



Escreva Suas Próprias Letras Orientadas a Dados

Utilize SameDiff para analisar os estilos líricos de dois músicos e invente um dueto para eles.

Resumo

SameDiff compara dois ou mais arquivos de texto e informa o quanto são similares ou diferentes. Ele te ajuda a ver as diferenças e similaridades nas palavras usadas em cada arquivo então você pode aprender sobre análise quantitativa de texto. Esta atividade prática ajuda aos participantes construírem seu alfabeto de dados pela comparação de letras de dois músicos e a criação de uma nova música que eles podem escrever juntos.

Objetivos de Aprendizagem

- Aumentar a habilidade de analisar dados de texto
- Entender que comparar duas coisas é uma forma poderosa de encontrar histórias em dados
- Conscientizar sobre o tipo de perguntas você pode/deve fazer aos dados de texto
- Entender que análises de algoritmos podem revelar informações interessantes sobre seus dados

Fazer a Atividade

Resolvendo um Problema

Analizar grandes textos manualmente é difícil. Uma maneira de entender um “corpus” de um texto é compará-lo com outro, ou compará-lo com suas outras partes. Cientistas da computação buscam alternativas para ajudar nisso, inventando várias receitas, ou “algoritmos”, que podem comparar dois corpora. SameDiff executa alguns destes algoritmos para que você possa experimentar comparar dois grandes pedaços de texto.

Compartilhe Exemplos Inspiradores

Enormes dados de textos estão à nossa volta. Hoje você pode baixar todos os e-mails da Secretária de Estado Americana Hilary Clinton, telegramas diplomáticos no Wikileaks, ou todas as histórias de Sherlock Holmes do projeto Gutenberg. Analisar e visualizar estes enormes textos é uma coisa comum atualmente, de maneira séria ou divertida. Mostre partes do “Color Signatures” de Jaz Parkinson (<http://jazparkinson.tumblr.com/>) que compara as cores mencionadas em diferentes livros, e “Rap Research Lab” (<http://rapresearchlab.com/>) de Tahir Hamphill.

Duração

30 minutos

Público

3 - 100 people. Ages 12+.

Nós projetamos esta atividade para aproximadamente 30 pessoas, mas você pode fazer com mais ou menos. Concebido para ensino básico, educação superior, organizações jornalísticas, organização sem fins lucrativos e oficinas em comunidades sociais. Não é necessária nenhuma experiência prévia.

Espaço

- Um projetor conectado a um computador
- Organização em grupos de 3 em torno de um computador
- Mesas ou chão com espaço para colocar papel ou adesivo para pregar papéis nas paredes, para que os participantes possam desenhar

Materiais

- Computadores
1 para cada 3 participantes
- Grandes pedaços de papel
aproximadamente 60 cm x 90 cm
- Lapis de cor/giz de cera grossos ou canetões

Fazer a Atividade (contínuo)

Introduza a Ferramenta

Abra o SameDiff (<https://databasic.io/samediff>) e escolha Beyoncé e Aretha Franklin nos exemplos. Na página de resultados explique que a coluna da esquerda mostra palavras únicas da Beyoncé, enquanto a coluna da direita mostra palavras únicas de Aretha Franklin. Estas são suas diferenças. A coluna do meio mostra as palavras que elas têm em comum. Chame a atenção para o topo da página de resultados onde diz: “Esses dois documentos são bem similares”. SameDiff usa um algoritmo chamado “similaridade de cosseno” para lhe dar uma pontuação de similaridade. Similaridade de cosseno funciona criando uma lista de palavras de Beyoncé e uma lista de palavras de Aretha Franklin. Ele conta a frequência com que cada termo aparece em cada documento e então compara o quanto estas duas listas estão próximas uma da outra. Este é um algoritmo útil para análise de texto.

Inicie a Atividade

1. Participantes têm 15 minutos.
2. Participantes trabalham em equipes de 3.
3. Cada equipe usa SameDiff para comparar as letras de dois músicos. Como a colaboração musical é muito comum, escolha dois artistas e imagine como seria uma música escrita por esses dois artistas.
<https://databasic.io/samediff>
4. Cada equipe escreve a letra de sua música em um grande pedaço de papel com lapiseira e canetas.
5. Equipes ganham pontos extras se: (a) sua letra tem rima e/ou (b) a letra tem uma música para ser cantada e/ou (c) eles apresentam em estilo karaoke.

Compartilhe

Dê 1 minuto para cada grupo compartilhar sua nova música. Algumas questões e temas para procurar e focar durante a discussão:

- Você notou temas em comum?
- As letras mais interessantes são resultantes de artistas que trabalham de maneira muito diferente?
- A comparação é um meio poderoso para encontrar histórias em dados
- Trabalhar com dados pode ser divertido

Lembretes

- Algoritmos são simplesmente um conjunto de passos que você (ou um computador) fazem para resolver um problema. Por exemplo, quando você perde suas chaves você executa um algoritmo para procurar por elas - primeiro você olha os bolsos, a fechadura da porta, etc.
- Uma similaridade de cosseno 1,0 significa que são exatamente iguais; 0,0 significa totalmente diferentes.

Glossário

Algoritmo

Um conjunto de etapas que você (ou um computador) fazem para resolver um problema. Por exemplo, quando você digitar um termo de pesquisa, o Google executa um algoritmo para descobrir quais páginas devem ser mostradas.

Corpus

Uma coleção de textos escritos. Por exemplo, todas as letras de músicas de Katy Perry. Seu plural é “corpora”.

Similaridade de Cosseno

A pontuação da Similaridade de Cosseno busca informar a similaridade entre dois documentos baseada no número de vezes que uma palavra é usada em cada um.