

TÓPICOS AVANÇADOS EM INTELIGÊNCIA ARTIFICIAL

BACHARELADO EM SISTEMAS DE INFORMAÇÃO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
UNIDADE ACADÊMICA DE SERRA TALHADA

PROFESSOR: DOUGLAS VITÓRIO (douglas.alisson17@gmail.com)

MINERAÇÃO DE OPINIÃO

AULA 02



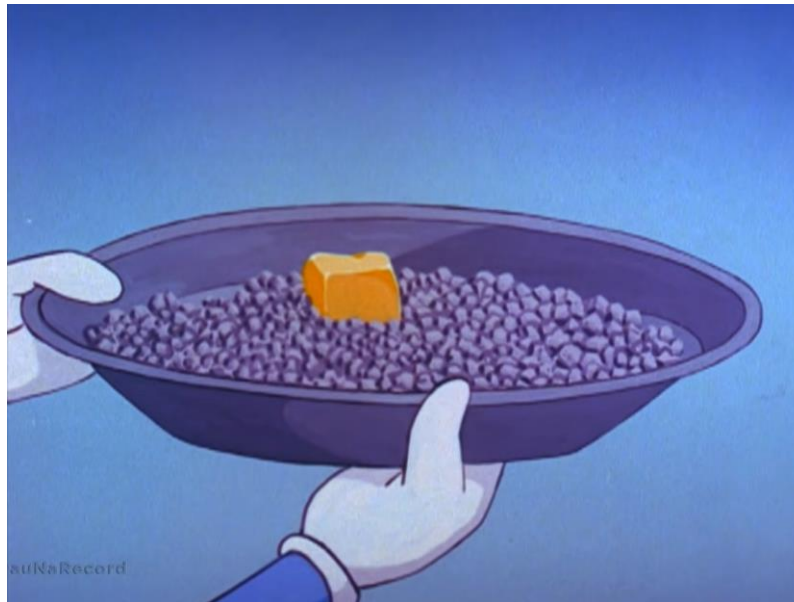
MINERAÇÃO

O QUE É MINERAR OU GARIMPAR?



MINERAÇÃO

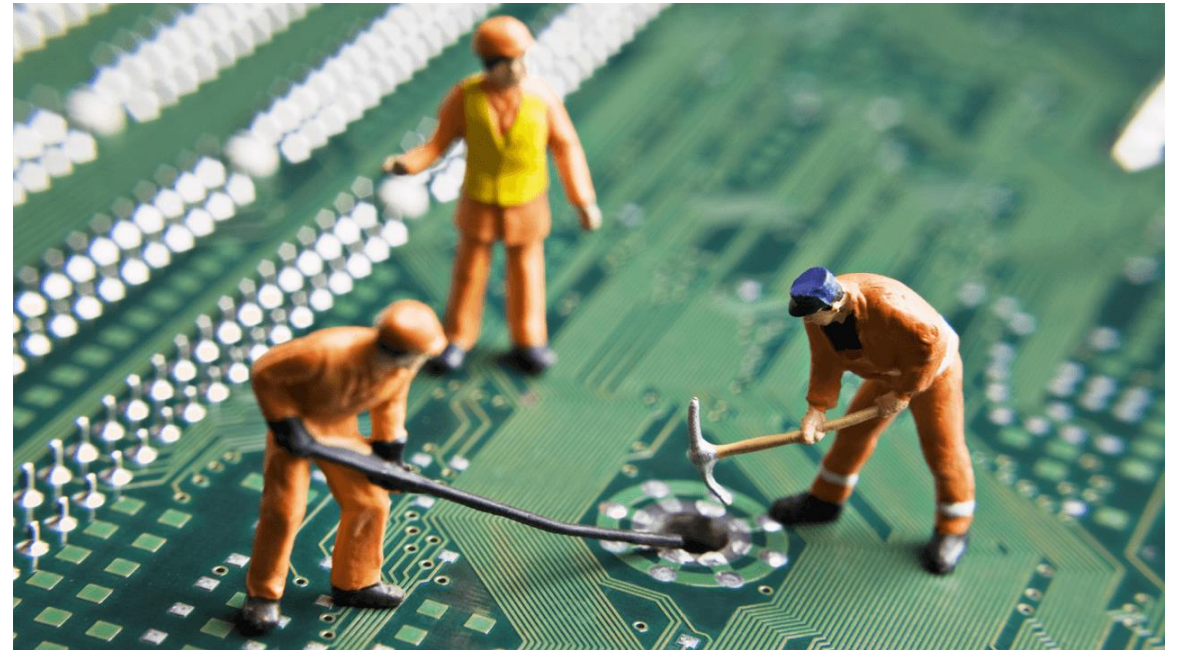
- Ato ou efeito de minerar;
- Ação de explorar minas e retirar minério;
- Ação de depurar (limpar) o minério que foi retirado das minas.



MINERAÇÃO

Da mesma forma, podemos minerar:

- Dados
- Texto
- Opiniões...



MINERAÇÃO DE DADOS

Mineração de Dados (*Data Mining*):

- O processo **automatizado** de descoberta de novas informações a partir de **grandes massas de dados**
- É uma área extensa e interdisciplinar



MINERAÇÃO DE DADOS



EXEMPLOS

Descoberta de **padrões** de comportamento de clientes:

- Uma das maiores redes de varejo estadunidenses descobriu, em seu gigantesco armazém de dados, que a venda de fraldas descartáveis estava associada à de cerveja;
- Pessoas que iam comprar fraldas aproveitavam a ida ao mercado e levavam algumas latinhas para casa;
- Então as cervejas passaram a ser colocadas do lado das fraldas descartáveis.

EXEMPLOS

Identificação de estilos de ações fraudulentas em cartões de crédito e seguradoras:

- Detecção de desvios a partir do comportamento de compra prévio de cada cliente;
- Buscava-se por situações de compra cujos dados divergissem das características do padrão de compra de cada cliente.

MINERAÇÃO DE TEXTO



MINERAÇÃO DE TEXTO

Mineração de Texto (*Text Mining*):

- O termo “mineração de texto” foi mencionado pela primeira vez em 1995, por Feldman e Dagan, como “**análise de texto apoiada por máquina**”.

Outros termos:

- *Intelligent Text Analysis*
- *Text Data Mining*
- *Knowledge-Discovery in Text (KDT)*

POR QUE MINERAR TEXTO?

- De acordo com estudos, 85-90% dos dados corporativos estão armazenados em formato não estruturado, isto é, em texto.
- Dados armazenados e analisados por grandes empresas vão mudar de petabytes para exabytes em num futuro próximo, sendo texto um dos formatos de armazenamento mais comuns.
- O grande crescimento das mídias sociais e do **Conteúdo Gerado pelo Usuário (CGU)** na Internet também fornecem uma grande quantidade de informação que permite descobrir experiências, opiniões e sentimentos de usuários (clientes, fãs, eleitores, alunos...).

DO QUE SE TRATA ESSE TEXTO?

Mineração de texto, conhecida também como mineração de dados textuais e semelhante à análise textual, refere-se ao processo de obtenção de informações importantes de um texto. Informações importantes são obtidas normalmente pela elaboração de padrões e tendências através de meios como o padrão estatístico de aprendizagem. Geralmente a mineração de texto envolve o processo de estruturação do texto de entrada (frequentemente análise, junto com a adição de algumas características linguísticas derivadas e com a retirada de outras, e com a subsequente inserção em um banco de dados), de derivação de padrões dentro da estrutura de dados e, por fim, de avaliação e interpretação do resultado. Geralmente, “importante” em mineração de texto refere-se a algumas combinações de relevância, originalidade e interesse. Tarefas típicas de mineração de texto incluem categorização e agrupamento de texto, extração de conceito/entidade, produção de taxonomias granulares, análise de sentimentos [1][2], resumo de documentos e modelagem de relações entre entidades (ex., aprender relações entre entidades nomeadas).

A análise de texto envolve informações de recuperação, análise lexical a fim de estudar a frequência de distribuição de palavras, reconhecimento de padrões, identificação/anotação, extração de informações, técnicas de mineração de dados que incluem link e associação de análises, visualização e analítica preditiva. O objetivo maior é transformar o texto em dados para análise, por meio da aplicação do processamento de linguagem natural (PLN) e de métodos analíticos.

Uma aplicação comum é examinar um conjunto de documentos escritos em uma linguagem natural e, ou modelar o conjunto de documentos para fins de classificação preditiva ou preencher um banco de dados ou índice de pesquisa com as informações extraídas.

DO QUE SE TRATA ESSE TEXTO?

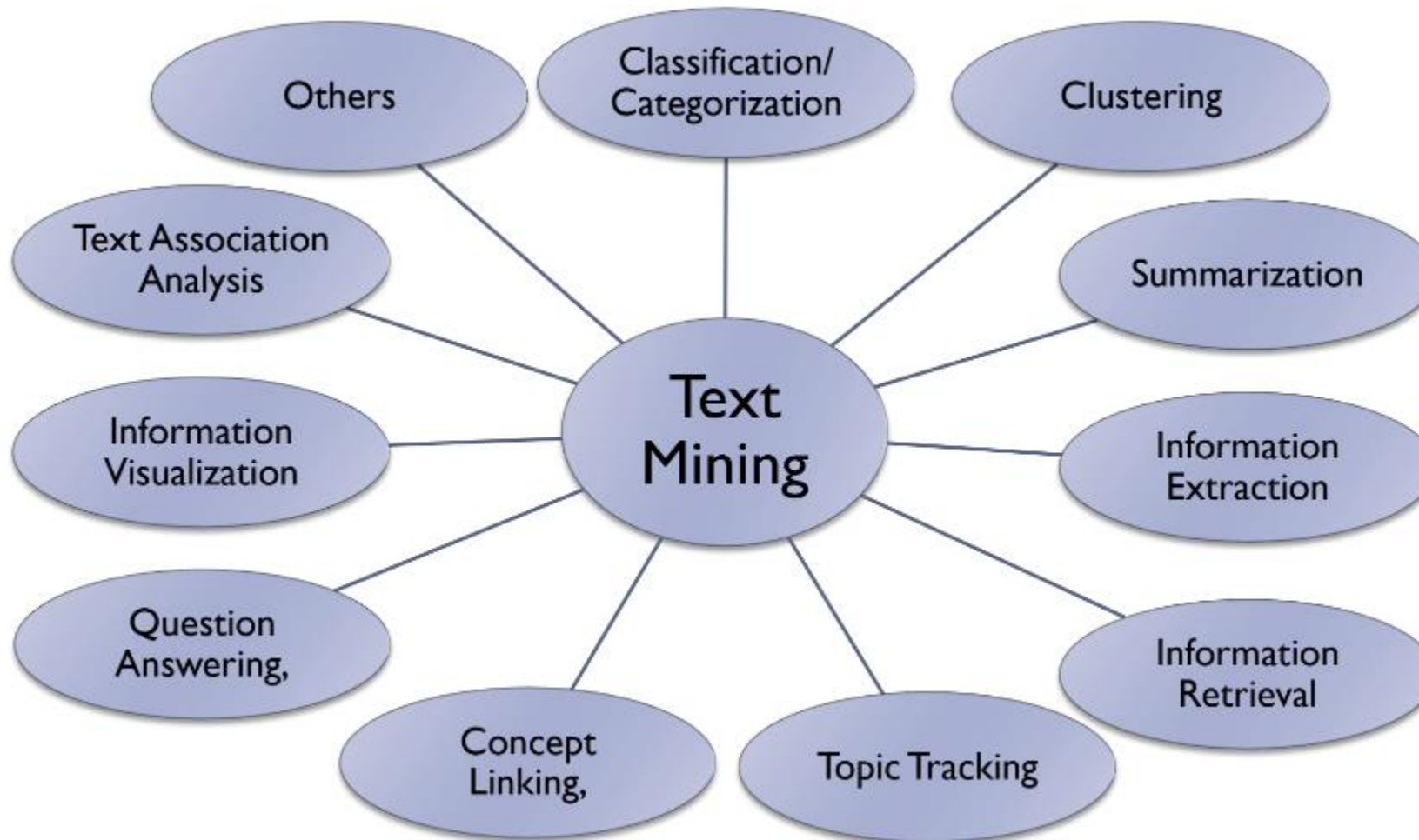
NUVEM DE PALAVRAS

A **nuvem de palavras** é um recurso gráfico, usado principalmente na Internet, para descrever os termos mais frequentes de um determinado texto.

O tamanho da fonte em que a palavra é apresentada é em função da frequência da palavra no texto.

<https://www.jasondavies.com/wordcloud/>

TAREFAS DE MINERAÇÃO DE TEXTO



TAREFAS DE MINERAÇÃO DE TEXTO

Classificação ou categorização	Identifica os principais temas de um texto e o classifica em um conjunto pré-definido de classes ou categorias. Aprendizado supervisionado.
Agrupamento	Utilizado para agrupar textos similares. Difere da classificação de texto no sentido que os textos são agrupados “on the fly”. Não se conhece previamente as classes ou categorias. Aprendizado não supervisionado.
Sumarização	Cria uma versão condensada de um texto ou “corpus” contendo os principais tópicos.
Extração de Informação	Identifica frases “chave” e relacionamentos dentro do texto procurando sequências predefinidas no texto através de correspondência de padrões.
Recuperação de Informação	“Pistas” são dadas sobre os documentos que deseja recuperar a partir da coleção e, em seguida, documentos correspondentes as pistas são apresentados como respostas para a consulta.
Rastreamento de Tópicos, Vinculação de Conceitos, Perguntas Respondidas, Visualização de Informações, Análise de Associações de Texto e etc.	

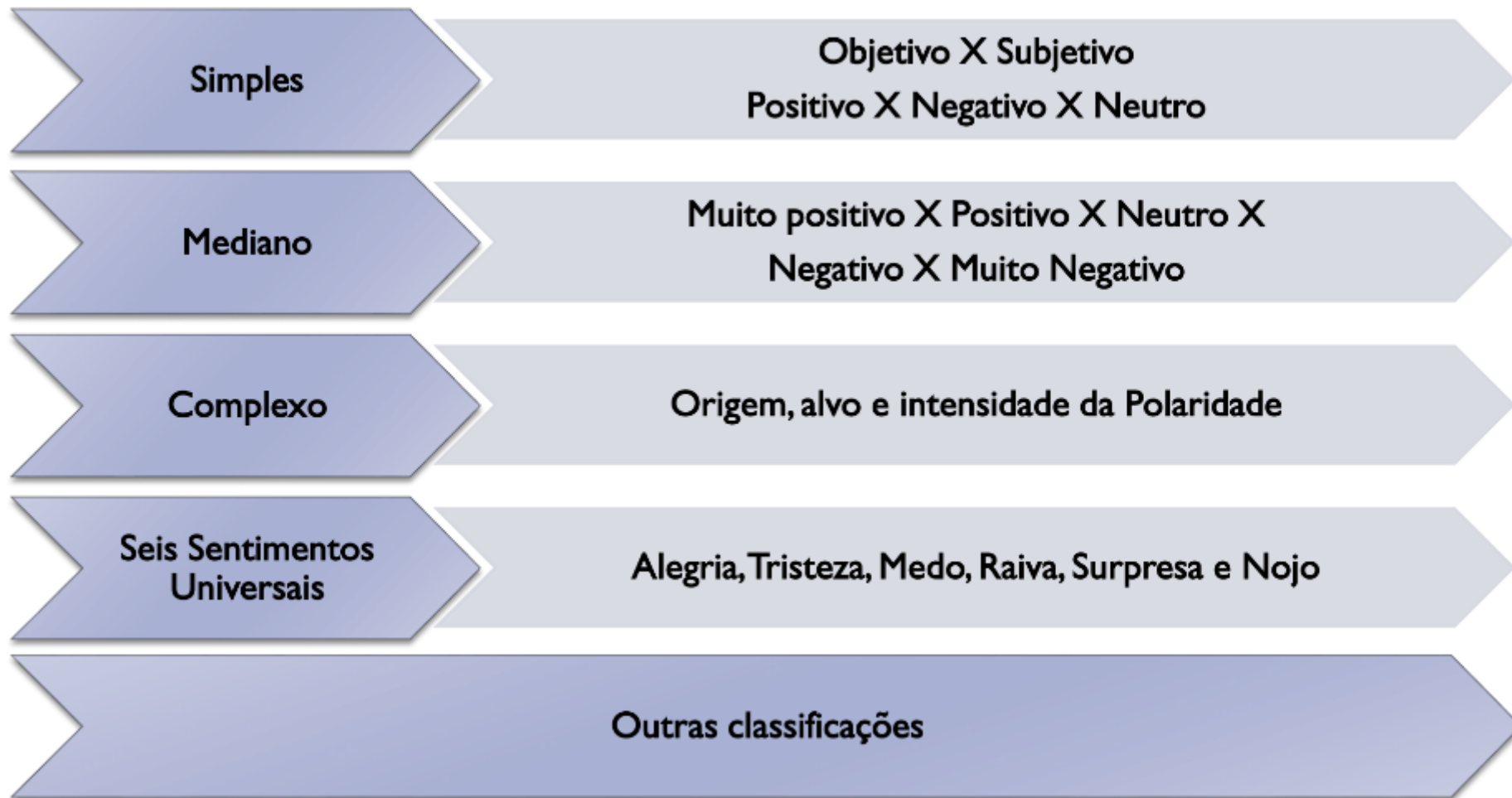
MINERAÇÃO DE OPINIÃO

Mineração de Opinião (*Opinion Mining*):

- Campo de estudo que analisa o **sentimento**, a **opinião**, as atitudes, as avaliações, as emoções de pessoas acerca de entidades (empresas, pessoas, partidos políticos...), expressos em texto.
- Também chamada de **Análise de Sentimento**, Extração de Opinião, Mineração de Sentimento...



CLASSIFICAÇÃO DE OPINIÃO



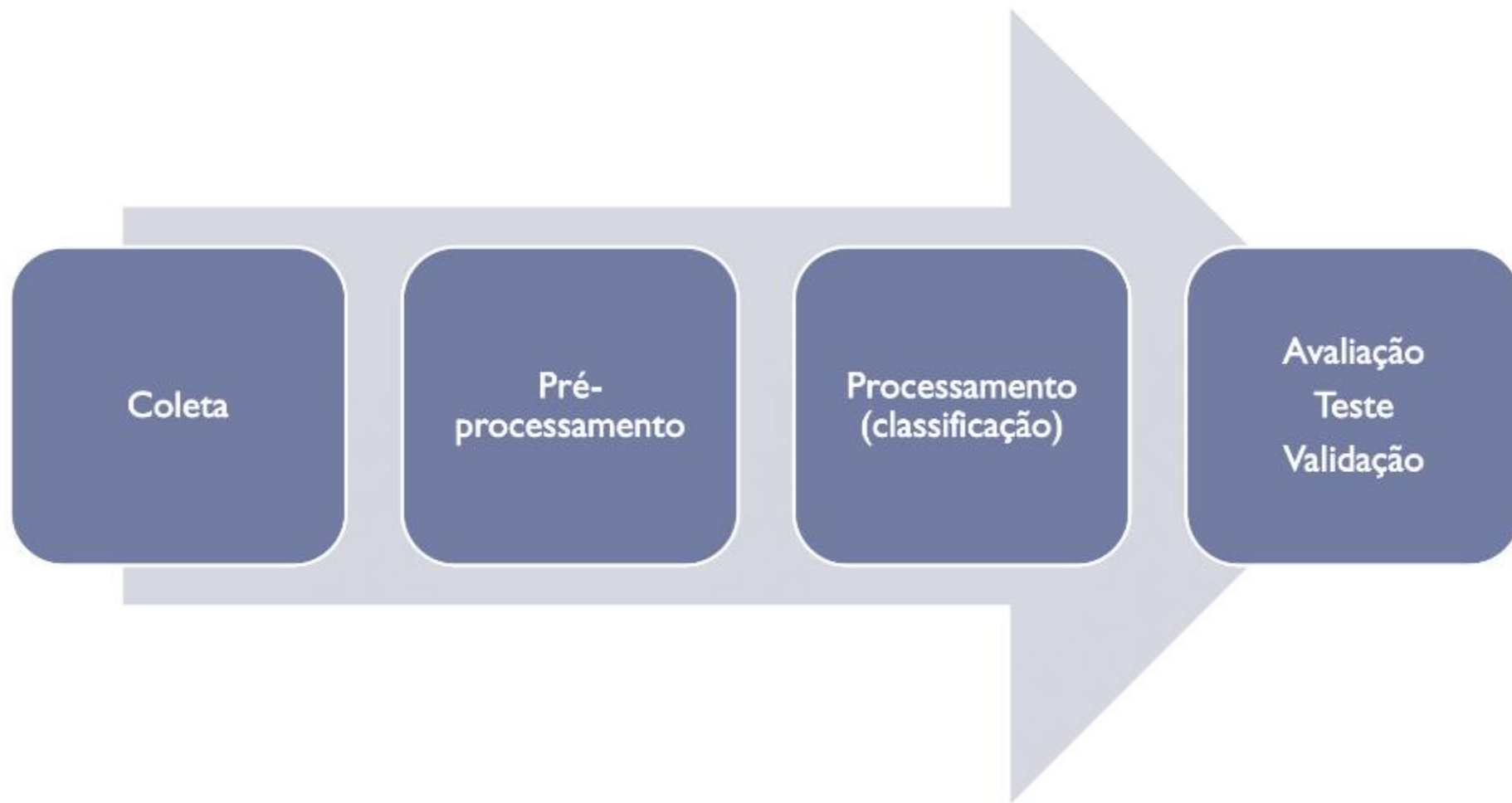
OPINIÕES

@EikeBatiiista Como podemos ajudar o senhor? ;-)

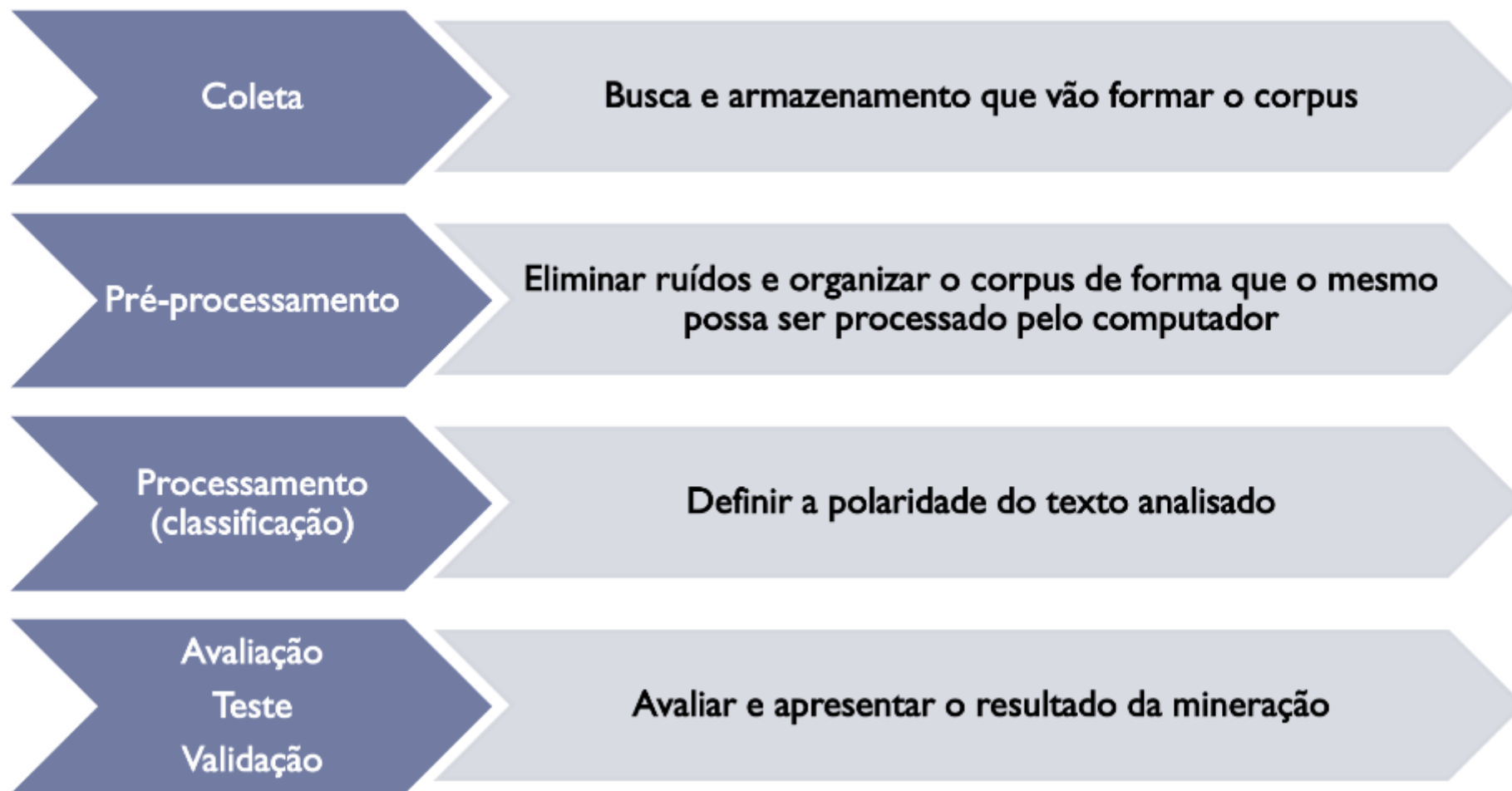
Alô, @BancodoBrasil! O app de vocês para Android tá muito louco! Entrei na minha conta e vi o extrato de OUTRO CLIENTE! O.o

que deus te de em dobro, toda a velocidade de internet q vc me deu
@vivo

PROCESSO DE MINERAÇÃO DE OPINIÃO



PROCESSO DE MINERAÇÃO DE OPINIÃO



EXEMPLO DE APLICAÇÕES

- Análise sobre produtos e serviços
- Popularidade de personalidades
- Repercussão sobre ações políticas e populares
- Opiniões sobre acontecimentos midiáticos...