EDUARDO BRITO & GULAR SAMADOVA & ELVIN MIRZAZADA

K-D TREES FOR SPACE INDEXING ON LARGE EMPLOYMENT DATABASES

MOTIVATION

Large companies, with thousands of employees, may find it difficult to effectively model, query and search through their human resources' databases in order to find the best matching individuals for new tasks, projects, or teams.

Name	Employees
Walmart WMT	2,200,000
Amazon AMZN	1,271,000
Volkswagen VOW3.DE	662,653
FedEx FDX	570,000
Deutsche Post DPW.DE	565,053

DOZENS OF DIFFERENT POSITIONS

HUNDREDS OF QUANTIFIABLE SKILLS

PROBLEM MODELING

This question of finding the best matching people for a specific task can be adapted to a range search problem in a multidimensional space. Each individual is modelled as a point with K dimensions/coordinates, corresponding to all the available/quantifiable skills.

ROLES BINARY VALUES EX: {0,1}

SKILLS INTEGER VALUES EX: [0,100]

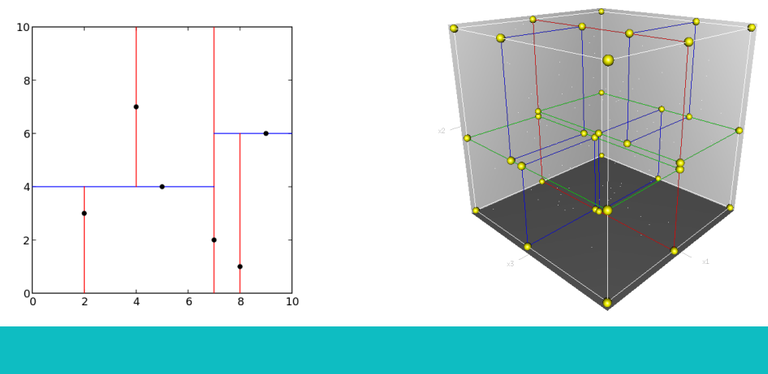
QUERYING

After modelling the database, the goal is to design a system that can answer quantifiable questions as range search queries to find the points that fit into the desired search space.

FIND ALL THE **DEVELOPERS** PROFICIENT IN **PYTHON [70,90]** THAT CAN WORK IN AN **AGILE ENVIRONMENT [40,80]** WITH GOOD **ENGLISH SKILLS [60,100]**

K-DIMENSIONAL TREE

After identifying the targeted columns of the database as the corresponding skills/positions of the individuals involved in the query, these will be the axes of the search space. The K-D Tree is then built based on these splitting hyperplanes that allow to iteratively sort and divide the space into two parts.



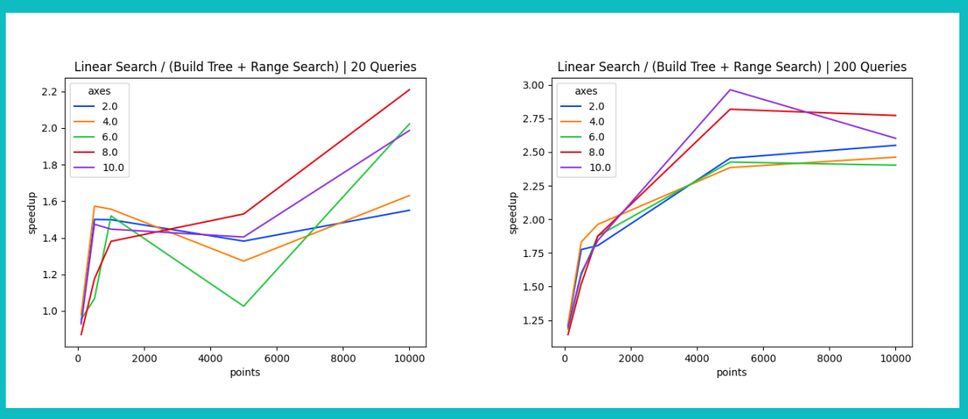
OTHER APPLICATIONS

The same approach can be applied to a variety of other problems where the information can be modelled as a set of points in a multidimensional space. It is highly beneficial especially for larger databases, a greater number of axes, and a good quantity of successive queries. Real-time systems could profit from this performance delivering faster answers for these spatial indexing questions.

GAMING **E-COMMERCE** **REAL ESTATE**

RANGE SEARCH PERFORMANCE

Having the K-D Tree, it is then possible to perform the desired range search queries in a similar fashion as the building process. Tests were conducted to measure the speedup gains relative to the traditional linear search, with successive queries on different parameters.



FUTURE WORK

An extension to this project could be to allow more complex queries with simultaneous creation of multiple K-D Trees and the combination of the results. Another idea is to include an additional dimension and treat the points as nodes in a graph, giving some meaning to the edges and solving related graph problems.

MULTI K-D TREES

GRAPH THEORY

REFERENCES

- Bentley, J. L. (1975). "Multidimensional binary search trees used for associative searching".
- Chandran, Sharat. Introduction to kd-trees. University of Maryland Department of Computer Science
- https://github.com/edurbrito/algo-project