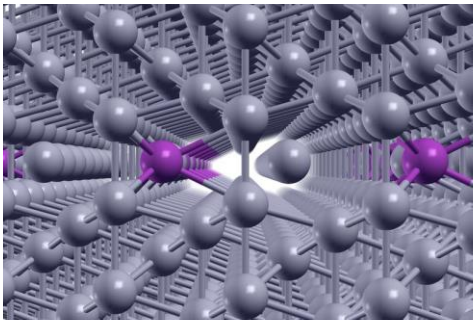
i-KNOW DATA SCIENCE CHALLENGE

sponsored by
T-Mobile Austria

task and data set provided by
Materials Center Leoben

DATA SCIENCE CHALLENGE – TASK ANNOUNCEMENT

An alloy is a mixture of metals or a metal and another element. By selecting which elements are mixed and in which proportion, alloys can be custom-tailored to meet diverse application requirements. The distribution of alloying elements plays a major role in determining the material properties. It is governed by the segregation energy, which in turn depends on the local coordination of atoms.



Advanced computing codes can compute the segregation energy based on the atom coordinates, yet at a considerable computational cost (i.e., days of computing on large clusters for a single alloy configuration). Although the segregation energy has been computed for many alloys, a much bigger number of potential alloys remains unexplored. One of the grand challenges in materials science is to identify the most promising alloy candidates for a set of basic elements and a given target segregation energy.

Our data set contains selected configurations of a specific alloy family together with the computed segregation energy. Is it possible to identify the most promising alloy candidates without explicitly computing all segregation energies? In case the scientific background of the problem sparks your interest, a detailed paper can be found here [1], though it is definitely not a requirement for the Data Science Challenge. The data can be treated as originating from a non-physical system, if desired.

The presented data set gives the segregation energy for configurations described by 60 atom positions each. There are 222 configurations in the full data set, which consists of two parts. Part one, *training_set.xml*, has 200 entries with sets of atom positions and the corresponding pre-calculated values for segregation energy. Your task is to devise a solution for predicting the segregation energy based on the atom positions.

```
<entry_n>
    <energy>[e]</energy>
    <coordinates>
        <c>[x] [y] [z]</c>
        <c>[x] [y] [z]</c>
           .
           :
    </coordinates>
</entry_n>
```

Part two, *validation_set.xml*, contains the remaining 22 entries with atom positions but no segregation energy values. Based on the solution for the first part of the data set, you will have to determine the entry numbers and the predicted energy values for part two of the data set, as shown in *return_file_template.csv*. Until September 20, 2017, 23:55 MEZ (UTC+1) we will accept one preliminary submission per participant (team or individual) in the form of the aforementioned csv-file containing the 22 entries for the second part of the dataset. Please submit your answers via e-mail (dsc@know-center.at). The results will be ranked based on the mean squared error. This intermediate submission is purely optional. If you do not wish to be a part of preliminary ranking, you do not have to turn in a solution by that date.

Until September 30, 2017, 23:55 MEZ (UTC+1) you can submit your final solution as csv-file. Once again we will rank the submissions by the mean squared error. The 5 top participants will be asked to provide their source code or demonstrate their solution during a live session with the jury and describe their approach, applied methods and tools. We strongly encourage publishing your source code in a public repository (github, zenodo, etc.). After verifying the results, we will provide you with the full dataset and announce the winners.

Best of luck to you and happy number-crunching!

[1] Daniel Scheiber, Vsevolod I. Razumovskiy, Peter Puschnig, Reinhard Pippan, Lorenz Romaner, Ab initio description of segregation and cohesion of grain boundaries in W–25at.% Re alloys, Acta Materialia, Volume 88, 2015, Pages 180-189, ISSN 1359-6454, http://dx.doi.org/10.1016/j.actamat.2014.12.053.

Know-Center GmbH || Research Center for Data-Driven Business and Big Data Analytics ||
Inffeldgasse 13/6 || 8010 Graz || Austria || office@know-center.at || Tel.+43 316 873 30800 ||
Fax+43 316 873 1030800 || Firmenbuchgericht Graz || FN 199 685 f || UID: ATU 50367703 ||
http://www.know-center.at || © Know-Center GmbH 2017-09-11 ||

Funded through the COMET-program (Competence Centers for Excellent Technologies), we thank our funding bodies:
COMET bmvit bmwfw FFG Das Land Steiermark SFG