Extração e Tratamento de Texto

L3P – Laboratório de Políticas Públicas Participativas

Luis Felipe Rosa de Oliveira

Data:14/10/2016

NetVizz

- É uma ferramenta que extrai dados de diferentes seções do Facebook. (https://apps.facebook.com/netvizz/)
- Em particular grupos e páginas, para propósito de pesquisas.
- Os arquivos podem ser facilmente analisados em outros softwares.

NetVizz - Módulos

- **Dados de Grupo** – Cria redes e planilhas sobre a atividade dos usuários nos posts em grupos.
- **Dados de Página** – Cria redes e planilhas sobre a atividade dos usuários nas postagens em páginas.
- **Rede do Like de Páginas** – Cria rede das páginas conectadas através dos likes entre elas.
- **Pesquisa** – Função de Busca do Facebook
- **Status do Link** – Provê estatísticas dos links compartilhados no Facebook.

YouTube Data Tools

- É uma coleção de ferramentas para extração de dados do Youtube via API (https://tools.digitalmethods.net/netvizz/youtube/).

Youtube Data Tools - Módulos

- **Informação do Canal** – Este módulo retorna diferentes tipos de informações sobre um canal a partir de um ID de canal.

- **Rede de Canais** – Este módulo faz uma rede de canais conectados que estão em destaque (“Featured”) e por número de inscritos separados por canais retornados de uma busca ou especificados previamente.

- **Lista de Vídeos** – Retorna informações e estatísticas de uma lista de vídeos de uma das quatro fontes: vídeos de uma canal específico, uma playlist, vídeos retornados por uma busca, ou especificados em uma lista de IDs.

- **Rede de Vídeos** – Este módulo cria uma rede de relações entre vídeos via “related videos” do YouTube, iniciando de uma pesquisa ou uma lista de vídeos.

- **Informações do Vídeo e Comentários** – Este módulo retorna as informações básicas de um vídeo e provê os comentários e informações sobre eles.

Codificação do Corpus de Texto

- Variáveis e Conjuntos de Texto.
  - Variáveis são separadas por ****** *exemplo**
  - Conjunto de Texto são separados por ****** Exemplo**
- É importante limpar o texto
  - **Palavras incorretas**
  - **Palavras sem acento**
  - **Emoticons**
  - **Símbolos**

Codificação do Corpus de Texto