

Metro Wordle: An Interactive Visualization for Urban Text Distributions Based on Wordle

Chenlu Li, Xiaoju Dong, Xiarou Yua
Visual Informatics
2018

Introdução

- Método tradicionais para recuperar localizações: mapas e ferramentas de buscas
- Mapas online (ex.: Google Maps):
 - *suportam buscas interativas, entretanto, faltam informações detalhadas.* Quando se faz uma pergunta, em geral o retorno é uma lista de documentos de texto com pouca informação geográfica
 - *com os métodos atuais, é difícil comparar grandes quantidades de POI's usando os dados textuais e a informação geográfica ao mesmo tempo.*

Objetivos da Visualização

- Permitir o acesso aos POI's através de palavras-chave (na visualização principal, somente um certo número de palavras-chave aparecem devido à limitação de espaço; todavia, através dessas palavras e com ajuda de filtros, o usuário pode localizar qualquer POI na cidade).
- Dispor informações textuais e geográficas de modo que possam ser observadas na mesma representação; dessa forma, características e relações espaciais podem ser descobertas.
- A visualização deve ser esteticamente interessante e “user-friendly”.

Trabalhos Relacionados

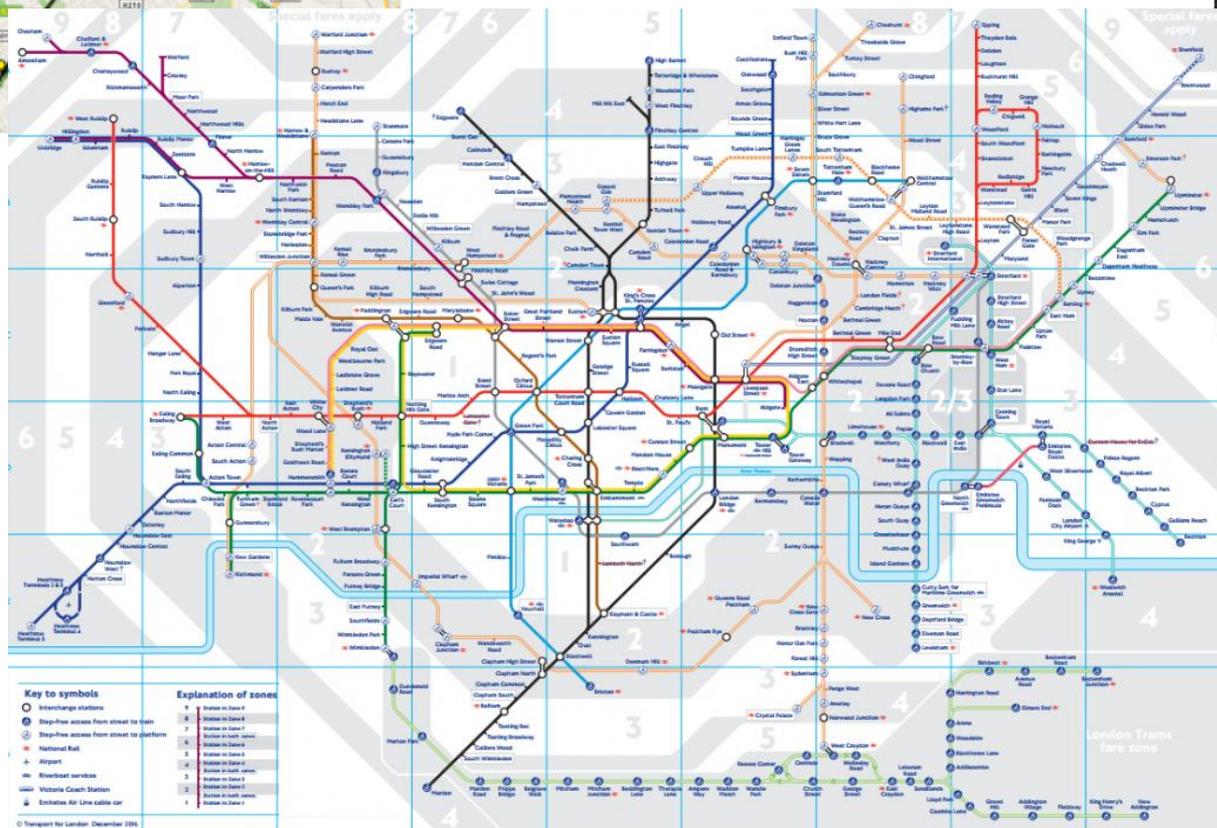
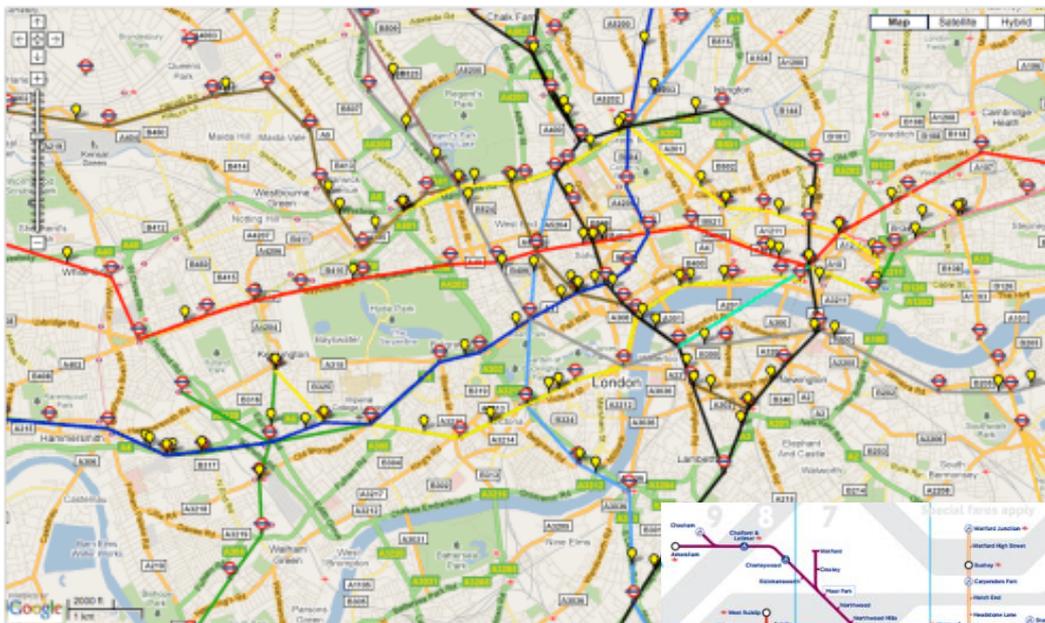
Wordle

- É uma técnica que cria representações visuais estéticas e impressionantes de entradas textuais.
- É uma extensão de tag clouds: tag clouds distribuem as palavras horizontalmente, linha por linha, com muitos espaços remanescentes entre as palavras. Wordle, contudo, permite que as palavras sejam dispostas em orientações distintas, visando obter um layout mais compacto.

Trabalhos Relacionados

Automatic Layout of Metro Map

- Um mapa de metrô é um desenho esquemático do sistema de metrô que enfatiza a conectividade topológica da rede, e não a posição geográfica.
- Desenhar um mapa de metro manualmente é uma atividade que demanda muito tempo. Sendo assim, técnicas para desenhar o layout do metrô de forma automática têm sido desenvolvidas:
 - Barkowsky, 2000: apresentou um método de simplificação da forma geográfica com o uso de “discrete curve evolution”.
 - Hong et al., 2006: aplicou “spring algorithm” para determinar a posição das estações do metrô.
 - Nöllenburg e Wolff, 2011: usaram “Mixed Integer Programming (MIP)” para desenhar o layout do metrô (como uso de restrições, é possível garantir a qualidade do layout).



Task 1: Obtain the Keywords of POI's

- Como o espaço do mapa é limitado, não é possível utilizar o nome dos POI's dentro do mapa. Sendo assim, foram obtidas palavras-chave que representam os POI's (obs.: essas palavras-chave devem ser, também, atrativas para o usuário).
- Se os POI's são restaurantes, por exemplo, a palavra-chave pode ser o prato mais importante daquele restaurante; se os POI's são lugares para passear, as atrações podem ser as palavras-chave.

Task 1: Obtain the Keywords of POI's

- As informações sobre os POI's foram obtidas na Internet.
- A partir dessas informações, palavras descritivas foram extraídas de comentários de usuários ou tags.
- Como não era possível (por limitação de espaço) usar todas as palavras descritivas, palavras similares foram clusterizadas.
- Essa clusterização baseava-se em similaridades semânticas e de localização, sendo assim, somente palavras que tinham similaridade semântica e posição próxima, ficavam no mesmo cluster (obs.: a clusterização foi feita por k-means).
- Para cada cluster foi dada uma única palavra e um peso, baseado no número de palavras no cluster. Esse peso foi usado posteriormente para definir o tamanho da palavra no mapa.

Task 2: Abstract the Metro Map



https://en.wikipedia.org/wiki/Shanghai_Metro

Task 2: Abstract the Metro Map

- Nesse trabalho, o mais importante é a estrutura topológica do metrô. Para tanto, o mapa do metrô foi abstraído em um grafo topológico.
- Primeiramente os nós-chave foram extraídos do mapa do metrô.
- As condições para defini-los foram:
 - A estação é uma estação de partida ou uma estação terminal
 - A estação é uma estação de transferência
 - A estação conecta duas linhas entre as quais o ângulo não está entre $0-45^\circ$ nem $135-180^\circ$

Task 2: Abstract the Metro Map

- Sendo assim, o mapa foi transformando em um grafo, cujos vértices são os nós-chave e a ligação entre as estações corresponde as arestas.
- O problema do layout do metrô é então transformado em um problema de desenho de grafo cujo objetivo é encontrar uma representação geométrica adequada para o grafo apresentado.
- Após obter o layout otimizado do grafo, as estações ignoradas são inseridas nos segmentos de reta no qual pertencem, a uma distância correta de seus nós-chave.
- Fazendo isso, a rede do metrô é simplificada e serve de base para o próximo processo de otimização do layout

Task 3: Statistical Analysis for Subareas

- As keywords foram marcadas com localizações (a latitude e longitude dois POI's relacionados).
- Usando a localização de um ponto e os limites da região, é possível descobrir se um ponto está dentro ou fora da região e descobrir a qual subárea cada ponto pertence no espaço geográfico real.
- A partir disso, uma posição relativa é guardada e baseada nessa posição relativa, os pontos e palavras serão colocados dentro do mapa distorcido e ficarão na mesma subárea no mapa esquematizado e no mapa real.
- Nessa etapa, também, calcula-se a área de cada região (A_{2j}).

Task 4: Automatic Layout of Metro Map

- Nessa etapa, o mapa do metrô deve ser distorcido para satisfazer as seguintes regras:
 - R1: todos os segmentos de reta devem ter orientação: horizontal, vertical ou 45-diagonal.
 - R2: A topologia da rede do metrô e a posição relativa entre as estações devem ser mantidas para coincidir com o mapa mental do usuário.
 - R3: Curvas ao longo de linhas individuais devem ser evitadas. Caso não seja possível, ângulo obtusos são preferidos do que ângulo agudos.
 - R4: O comprimento dos segmentos de reta entre dois vértices obtidos do grafo abstraído, deve ser mantido o mais uniforme possível.
 - R5: O tamanho das áreas deve ser balanceado com o número de POI's nela.
- R1-R4 já haviam sido estabelecidas por K. Garland (1994) quando desenhou o mapa do metrô de Londres.
- R5 foi definida nesse artigo para que na área de cada região caiba a sua informação textual.

Task 4: Automatic Layout of Metro Map

- Usando essas regras, buscou-se um layout que se compromete com as localizações geográficas das estações, a percepção do usuário e a quantidade de texto que será inserida nas regiões.
- O algoritmo MIP foi usado para desenhar o layout. Sendo assim, foram apresentadas ao MIP restrições (hards e softs) que fazem com que as regras R1-R5 sejam atendidas.
 - H1: para cada aresta, o segmento de reta deve ser octilinear.
 - H2: para cada vértice, a ordem circular de seus vizinhos deve permanecer como a do input.
 - H3: para cada aresta, o segmento de reta deve ter um comprimento mínimo.
 - H4: cada aresta deve ter uma distância de cada aresta não incidente para evitar intersecção.
 - S1: as linhas deve ter poucas curvas e ângulos de curva devem ser o maior possível.
 - S2: para cada par de vértices adjacentes, suas posições relativas devem ser preservadas.
 - S3: o comprimento total de arestas deve ser pequeno.
 - S4: para cada área, a relação entre a soma dos pesos de todas as palavras naquela região e o tamanho da região deve ser o mais próximo possível de 1.
- As “hard constraints” (H1-H4) servem para alcançar R1-R2 e para garantir a legibilidade do mapa do metrô.
- As “soft constraints” servem para alcançar R2-R5.

Task 4: Automatic Layout of Metro Map

- Por ser nova, o artigo explorou com mais detalhes a S4 (as outras restrições são descritas em outros artigos).
- S4: para cada área, a relação entre a soma dos pesos de todas as palavras naquela região e o tamanho da região deve ser o mais próximo possível de 1.

$$A1_j = \sum_i (n_i * S / K_j)$$

A1_j: espaço ocupado por todas as palavras naquela área

n_i: número de letras em uma palavra

S: tamanho do caractere

K_j: parâmetro de clusterização da área

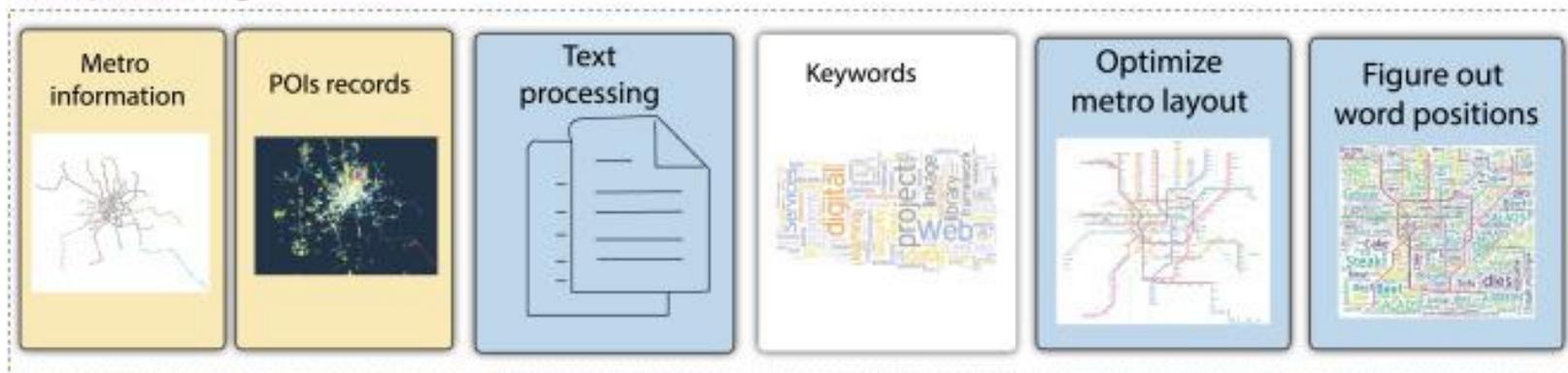
A2_j: é o tamanho da área que foi determinado no task 3

A1_j e A2_j devem ser o mais próximas possível.

Task 5: Word Embedding

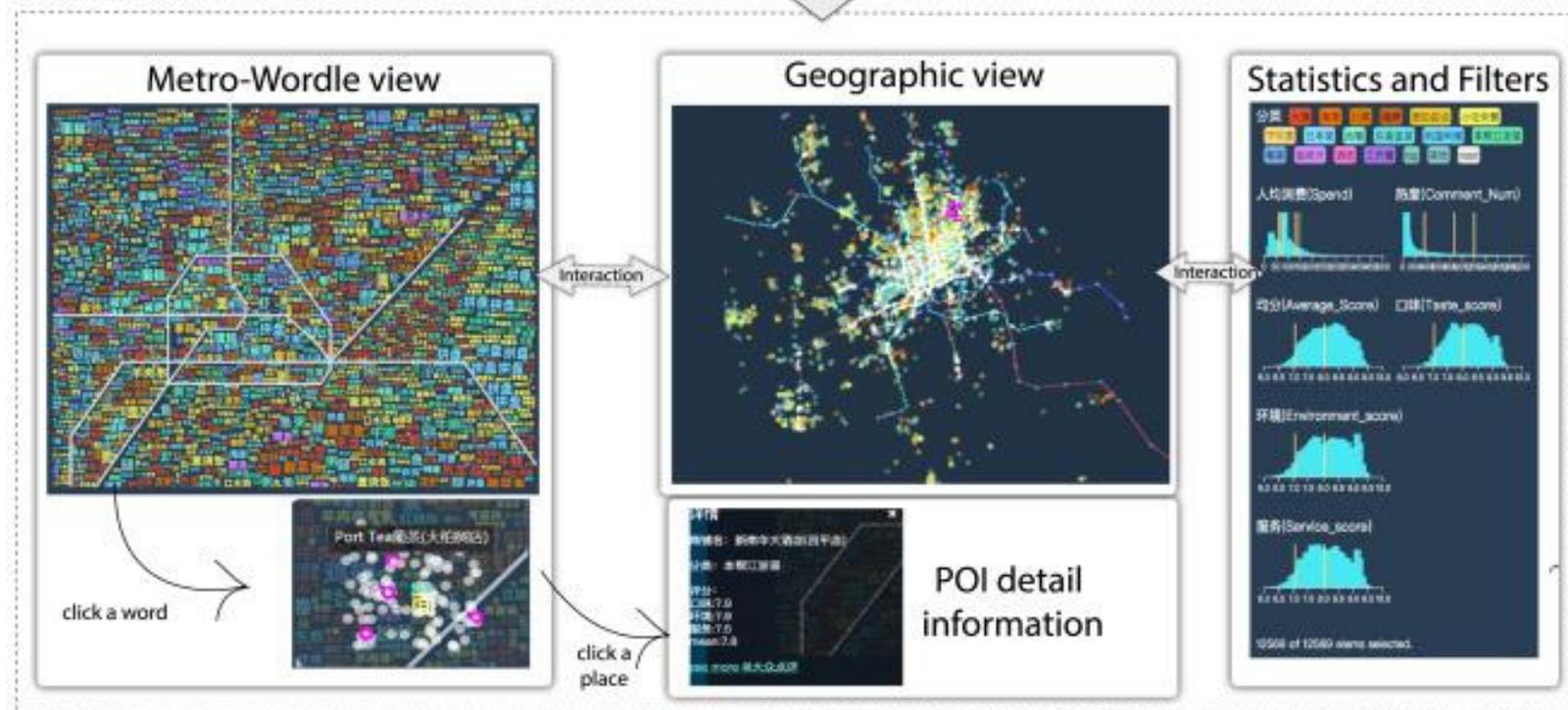
- As palavras eram posicionadas no mapa conforme a posição determinada no task 3. Essa é considerada a posição inicial.
- Se houvesse sobreposição com outras palavras ou linhas do metrô, tentava-se outra posição mexendo a palavra ao longo do limite da região e ao redor da posição inicial.
- Se não fosse possível achar uma melhor posição próxima, e tamanho da fonte era diminuído.

Data processing



Visualization

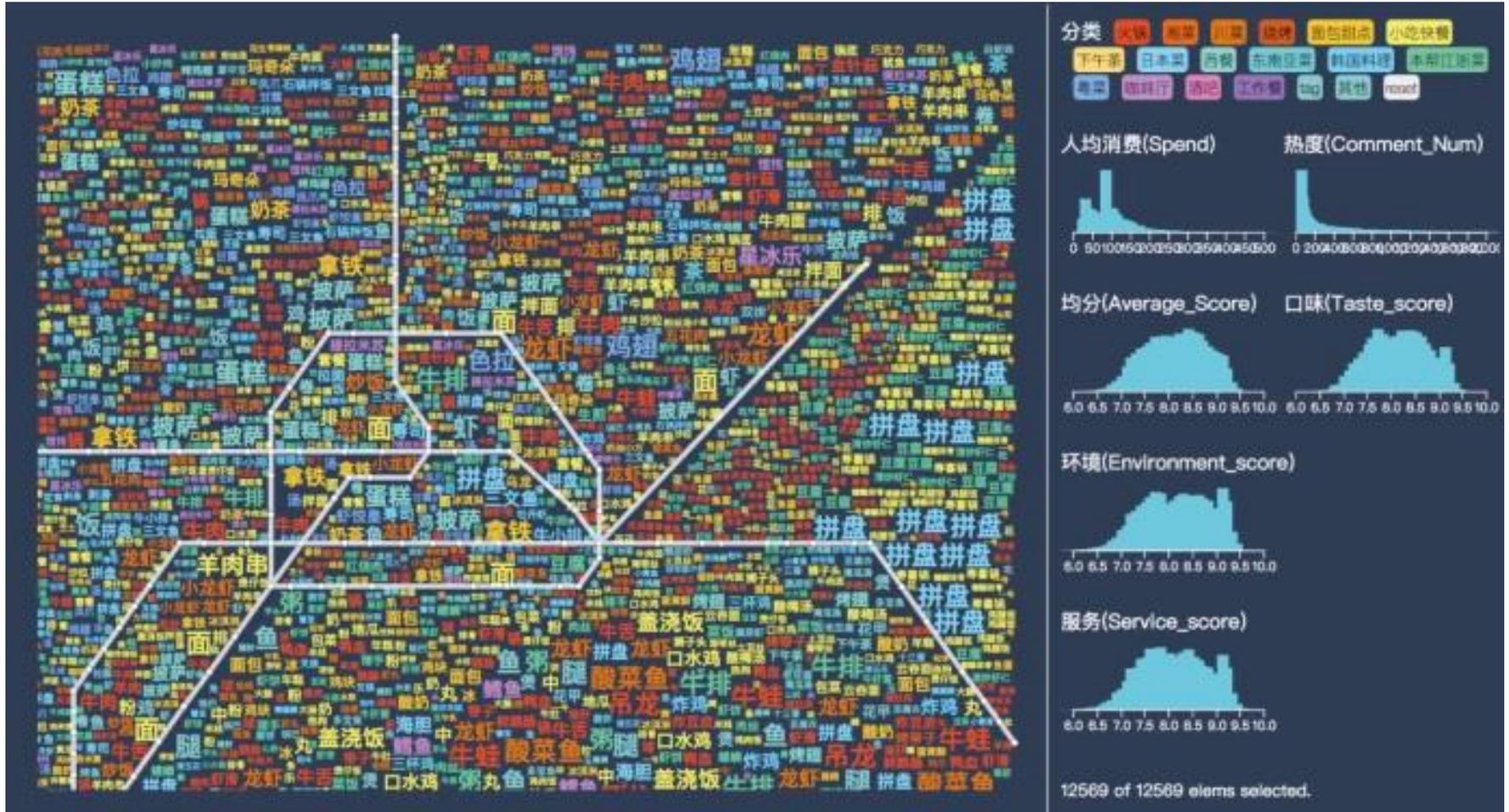
Processed Data



São obtidas 4 visualizações

- Metro-Wordle View
- Geographic View
- Statistic and Filter Module
- POI's detail View

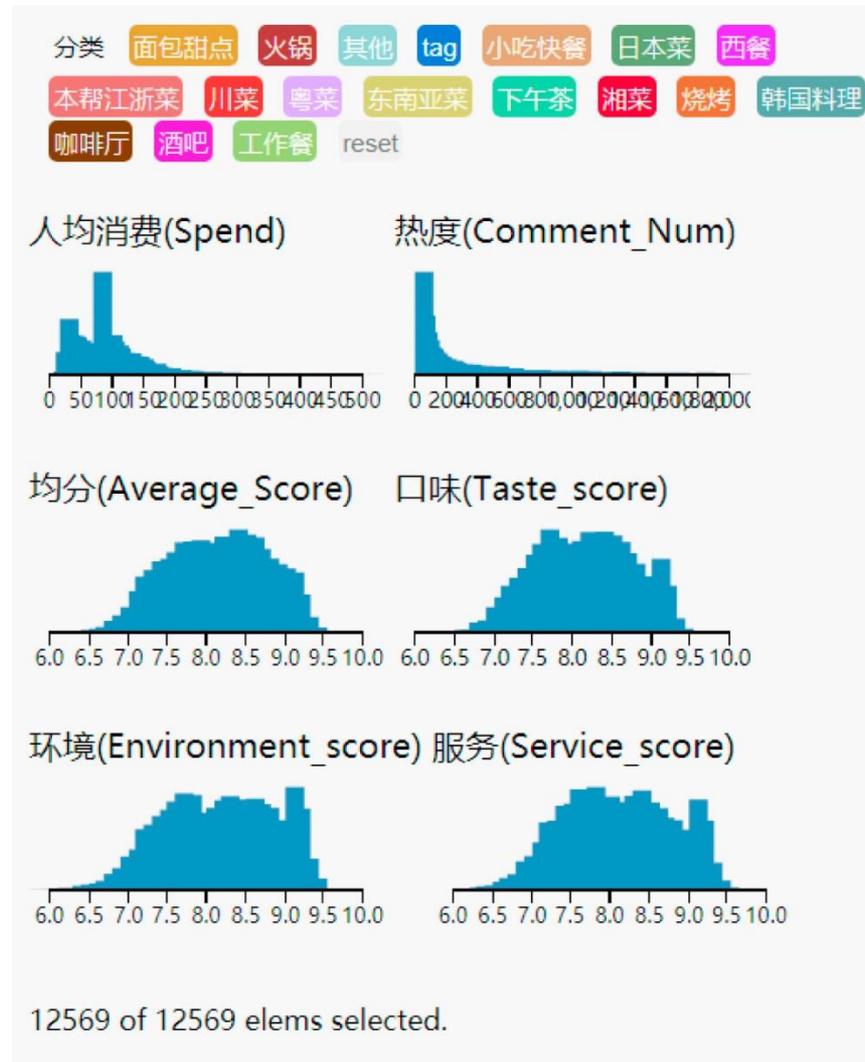
Metro-Wordle View



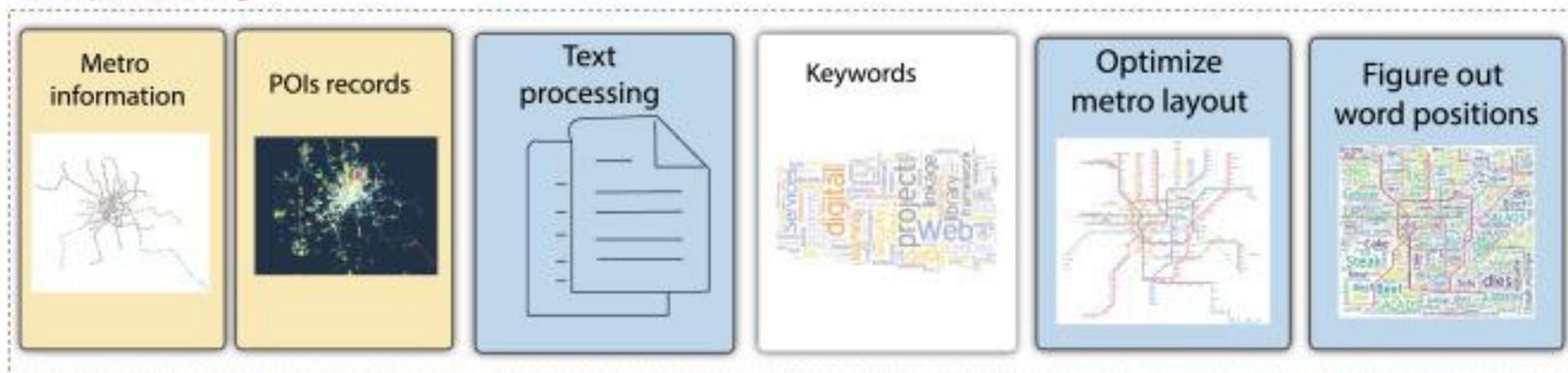
Geographic View



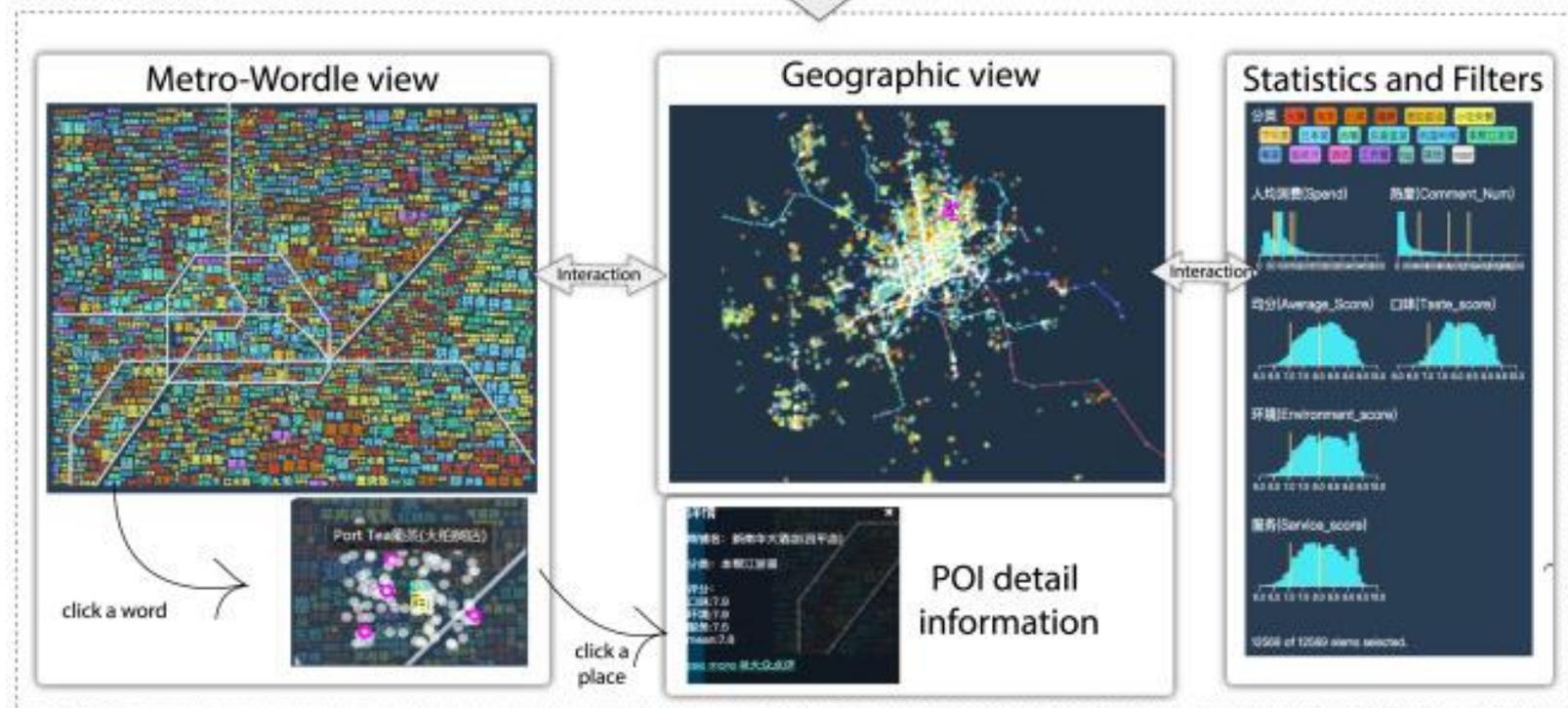
Statistics and Filter Module



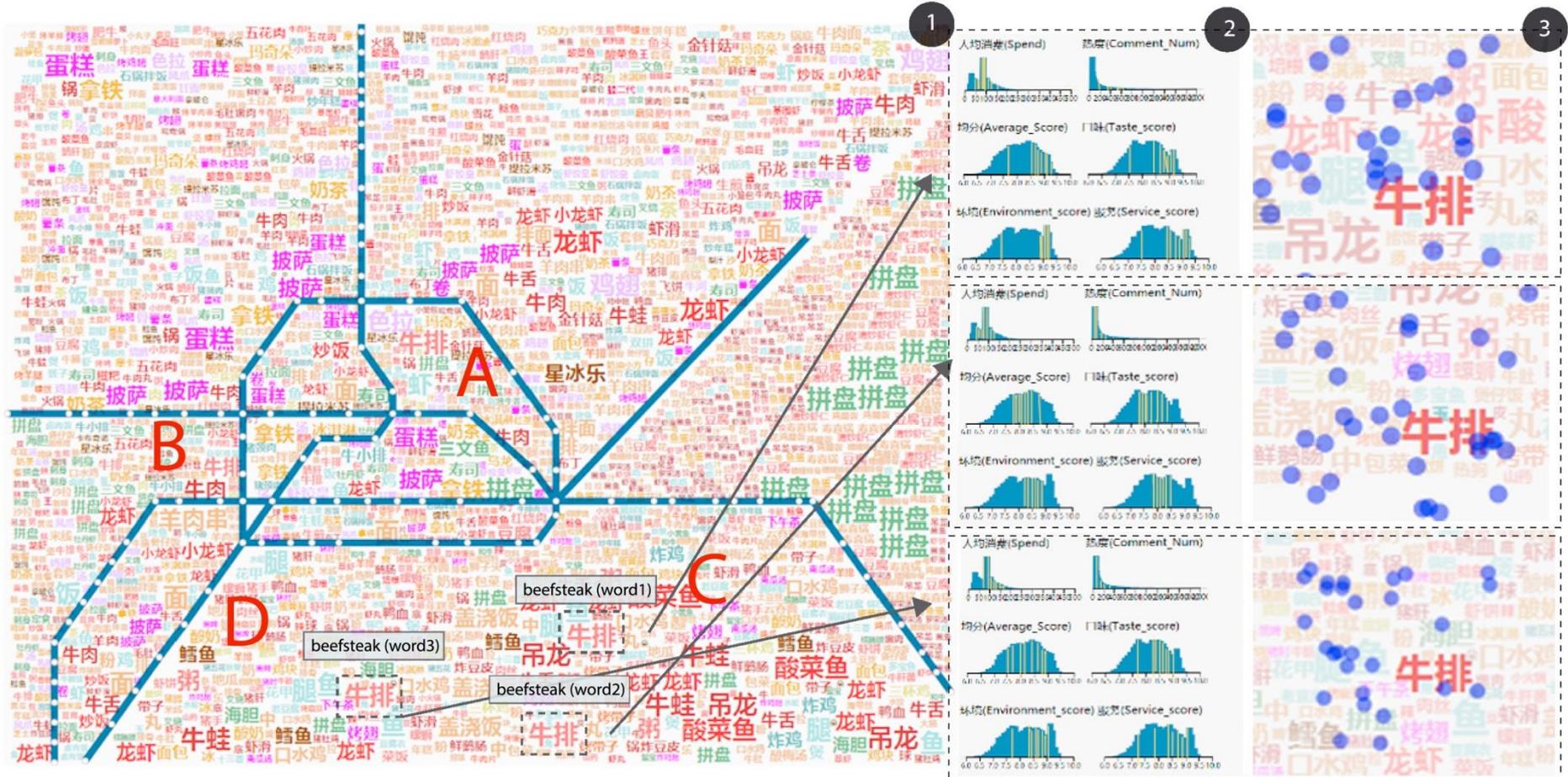
Data processing



Visualization



Estudo de Caso – Restaurantes em Shangai



Conclusão

- Análise semânticas e estatísticas foram utilizadas para processar o texto.
- MIP algorithm foi utilizado para desenhar o mapa do metrô automaticamente.
- A visualização é composta por: geographic view, metro-wordle view, filters e detail information view.
- O usuário pode interagir para maior exploração.
- O método pode ser aplicado para vários POI's e outras línguas.

Análise Crítica

- Na minha opinião o trabalho é extremamente interessante e, se caso disponibilizado, é uma boa alternativa para recuperação de pontos de interesse.
- A estética final é atrativa.
- O método acaba sendo aplicado somente para cidades grandes que possuem linhas de metrô e muitos pontos de interesse; como poderia ser feito para cidades menores?
- Em muitos momentos, achei o texto confuso. Por exemplo, a explicação de que o que aparece no mapa são palavras-chave e que elas agrupam restaurantes que estão próximos, demorou a ser apresentado, deixando o conteúdo desconexo no início.

Perguntas

1. Qual o benefício de se utilizar “wordles” e não “tag clouds”?
2. A soft constraint S4, apresentada ao MIP, é nova em relação as outras. Explique o que ela significa e sua importância.
3. Para inserir as palavras no mapa distorcido, um outro algoritmo é utilizado. Qual é tal algoritmo?