DataPlay - Industry Practitioner

Jacob Yim

DataPlay Summary

- SQL has issues!

    - Discourages trial and error while building quantified queries

    - Lacks *syntax locality*: small changes to query semantics sometimes result in large changes to query structure

    - Does not display non-answers (tuples not satisfying a query): difficult to interpret query based on output and tell whether it is truly working

- Introducing DataPlay, a mixed-initiative query tool

    - Direct manipulation: graphical query language represents query as a tree, suggests changes

    - Auto-correction: answers and non-answers are displayed to provide live feedback. Can be marked as correct/incorrect, and the system will suggest changes to fix the query

Background

- Our company, Nyamazon.com, is a small startup selling books on an online marketplace
- However, we are looking to expand to sell many, many other things

Problem Statement

- Our small team has limited experience with database querying and SQL

    - In the past, we've had difficulty using complex SQL queries to perform analysis on our customer data

- We would like to perform analyses on our customers' searches and purchases using our new expanded catalogue

How can DataPlay help?

- Currently, using complex SQL queries to perform our analysis has proven difficult and time-consuming

- Our analysis often involves quantified queries

  - e.g. find all IDs of customers who made at least one purchase of cat food

- DataPlay may save us time and money by helping us iteratively build queries

- Direct manipulation may be especially helpful for our team members who are less familiar with SQL

DataPlay Pros and Cons

- Pros:

    - Save time and money building queries for data analysis

    - Graphical user interface could make using and understanding queries easier for team members less experienced with SQL

- Cons:

    - Learning curve

    - Performance on large datasets is unclear

    - Software potentially not ready for widespread use