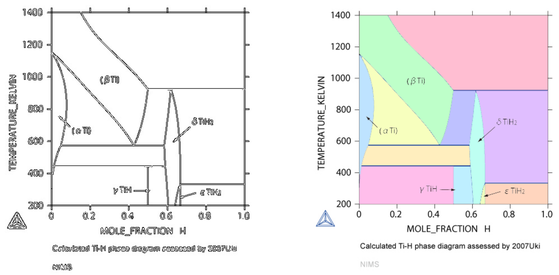
CIF21 DIBBs: EI: Materials Data Engineering (MaDE) Laboratory

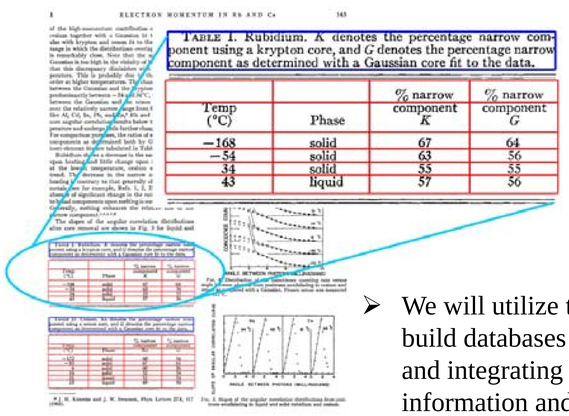
Venu Govindaraju (PI), Krishna Rajan, Thomas Furlani, Srirangaraj Setlur, Scott Broderick
University at Buffalo – The State University of New York

Document Image Processing Toolkit

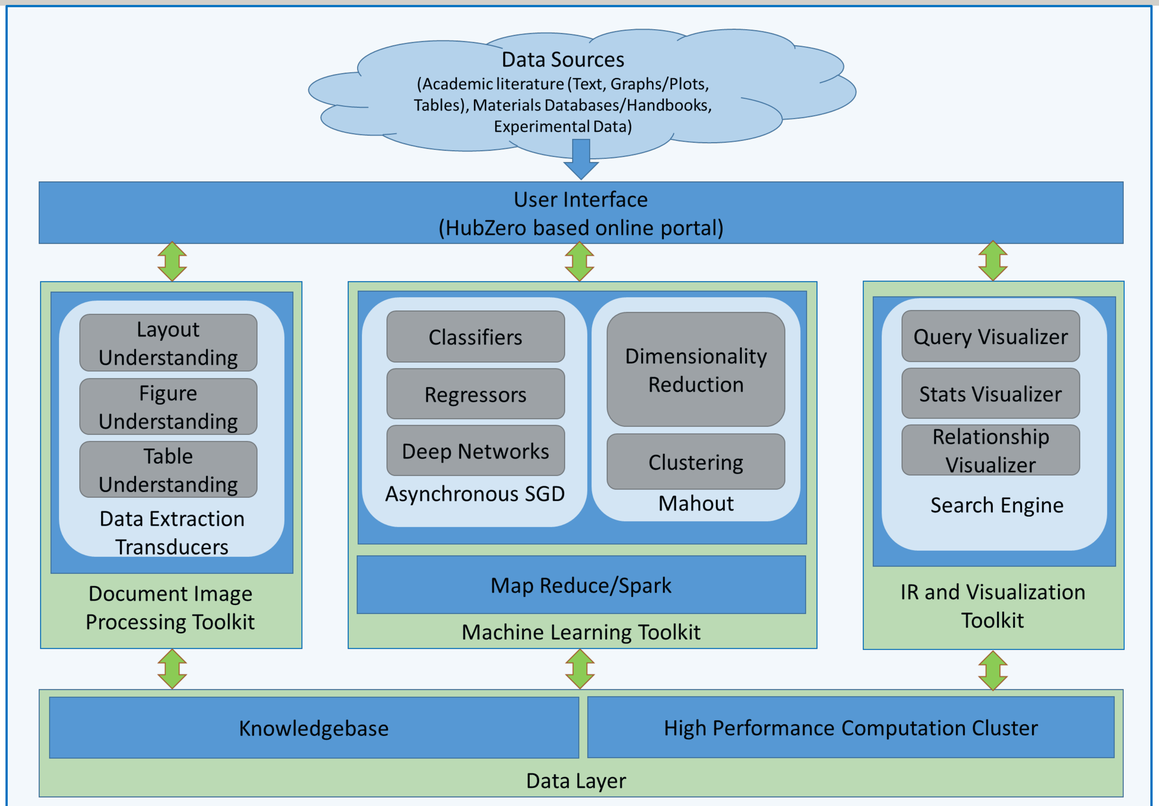
- The building blocks enable researchers in a variety of scientific domains to utilize the power of machine learning for extracting rich information from scientific publications, technical handbooks and databases that can facilitate innovative data analytics approaches.
- The building block modules of image processing and predictive machine learning algorithms will provide users with the ability to query and visualize interesting data as well as run simulations and experiments.



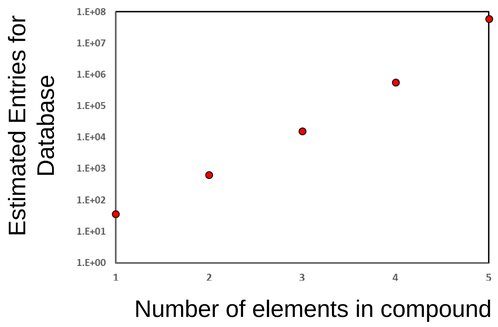
- (Above) Edge detection and region filling to index phase diagrams to support search and query on alloy properties
- (Below) A sample of our table and caption detector on a Physics article published in 1968



- We will utilize this approach to build databases from legacy data, and integrating graphical information and tabular data.
- As an example, we will use image processing algorithms to develop digitized and queryable phase diagrams which we will then link with machine learning property predictions.



Data Layer

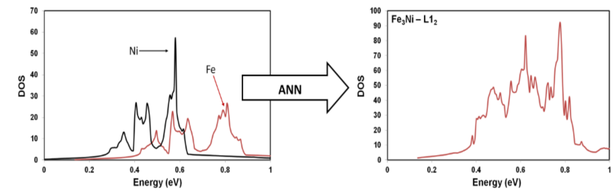


- Given the increased complexity in physics, the number of entries in databases typically decreases with increased chemical complexity
- This project addresses that challenge and increases entries with complexity.

- Our database is based on machine learning of the physics without explicit definition. Therefore, we are able to model properties as the complexity increases (ex. Density of states spectra) --- addresses current design limitation where high density of properties are known for limited regions in the chemical design space
- Our database will contain ~30 elemental entries, but ~ 15,000 ternary entries, and ~ 500,000 quaternary entries.

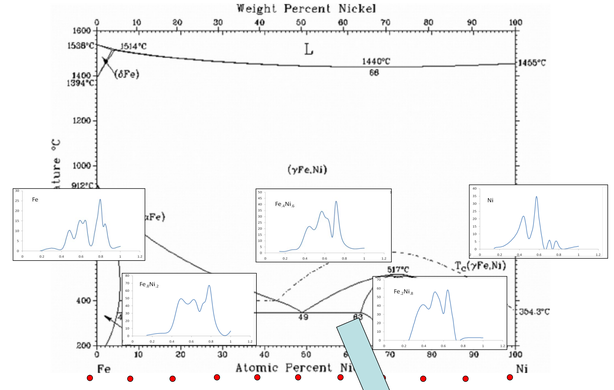
Machine Learning Toolkit

- Databases and literature contain many scientific diagrams which are not used as an input for property predictions
  - Ex. Phase diagrams and density of states (DOS) spectra
- Our approach uses these diagrams to predict properties.

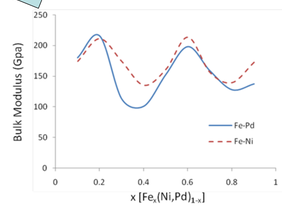


- Example of machine learning for accelerated modeling of material properties by modeling multi-component alloy DOS as function of elemental DOS
- This approach addresses the limitations of first principles approaches by not requiring specific definition of pair potentials.
- We then extract properties from these spectra.

Modeling Across Compositional Spreads



- This approach allows us to do massive numbers of calculations across all compositions without requiring additional first principles.
- Even in approaches performing large numbers of calculations, the values are still discontinuous – our approach allows for continuous property mappings.



Property Extraction Across the Compositional Space

Award Number : 1640867