W-\frac{\partial T^*}{\partial W}U=Q_Q^*$$

$$\frac{d}{dt}\left(\frac{\partial T^*}{\partial q}\right)-\frac{\partial T^*}{\partial q}+\frac{\partial V_{zg}}{\partial q}U=Q_q^* \qquad (11)$$



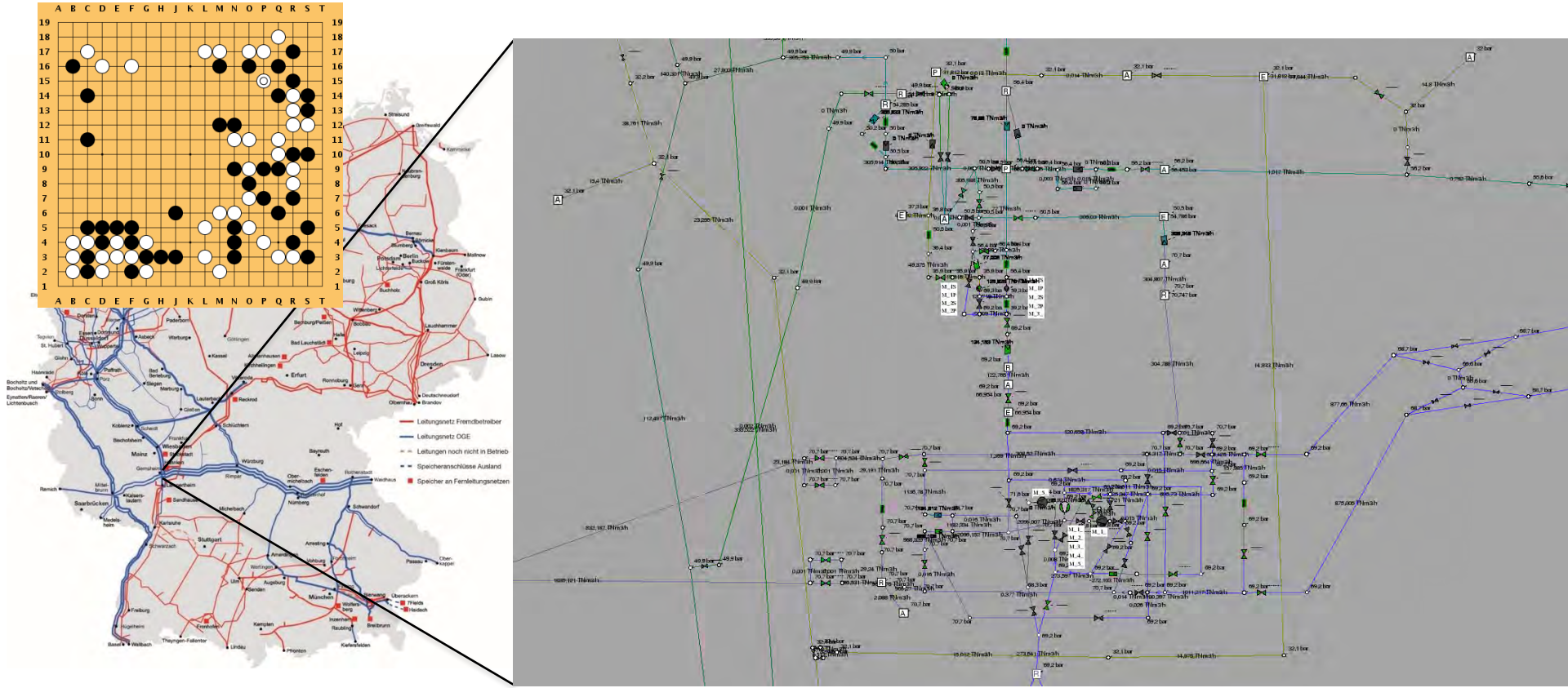


Mathematical Modeling and Numerical Simulations of Javelin Throw
J. Maryniak, E. Ładyżyńska-Kozdraś, E. Golihska
J. of Human Movement, Vol. 1 (2009), 16-20

Javlin throwing: Sports

Javelin throwing: Real life practice

Complex Networks



3 reasons why a question in industrial optimization might be difficult to answer

1. The question is not well defined,
   i.e., the modeling is intricate. Very often, in industry, problems are involved and multi-layered. Determining a precise definition of the problem, the input and output data, and mapping this to a mathematically well-defined computable optimization problem can be challenging.

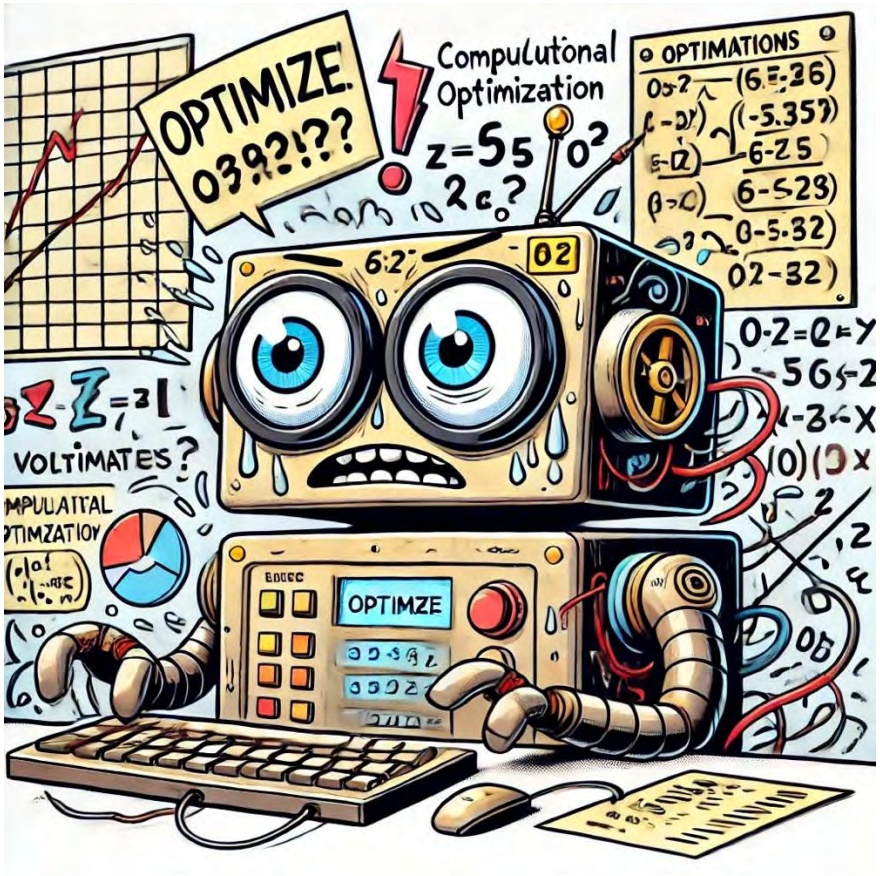
   - Given a solution can you decide if it is feasible to the original problem?
   - Given two solutions can you decide which one is better?

2. The data needed to solve the problem is not fully available.
   Many companies struggle hard to consolidate their IT. Getting out precise numbers is often surprisingly hard. One fundamental reason is decomposition, which has been necessary, at least in the past, to counter complexity. As a result, everyone only sees either a very little or very simplified part of the whole picture, and it is very hard to impossible to collect and the data into a coherent set.

   - Can you check whether the data is complete, correct, and consistent?
   - Is data on possibilities available?

3. The resulting problem is computationally hard to solve.
   Since the complexity class of discrete optimization problems often is NP-hard or worse, this is not surprising. However, experience shows, that solving particular instances works surprisingly well and that usually, the main reason for the inability to solve a problem is its size. For example, the likes of SAP, Amazon, Google, Huawei all have extremely large-scale supply-chain-type problems at hand. But not so many others. And there are surprisingly few small challenging real-world problems unless the time allowed for solving is very short.

Dall-E: Please draw a Computational Optimization Robot at Work



Fragen

有問題嗎

คำถาม

?

質問

Вопросы

Questions

Câu hỏi

Thank you very much!