

Nama: Tessa Kania Sagala

NIM: 122450040

Tugas resume jurnal dengan judul Making data visualization more efficient and effective: a survey  
Visualisasi Data & Informasi kelas RA

Visualisasi data sangat penting dalam dunia bisnis berbasis data saat ini, yang telah banyak digunakan untuk membantu pengambilan keputusan yang berkaitan erat dengan pendapatan besar banyak perusahaan industri. Namun, karena tingginya permintaan pemrosesan data berdasarkan volume, kecepatan, dan kebenaran data, muncul kebutuhan akan ahli database untuk membantu visualisasi data yang efisien dan efektif.

Selain itu, visualisasi data juga telah banyak digunakan dalam banyak aplikasi yang berhubungan dengan database, seperti Excel, Google Spreadsheet, Desktop Visualisasi Data Oracle, IBM DB2, Tinjauan Cepat Amazon, Microsoft Power BI, dan banyak lainnya.

### **The Pipeline of Data Visualization**

1. **Data Import** adalah mengambil data yang diperlukan dari sumber data yang diinginkan.
2. **Data Preparation** adalah mempersiapkan data yang diimpor untuk visualisasi, misalnya dengan menormalkan nilai, mengoreksi entri yang salah, dan menginterpolasi nilai yang hilang.
3. **Data Manipulation** adalah memilih data yang akan divisualisasikan (atau pemfilteran dari komunitas visualisasi) dan mungkin dengan operasi umum lainnya seperti penggabungan dan pengelompokan.
4. **Mapping** adalah memetakan data yang diperoleh dari proses di atas ke primitif geometris (misalnya titik dan garis), beserta atributnya (misalnya warna, posisi, dan ukuran).
5. **Rendering** adalah mengubah data geometris di atas menjadi representasi visual.

Berdasarkan pipeline ada tiga pendekatan yang membuat visualisasi data lebih efisien dan efektif, namun tetap relevan bagi peneliti database.

1. **Spesifikasi Visualisasi:** Spesifikasi visualisasi menyediakan berbagai cara agar pengguna dapat menentukan apa yang mereka inginkan, dikarenakan:

2. **Pendekatan Efisien untuk Visualisasi Data:** Agar dapat melibatkan pengguna secara efektif dalam alur berulang, proses pembuatan visualisasi data harus efisien dan terukur, terutama untuk dua komponen, “Manipulasi Data” dan “Pemetaan”.

3. **Rekomendasi Visualisasi Data** Menentukan visualisasi secara tepat sulit dilakukan, bahkan bagi para ahli, hanya karena pemahamannya data apa yang akan divisualisasikan, cerita mana yang harus diceritakan, dan bagaimana memvisualisasikan. Oleh karena itu, penting agar sistem visualisasi dapat memandu pengguna secara cerdas dengan memberikan rekomendasi.

### **Spesifikasi visualisasi data**

Secara umum, bahasa visualisasi data terdiri dari tiga bagian: data, tanda (atau isyarat visual), dan pemetaan di antara keduanya.

#### **– Data**

- **Records:** data yang perlu divisualisasikan.
- **Transformasi:** operasi—seperti grup, bin, filter, dan pengurutan—digunakan untuk mengubah rekaman data tertentu.

#### **– Tanda (atau isyarat visual)**

- **Jenis:** representasi visual untuk rekaman data, seperti batang, garis, atau titik.
- **Ukuran:** lebar, tinggi visualisasi.
- **Legenda:** informasi legenda.
- **Aneka ragam:** properti lainnya, seperti lebar dan warna batang.

– **Pemetaan:** memetakan data ke tanda yang sesuai.

Operasi visual berbasis GUI biasanya diterjemahkan ke dalam bahasa visualisasi data.

### **Kategorisasi bahasa visualisasi data**

Strategi yang umum digunakan untuk mengkategorikan bahasa visualisasi data didasarkan pada ekspresifnya semakin rendah tingkat suatu bahasa, semakin ekspresif bahasa tersebut. Bahasa tingkat tinggi merangkum beberapa detail tingkat rendah dengan memberikan default yang masuk akal dan menambahkan lebih banyak batasan semakin tinggi tingkat bahasanya, semakin mudah digunakan.

**Visualisasi Data Interaktif** Racionalitas di balik visualisasi data interaktif adalah bahwa dalam banyak kasus, visualisasi data adalah proses eksplorasi, di mana pengguna harus terus menyempurnakan spesifikasi (misalnya, menambah/menghapus/mengubah atribut, mengubah jenis bagan) dari visualisasi yang dieksplorasi saat ini hingga mendapatkan hasil yang diinginkan. visualisasi yang diinginkan dalam proses eksplorasi.

### **Spesifikasi yang tidak ditentukan**

Visualisasi tidak ada artinya jika tidak dapat memberikan gambaran tentang data. Namun, dalam banyak kasus, pengguna tidak benar-benar mengetahui semua aspek data yang ada, karena datanya mungkin besar dan datanya sering diperbarui. Oleh karena itu, hal ini menimbulkan persyaratan untuk mendukung spesifikasi yang tidak ditentukan. Secara umum, untuk spesifikasi yang tidak ditentukan, pengguna hanya memberikan beberapa “petunjuk”, dan merupakan tugas sistem visualisasi untuk menafsirkan masukan yang tidak ditentukan, dengan (mungkin) cara yang berbeda. Jenis petunjuk pertama adalah “berbasis referensi”, di mana pengguna memberikan visualisasi referensi sebagai benih dan sistem menyarankan visualisasi berdasarkan referensi tersebut. Jenis petunjuk kedua adalah “berbasis kata kunci”, dalam gaya Google. Tepat menerima sasaran tampilan data pengguna pada kolom yang diinginkan, misalnya, “menyajikan penundaan keberangkatan dan hubungan terjadwal”. Dengan kata lain, APT menentukan kolom untuk divisualisasikan dan kemudian merekomendasikan visualisasi yang memenuhi tujuan.

Jenis petunjuk ketiga adalah “berbasis bahasa alami”, yang mempertimbangkan konteks masukan pengguna dan status sistem dalam siklus eksplorasi data, bukan petunjuk satu kali dalam petunjuk “berbasis kata kunci”.

### **Pendekatan efisien untuk visualisasi data**

1. Visualisasi data yang tepat
2. Mengintegrasikan Sistem Visualisasi dengan DBMS
3. Column Stores
4. Indexes
5. Perhitungan Paralel
6. Prediksi dan Pengambilan Awal

Pada prefetch dan prediksi dikategorikan lagi menjadi dua jenis yaitu:

1. Visualisasi yang Sedang Dieksplorasi.
2. Data Historis Ketika data historis tersedia, tentu saja, sistem dapat melakukan inferensi yang lebih rumit namun bermakna dibandingkan memilih arah secara acak seperti yang dibahas di atas.

### **Perkiraan visualisasi data**

1. Berbasis AQP. Cara mudah untuk menghasilkan perkiraan visualisasi dalam waktu interaktif adalah dengan memanfaatkan teknik AQP. Menggunakan subkumpulan data yang representatif dapat memberikan perkiraan visualisasi interaksi online kepada pengguna dengan mengorbankan kualitas.

2. Berbasis Pengambilan Sampel Tambahan. Gagasan utama visualisasi perkiraan dengan pengambilan sampel tambahan adalah sistem menghasilkan visualisasi perkiraan berdasarkan sampel kumpulan data yang representatif dengan cepat. Kemudian, sistem meningkatkan ukuran sampel dari waktu ke waktu untuk terus meningkatkan kualitas visualisasi. Pengguna biasanya dapat memperoleh beberapa wawasan awal dari perkiraan visualisasi dan memutuskan untuk menghentikannya jika kualitas visualisasi cukup untuk memverifikasi wawasan ini.
3. Berbasis Persepsi Manusia. Terkadang, menambah ukuran sampel tidak selalu meningkatkan kualitas visualisasi. Alasan eksternalnya adalah jumlah piksel layar yang terbatas, dan alasan internalnya adalah keterbatasan kognitif persepsi manusia dalam mengidentifikasi detail-detail kecil. Oleh karena itu, dimungkinkan untuk memperkirakan sistem visualisasi sistem untuk menghasilkan hasil perkiraan berdasarkan sampel yang representatif tetapi dengan dampak minimal terhadap 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 progresif

Secara umum, mereka membangun struktur hierarki dengan menggabungkan data pada tingkat yang berbeda, misalnya, ukuran wadah yang berbeda, rentang nilai temporal yang berbeda, zona nilai spasial yang berbeda. Kemudian, struktur hierarki ini digunakan untuk mendukung eksplorasi visualisasi progresif pengguna.

### Rekomendasi berdasarkan spesifikasi

1. Spesifikasi tidak lengkap. Sistem rekomendasi visualisasi dengan spesifikasi kosong tidak memerlukan masukan pengguna, sedangkan sistem rekomendasi dengan spesifikasi parsial menerima input spesifikasi elemen visualisasi parsial pengguna untuk visualisasi yang diinginkan. Misalnya, APT menerima sasaran tampilan data pengguna sebelumnya. Satu-satunya perbedaan antara spesifikasi kosong dan parsial adalah spesifikasi kosong harus memangkas ruang pencarian berdasarkan batasan yang ditentukan pengguna saat menghitung elemen visualisasi untuk menghasilkan kandidat visualisasi. Misalnya visualisasi yang tidak mengandung kolom *C* difilter dari kandidat visualisasi jika pengguna menentukan kolom *C* seperti kolom yang diminati di Voyager, dan visualisasi yang bukan diagram garis atau tidak memuat kolom “*listrik*” difilter dari kandidat visualisasi saat pengguna mengetik “*tunjukkan diagram garis tentang listrik*” di DeepEye. Dan tidak ada perbedaan saat memeringkat visualisasi kandidat.
2. Spesifikasi berbasis referensi. Beberapa sistem rekomendasi visualisasi merekomendasikan visualisasi berdasarkan data referensi atau visualisasi referensi. Biasanya, sistem akan merekomendasikan visualisasi yang serupa atau berbeda dari referensi yang diberikan dalam aspek tertentu.
3. Rekomendasi berbasis perilaku. Sistem rekomendasi berbasis perilaku menangkap perilaku pengguna saat ini sebagai masukan, kemudian menyimpulkan tugas yang diinginkan pengguna dan merekomendasikan visualisasi yang berguna berdasarkan tugas mereka. MEMANEN adalah sistem rekomendasi visualisasi berbasis perilaku. Ini merekomendasikan visualisasi berdasarkan tugas pengguna yang disimpulkan dari perilaku mereka. Karena sulit bagi pengguna untuk mendeskripsikan maksudnya dengan jelas dan tugas pengguna berkembang seiring proses eksplorasi, HAR-VEST menebak maksud pengguna berdasarkan perilaku mereka.

4. Rekomendasi yang dipersonalisasi. Sistem rekomendasi yang dipersonalisasi menangkap riwayat perilaku pengguna sebagai masukan untuk merekomendasikan visualisasi menarik yang dipersonalisasi.