

Resume Visualisasi Data dan Informasi – Minggu I

Rian Bintang Wijaya – 122450094 – RA

Definisi:

Visualisasi Data yakni merupakan upaya mengolah data abstrak yang ditransformasikan menjadi bentuk visual dengan berbagai cara seperti bentuk, warna sehingga mempermudah dalam penjelasan data, mengingat manusia lebih mudah menangkap informasi visual yang menarik. Hal ini mempermudah Data Scientist dalam penjelasan data dan penginterpretasian dari hasil analisis data.

Tahapan Visualisasi Data:

1. Import Data,
yakni mengumpulkan data dari sumber data
2. Preparasi Data,
yakni menyiapkan data untuk lebih mudah diolah dan dibaca sehingga mendapatkan hasil yang lebih baik untuk penyajian data dalam visualisasi (missing values, normalizing values)
3. Manipulasi Data,
yakni memilah data yang ingin divisualisasikan dengan berbagai opsi seperti join dan grouping
4. *Mapping*,
yakni memetakan data ke dalam bentuk geometri untuk mempermudah pembacaan seperti garis, titik dan persegi panjang dengan penambahan atribut untuk pembeda yang lebih jelas yakni warna, posisi dan ukuran
5. *Rendering*,
yakni menjadikan secara lengkap dalam bentuk representasi data visual menyesuaikan dengan sajian bentuk data yang dimiliki

Spesifikasi Visualisasi Data

1. Data,
 - Record yakni data yang akan divisualisasikan
 - Transformasi yakni pengoperasian dan pengolahan data menggunakan berbagai cara seperti filter, join dan sort untuk data yang lebih spesifik
2. Tanda/marks
 - Jenis, bentuk visual berupa bar, lingkaran, garis ataupun titik yang merepresentasikan data
 - Ukuran, lebar dan tinggi dari bentuk visualisasi data yang menunjukkan perbedaan makna ukuran data mana yang lebih besar ataupun kecil
 - Legenda, yakni informasi penjabar dari bentuk visual, misalkan penjelasan perbedaan warna ataupun jenis ketebalan garis yang mewakili informasi tertentu
3. Pemetaan, yakni penyesuaian korelasi data dengan tanda/marks

Bahasa Visualisasi Data

Visualisasi memiliki tingkatan bahasa yang menunjukkan level visualisasi data terkait yang bergantung pada bagaimana cara mengekspresikan dan menyajikan data.

1. Low-level (Prefuse, Flare, Protovis, D3, Vega, Reactive Vega)
2. High-level (ggplot, vega-lite, altair, echarts, VizQL, ZQL)
3. GUI-based tools (Tableau, Excel, Keshif, iVisDesigner, Qlik, Lyra, Google Sheets, Amazon Quicksight, Microsoft Power, Data Illustrator, Google Fusion Tables)
4. Underspecified Language (zenvisage, DeepEye, APT, SW, Eviza, Evizeon)

Efisiensi Pendekatan dalam Visualisasi Data

Visualisasi Data kerap kali mengalami kendala dalam upaya mencari efisiensi pendekatan yang terbaik dalam bentuk transformasi data terutama yang kompleksitasnya tinggi ke dalam bentuk visualisasi dengan aproksimasi yang terbaik. Hal ini dipermudah dengan berbagai tools yang dapat melakukan render data menjadi bentuk visualisasi dan umumnya menggunakan manipulasi data terlebih dahulu pada SQL.

1. Query Translation,
Cara manual menggunakan SQL/DBMS lainnya dengan langsung menginputkan query untuk menargetkan visualisasi yang diinginkan misal menggunakan DDL berupa select, from, group by, join, where, dll.
2. Sistem Integrasi Visualisasi oleh DBMS
DBMS memungkinkan agar kita dapat menampilkan visualisasi data sederhana menggunakan tools maupun filter operasi yang tersedia contohnya seperti Ermac, yakni Data Visualization Management System (DVMS) yang mendukung hubungan antara data dan skala yang memvisualisasikan data record sebagai Logical Visualization Plan (LVP)
3. Column Stores,
yakni manajemen data berupa layout yang berdasarkan baris dan kolom untuk meningkatkan performa pembacaan data
4. Index
Indeks mempermudah dalam pembacaan data pada tabel yang berisi kolom dan baris untuk mendetailkan data mana yang perlu dilakukan olah data sehingga meningkatkan pengolahan lanjutan dari data terkait untuk divisualisasikan oleh tools
5. Komputasi Paralel
Komputasi paralel juga telah banyak digunakan untuk pemrosesan kueri dalam sistem visualisasi data. Kueri agregasi pada *data tiles* di imMens diparalelkan menggunakan representasi indeks padat dari *data tile*. SeeDB menjalankan beberapa kueri SQL dari kandidat visualisasi secara paralel selama peringkat visualisasi. Harald et al. menyediakan arsitektur *multi-threading* untuk eksplorasi visualisasi interaktif. Arsitektur ini mempertahankan utas aplikasi utama untuk menangkap permintaan interaksi pengguna dan beberapa utas visualisasi untuk setiap visualisasi guna memproses visualisasi dari utas tersebut. Lebih lanjut, apakah

utas utama dan utas visualisasi bersifat asinkron atau sinkron tergantung pada jenis permintaan interaksi pengguna.

6. Prediksi dan Caching Data

Eksplorasi data sangat penting dalam visualisasi data, di mana pengguna terus-menerus menjelajahi visualisasi yang menarik minat mereka untuk memahami apa yang perlu divisualisasikan. Kerap kali, visualisasi yang sedang dieksplor saat ini biasanya terinspirasi dari visualisasi sebelumnya yang memungkinkan pengguna untuk mendapatkan visualisasi berikutnya dengan mengubah parameter visualisasi saat ini atau memperbesar/memperkecil untuk mendapatkan informasi yang lebih detail. Memprediksi data berikutnya akan menarik bagi pengguna, kemudian melakukan prefetching atau caching data yang mungkin digunakan pada langkah berikutnya selama eksplorasi saat ini dapat mempercepat proses eksplorasi, dan teknik-teknik ini telah digunakan di banyak sistem visualisasi.

Aproksimasi Visualisasi Data

Data berkembang volume nya secara eksponen yang mengakibatkan pemrosesan data secara sederhana akan menyulitkan hasil pemrosesan data, hal ini dibantu dengan percepatan pemrosesan data dengan AQP (Approximate Query Processing) yang memiliki tiga perspektif.

1. AQP Based

Menggunakan subset representatif dari data dapat memberikan perkiraan visualisasi kepada pengguna interaksi online dengan mengorbankan kualitas.

2. Increment Sampling Based

mencoba melakukannya menghubungkan teknik kueri data tambahan dengan visualisasi data. Ide kunci dari perkiraan visualisasi dengan inkremental pengambilan sampel adalah sistem menghasilkan perkiraan visualisasi berdasarkan sampel representatif dari kumpulan data dengan cepat. Kemudian, sistem meningkatkan ukuran sampel dari waktu ke waktu untuk terus meningkatkan kualitas visualisasi. Pengguna biasanya bisa mendapatkan beberapa wawasan awal dari perkiraan visualisasi dan memutuskan untuk menghentikan jika kualitas visualisasi sudah cukup untuk memverifikasi wawasan ini.

3. Human Perception Based

Terkadang, meningkatkan sampel ukuran tidak selalu meningkatkan kualitas visualisasi. Alasan eksternalnya adalah jumlah pikselnya layarnya terbatas, dan alasan internalnya adalah kognitif keterbatasan persepsi manusia dalam mengidentifikasi detail-detail kecil. Oleh karena itu, sistem visualisasi perkiraan dapat menghasilkan hasil perkiraan berdasarkan representasi sampel tetapi dengan dampak minimal pada kualitas visualisasi. Pendekatan berbasis persepsi manusia menghentikan pengambilan sampel ketika tidak ada perbedaan nyata pada persepsi manusia antara perkiraan visualisasi saat ini dan visualisasi yang diperoleh dengan pengambilan sampel lebih lanjut

Visualisasi data adalah bidang yang berkembang pesat dengan banyak hasil penelitian baru dan sistem inovatif yang baru-baru ini dikembangkan. Penelitian dan praktisi dari berbagai bidang telah berkontribusi pada kesuksesan luar biasa visualisasi data, yang didorong oleh sebagian besar (jika tidak semua) domain dan aplikasi. Artikel ini terutama mengulas karya-karya terbaru dalam visualisasi data dari perspektif manajemen data. Secara khusus, kami telah menggambarkan secara komprehensif karya-karya dalam spesifikasi visualisasi, metode efisien untuk visualisasi data, dan rekomendasi visualisasi. Seperti yang disebutkan sebelumnya, sebagian besar sistem visualisasi data komersial unggul dalam hal kemudahan penggunaan dalam spesifikasi visualisasi data. Namun, banyak praktisi masih menghadapi masalah efisiensi dan rekomendasi dari sistem-sistem ini.