

# III ENCONTRO DE PESQUISA EM ARQUIVOLOGIA

17 e 18 de outubro de 2018

UFES - CCJE



## Sistema de Classificação Automática de Documentos

**Matheus de Araújo Nogueira – Vitória Virtual**

Elias de Oliveira – Universidade Federal do Espírito Santo

# Recuperação da informação

- ▶ A computação nos permitiu criar mais informações e compartilhar com qualquer pessoa.
- ▶ Assim gerando uma quantidade gigantesca de informação a todo momento.
- ▶ Como jornais, cartas, editais, revistas, relatórios técnico, etc.

# Recuperação da informação

- ▶ Sendo necessário organizar e classificar esses documentos, para futuras buscas e análises;
- ▶ Essa classificação é realizada manualmente;
- ▶ Para documentos que nascem em meio digital e segue para *softwares* de organização, isso não é um grande problema;
- ▶ Mas, e para acervos que não estão classificados?
- ▶ Ou, mudanças nos parâmetros da classificação?

# Exemplo de acervo

- ▶ O acervo do jornal A Tribuna disponibilizado online;
- ▶ Contém mais 300 mil páginas do jornal de 2003 a 2018;
- ▶ Está dividido em 12 classes(cadernos);

At2	Noticiário	Sobre Rodas	Jornal da Família
Tv Tudo	Mulher	Minha Casa	Informática
Especial	Imóveis	Classificados	Tudo a Ver

# Exemplo de acervo

- ▶ E desejo modificar as classes para:

Política	Economia	Segurança	Esporte
Cultura	Televisão	Teatro	Cinema
Informática	Música	Classificados	Educação
Saúde	Bem estar	Gastronomia	Eventos

# Exemplo de acervo – Tempo da tarefa

- ▶ 1 minuto para leitura e definição da classe;
- ▶ São necessárias 5000 horas de trabalho;
- ▶ 625 dias de trabalho(8 horas diárias).

**Como diminuir este  
tempo?**

# Solução

- ▶ Distribuindo a tarefa entre muitas pessoas;
- ▶ Ou, utilizando sistemas de classificação automática;

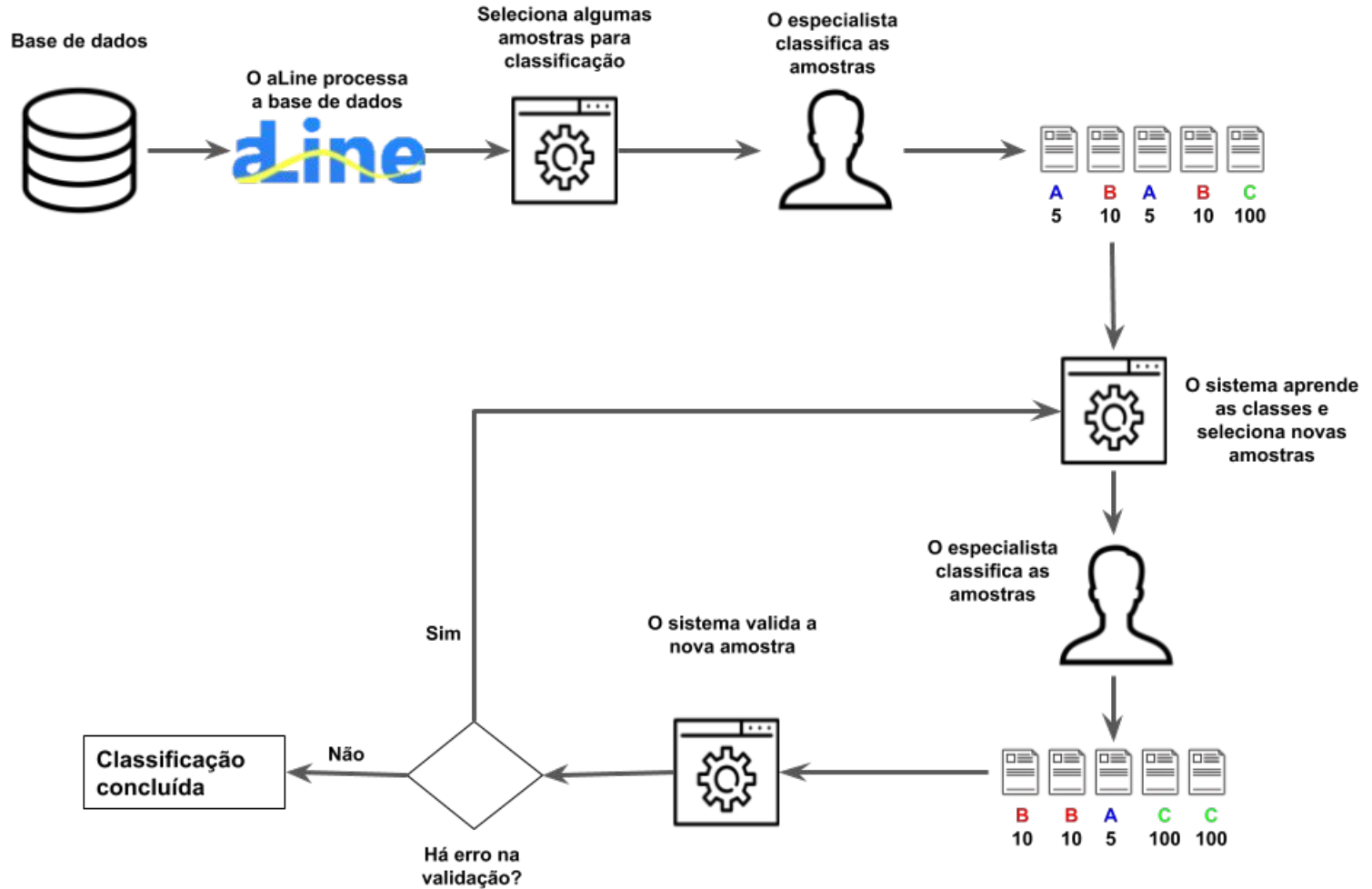




# aLine

- ▶ Extrai automaticamente textos e padrões dos documentos;
- ▶ Aprende com o arquivista, quais os padrões de classificação ele utiliza;
- ▶ Com base em poucas amostras classificadas;

# aLine



# Classificação da A Tribuna

- ▶ Foram classificadas 1000 páginas para amostra;
- ▶ Aproximadamente 30 horas de anotação;
- ▶ 7 dias de processamento de máquina;

Política	Economia	Segurança	Esporte
Cultura	Televisão	Teatro	Cinema
Informática	Música	Classificados	Educação
Saúde	Bem estar	Gastronomia	Eventos

# III ENCONTRO DE PESQUISA EM ARQUIVOLOGIA

17 e 18 de outubro de 2018

UFES - CCJE



## Obrigado!

Matheus de Araújo Nogueira

Sócio Fundador

[manogueira@vitoriavirtual.com.br](mailto:manogueira@vitoriavirtual.com.br)

[www.vitoriavirtual.com.br](http://www.vitoriavirtual.com.br)



Vitória Virtual  
SOLUÇÕES EM DATA MINING



UFES