

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO**

**DESENVOLVIMENTO E IMPLEMENTAÇÃO  
DE UM SISTEMA DE RECONHECIMENTO  
DE ESPECTROS DE RAIOS X  
CARACTERÍSTICOS DE MINERAIS**

**TRABALHO DE GRADUAÇÃO**

**Thiago Lopes Trugillo da Silveira**

**Santa Maria, RS, Brasil**

**2015**

**DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM  
SISTEMA DE RECONHECIMENTO DE ESPECTROS DE  
RAIOS X CARACTERÍSTICOS DE MINERAIS**

**Thiago Lopes Trugillo da Silveira**

Trabalho de Graduação apresentado ao Bacharelado em Sistemas de  
Informação da Universidade Federal de Santa Maria (UFSM, RS), como  
requisito parcial para a obtenção do grau de  
**Bacharel em Sistemas de Informação**

**Orientador: Prof. Dr. José Antônio Trindade Borges da Costa**

**Santa Maria, RS, Brasil**

**2015**

**Universidade Federal de Santa Maria  
Centro de Tecnologia  
Bacharelado em Sistemas de Informação**

A Comissão Examinadora, abaixo assinada,  
aprova o Trabalho de Graduação

**DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM SISTEMA DE  
RECONHECIMENTO DE ESPECTROS DE RAIOS X  
CARACTERÍSTICOS DE MINERAIS**

elaborado por  
**Thiago Lopes Trugillo da Silveira**

como requisito parcial para obtenção do grau de  
**Bacharel em Sistemas de Informação**

**COMISSÃO EXAMINADORA:**

**José Antônio Trindade Borges da Costa, Dr.**  
(Presidente/Orientador)

**Marcos Cordeiro d'Ornellas, Dr. (UFSM)**

**Ana Trindade Winck, Dra. (UFSM)**

Santa Maria, 14 de Dezembro de 2015.

*Aqueles que sempre me apoiaram...*

## AGRADECIMENTOS

Agradeço a todos os meus professores que, durante o Curso, muito ajudaram em minha formação pessoal e profissional. Em especial, agradeço ao meu orientador e amigo, Prof. Dr. José Antônio Trindade Borges da Costa, por seus ensinamentos antes e durante o desenvolvimento deste trabalho. Agradeço também à Profa. Dra. Ana Trindade Winck e ao Prof. Dr. Marcos Cordeiro d'Ornellas por terem aceito participar da avaliação deste trabalho e, sobretudo, por suas excelentes colocações e contribuições para com o mesmo.

Um especial agradecimento aos Professores Gustavo Castellano e Jorge Trincavelli da Facultad de Matemática, Astronomía y Física de la Universidad Nacional de Córdoba pela aquisição dos espectros utilizados neste trabalho.

Agradeço também a todos de minha família e, em especial, minha mãe, Ilse Trindade Lopes, e meu pai, Andre Trugillo da Silveira por estarem sempre presentes. Agradeço também a minha indefectível namorada, Samara Dias Osorio, por todo o amor, carinho e apoio que ela sempre me deu.

Por fim, agradeço a todos aqueles que não foram diretamente citados aqui mas que contribuíram com este trabalho e me ajudaram durante todo este tempo.

*“There is no finish line. So love the journey”*  
— DAVID WEEKLY

## RESUMO

Trabalho de Graduação  
Bacharelado em Sistemas de Informação  
Universidade Federal de Santa Maria

### DESENVOLVIMENTO E IMPLEMENTAÇÃO DE UM SISTEMA DE RECONHECIMENTO DE ESPECTROS DE RAIOS X CARACTERÍSTICOS DE MINERAIS

AUTOR: THIAGO LOPES TRUGILLO DA SILVEIRA

ORIENTADOR: JOSÉ ANTÔNIO TRINDADE BORGES DA COSTA

Local da Defesa e Data: Santa Maria, 14 de Dezembro de 2015.

Minerais são definidos como ocorrências naturais de elementos ou compostos químicos formados como o produto de uma sucessão de processos inorgânicos. A correta identificação de minerais permite o avanço de diversas áreas da ciência e da indústria. Técnicas de espectroscopia, em especial por dispersão de raios X, são particularmente úteis quando há a necessidade de se analisar *in situ*, não destrutivamente, grãos minerais de escala micrométrica. Uma vez que já foram identificados mais de 5000 minerais na crosta terrestre, é interessante investigar metodologias semi-automáticas que possam auxiliar geólogos e mineralogistas na tarefa de identificação. Assim, este trabalho apresenta a versão inicial de um sistema de apoio à decisão para reconhecimento de minerais através de seus espectros de raios X característicos. Nesse sistema, foram testadas diferentes maneiras de pré-processar os espectros, visando extrair características minimamente variantes às configurações de aquisição, e distintas métricas de similaridade (booleanas e não-booleanas), que ponderam diferentemente as características espectrais. Para avaliar o desempenho do sistema, dois cenários foram considerados, ambos usando espectros de amostras de minerais certificados como *gold standard*. No primeiro cenário, os testes foram realizados com outro conjunto de dados certificados e, no segundo, com dados pós-processados, sujeitos a distorções, visando identificar o potencial do *framework* desenvolvido. Os resultados obtidos no primeiro cenário indicam que as métricas não-booleanas tendem a ser mais acuradas do que as booleanas. Já no segundo cenário, as conclusões são contraditórias, indicando a necessidade de dar continuidade a este estudo em busca de uma combinação de ambas as métricas, a fim de conferir mais robustez ao sistema.

**Palavras-chave:** Espectroscopia. Identificação de minerais. Reconhecimento de padrões.

## **ABSTRACT**

Undergraduate Final Work  
Undergraduate Program in Information Systems  
Federal University of Santa Maria

### **DEVELOPMENT AND IMPLEMENTATION OF A SYSTEM FOR RECOGNITION OF CHARACTERISTIC X-RAY SPECTRA OF MINERALS**

**AUTHOR: THIAGO LOPES TRUGILLO DA SILVEIRA**

**ADVISOR: JOSÉ ANTÔNIO TRINDADE BORGES DA COSTA**

**Defense Place and Date: Santa Maria, December 14<sup>th</sup>, 2015.**

Minerals are defined as natural occurrences of elements or chemical compounds formed as a product of several inorganic processes. The correct identification of minerals enables the advancement of several areas of science and industry. Spectroscopy techniques, specially X-ray dispersive analysis, are particularly useful when non-destructive in situ analysis of micrometer sized grains is needed. Since more than 5000 minerals have been identified in Earth's crust, it is interesting to investigate novel semi-automatic methods to help geologists and mineralogists to accomplish that task. Thus, in the present work, the initial version of a decision support system for mineral recognition through its characteristic X-ray spectrum is introduced. Different ways to pre-process the X-ray spectra, aiming to extract characteristics that are minimally variable under acquisition configurations, and different similarity metrics (boolean and non-boolean) weighting differently the spectral characteristics, were tested. In order to evaluate the system performance, two scenarios were considered, both using spectra of certified samples of minerals as gold standard. In the first scenario another certified data set was used for testing, whilst, in the other, post-processed data, which may be distorted, were used aiming to assess the potential of the developed framework. Results from the first scenario indicate that non-boolean similarity metrics tend to be more accurate than the boolean ones. In second scenario, the conclusions are contradictory, pointing out the need of continuing this study, searching for a combination of both kinds of metrics, in order to impart robustness to the system.

**Keywords:** Spectroscopy, Mineral identification, Pattern recognition.



## LISTA DE FIGURAS

Figura 2.1 – Espectro de emissão visível do hidrogênio atômico. Adaptada de TUTOR-VISTA (2015). . . . .	17
Figura 2.2 – Espectros de emissão visíveis de diferentes elementos químicos. Adaptada de DIMAURO (2015). . . . .	18
Figura 2.3 – Espectro de emissão de uma lâmpada fluorescente compacta representado na forma de um gráfico de intensidade versus comprimento de onda. Adaptada de CORNER (2008) . . . . .	19
Figura 2.4 – Conversão de um sinal analógico em um registro digital. Adaptada de PLATFORM (2015). . . . .	19
Figura 2.5 – Foto de uma microsonda eletrônica da marca Jeol e sua representação esquemática. Adaptadas de MCSWIGGEN (2005). . . . .	21
Figura 2.6 – Espectro EDS do mineral Anortita ( $\text{CaAl}_2\text{Si}_2\text{O}_8$ ). . . . .	22
Figura 2.7 – Dois espectros do mineral Anortita obtidos sob diferentes resoluções. . . . .	23
Figura 2.8 – Fluxograma da metodologia proposta. . . . .	31
Figura 3.1 – Espectros do mineral Cromita das bases de dados (a) SEM e (b) EPMA. . . . .	37
Figura 3.2 – Espectros dos minerais (a) Anortita e (b) Zoisita da base de dados SEM e (c) Anortita da base de dados Be <i>window</i> . . . . .	40

## LISTA DE TABELAS

Tabela 2.1 – Tabela de energias de emissão de raios X com $K_{\alpha 1}$ em [0,~10]KeV (EDAX, 2015). . . . .	25
Tabela 2.2 – Métricas de similaridade booleanas. Adaptação de JOHNSON; WICHERN (2007). . . . .	27
Tabela 3.1 – Base de dados SEM. . . . .	34
Tabela 3.2 – Base de dados EPMA. . . . .	35
Tabela 3.3 – Comparação de métricas de similaridade booleanas para os Cenários I e II. . .	36
Tabela 3.4 – Comparação de métricas de similaridade não-booleanas para os Cenários I e II. . . . .	36
Tabela 3.5 – Bases de dados em MINERAL SPECTRA (2014). . . . .	38
Tabela 3.6 – Comparação de métricas de similaridade booleanas para os Cenários III e IV. .	39
Tabela 3.7 – Comparação de métricas de similaridade não-booleanas para os Cenários III e IV. . . . .	39

## LISTA DE ABREVIATURAS E SIGLAS

AD	Analógico-digital
EDS	<i>Energy dispersive spectrometer</i>
EPMA	<i>Electron probe micro-analyzer</i>
EMSA	Padrão de armazenamento de dados espectrais
MLA	Mineral Liberation Analyzer
QEMSCAM	Quantitative Evaluation of Minerals by Scanning Electron Microscopy
SEM	<i>Scanning electron microscopy</i>
WDS	<i>Wavelength dispersive spectrometer</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
<b>1.1 Contextualização e motivação</b> .....	13
<b>1.2 Objetivos</b> .....	15
1.2.1 Objetivo geral .....	15
1.2.2 Objetivos específicos .....	16
<b>1.3 Estrutura do texto</b> .....	16
<b>2 MATERIAIS E MÉTODOS</b> .....	17
<b>2.1 Espectroscopia e identificação de minerais</b> .....	17
<b>2.2 Pré-processamento</b> .....	22
2.2.1 Detecção de linhas espectrais.....	22
2.2.2 Determinação das intensidades relativas .....	23
<b>2.3 Métricas de similaridade</b> .....	26
2.3.1 Métricas de similaridade booleanas .....	26
2.3.2 Métricas de similaridade não-booleanas.....	28
<b>2.4 Visão geral da metodologia proposta</b> .....	30
<b>3 RESULTADOS E DISCUSSÕES</b> .....	32
<b>3.1 Métricas de avaliação</b> .....	32
<b>3.2 Experimentos com dados certificados</b> .....	33
<b>3.3 Experimentos com dados pós-processados</b> .....	38
<b>4 CONCLUSÕES</b> .....	41
<b>REFERÊNCIAS</b> .....	42

# 1 INTRODUÇÃO

## 1.1 Contextualização e motivação

Minerais são definidos como ocorrências naturais de elementos ou compostos químicos formados como o produto de uma sucessão de processos inorgânicos (HURLBUT, 1963). A tarefa de identificação de minerais, apesar de não ser nova, continua a desenvolver-se pelo acréscimo de novas tecnologias de caracterização e de novos métodos de processamento da informação que elas fornecem. Nesse campo, os métodos de auxílio à decisão, baseada em reconhecimento de padrões, prometem abreviar a busca e recuperação de registros dos milhares de minerais conhecidos.

O reconhecimento de minerais pela análise de dados obtidos através de espectrômetros por dispersão de energia (EDS), classicamente executado por especialistas (AKKAŞ et al., 2015), está relacionado à identificação de linhas espectrais – emissões de radiação em faixas de energia características dos elementos químicos presentes – e a determinação de suas intensidades relativas. Tal tecnologia é especialmente importante quando há a necessidade de se analisar, *in loco*, pequenas estruturas que não podem ser trivialmente submetidas a outros testes como de cor, brilho, traço, clivagem, fratura, dureza, densidade, etc (LISBOA, 1982).

Os minerais presentes na crosta terrestre podem ser classificados em, pelo menos, oito grandes grupos: silicatos, óxidos, fosfatos, sulfetos, sulfatos, carbonatos, halogenetos e elementos nativos (HURLBUT, 1963). Essa divisão em classes diz respeito às diferentes propriedades químicas e físicas que os diversos minerais compartilham ou não (REED, 1996), as quais ditam o seu uso nas mais diversas áreas da ciência e indústria. AKKAŞ et al. (2015) argumentam que é extremamente importante conhecer o grupo de minerais ao qual pertence um espectro problema, antes de qualquer identificação mais específica. Tal classificação é determinada pela presença ou ausência de certos elementos químicos. Na análise de espectros EDS, isso pode ser feito pela presença das respectivas linhas espectrais, isto é, pela identificação da ocorrência de picos de intensidade dos raios X emitidos nas energias correspondentes às transições eletrônicas destes elementos.

A identificação, via programas de computador, de picos de intensidade dos elementos químicos presentes em um espectro possibilita o desenvolvimento de novas ferramentas de apoio à decisão para reconhecimento automático de minerais (NEWBURY, 2009). Em vista

disso, nas últimas décadas, para diferentes tipos de espectroscopia, métodos de reconhecimento de padrões têm sido propostos visando recuperar, de uma base de dados, o(s) espectro(s) que mais se aproxima(m) do espectro problema (*query*). Alguns deste métodos são proprietários, por vezes embarcados em equipamentos de aquisição, como *Quantitative Evaluation of Minerals by Scanning Electron Microscopy* (QEMSCAM) (FEI, 2015a) e *Mineral Liberation Analyzer* (MLA) (FEI, 2015b).

LAVINE; DAVIDSON; MOORES (2002) notam que o desempenho de uma ferramenta de classificação de padrões em problemas relacionados à análise espectral varia muito de acordo com a metodologia empregada, e que é uma tarefa difícil a escolha de quais são as características mais relevantes dos dados espectrais para o problema de classificação. Para atacar essas dificuldades, Lavine e colaboradores propõem um método baseado em algoritmos genéticos para seleção e refinamento de componentes principais dos dados espectrais proporcionando a identificação acurada de seis tipos de polímeros.

Os procedimentos preliminares para identificação de minerais através de seus espectros de raios X não diferem muito. O *software* QEMSCAN quebra o espectro em partes previamente conhecidas, relacionadas a cada elemento químico – as energias correspondentes às suas transições eletrônicas. A adoção de tal processo, chamado de decomposição elementar, geralmente propicia melhores resultados do que se a análise fosse feita sobre o espectro por inteiro. Segundo OWEN (2014), QEMSCAN funciona bem, embora este não tenha sido projetado para trabalhar com diversas resoluções de aquisição. Além disso, tal *software* apresenta para o usuário as informações de casamento ou não-casamento dos espectros, sem apresentar um "grau de certeza" de sua resposta. Por outro lado, o MLA avalia as probabilidades de um espectro *query* corresponder com algum daqueles de minerais certificados da base de dados. Entretanto, o sistema tem a tendência de ponderar mais fortemente o pico de maior intensidade, o que pode ser uma característica inapropriada quando há inversão nas intensidades relativas em duas amostras de um mesmo mineral.

Para trabalhar com diferentes níveis de resolução, OWEN (2014) propõe um método embarcado que utiliza medidas de valor médio e desvio padrão para cada elemento dentro da faixa de energia (ou comprimento de onda) dos espectros analisados. Obviamente, para obtenção de tais indicadores, há a necessidade de se ter várias amostras de um mesmo mineral. Medidas de probabilidade de casamento são então calculadas para cada elemento presente nos minerais da base de dados e, por fim, são apresentadas as probabilidades totais para cada um dos minerais

que mais se aproximam do *query*. Ao final, o autor observa que a redução das diferenças entre os espectros a um único número, através do cômputo de um métrica de similaridade, produziu bons resultados.

Da mesma forma que na proposta de OWEN (2014), abordagens que utilizam algoritmos de aprendizagem de máquina, como as propostas por BENTZ et al. (1999); DING; COLPAN (2006); AKKAŞ et al. (2015), necessitam de várias amostras referentes a um mesmo mineral para identificação dos espectros *query*. Em contrapartida, SMITH; LEIBOVITZ (1986) propõem um sistema onde o usuário pode informar variações que ele mesmo considera aceitáveis nas concentrações de cada elemento do espectro analisado. A conversão de intensidades para concentrações é feita pelo sistema e, então, este apresenta ao usuário aqueles minerais que melhor se aproximam das informações dadas.

No presente trabalho, da mesma forma que em OWEN (2014), os espectros de emissão de raios X característicos produzidos em microsondas eletrônicas (SEM) ou em microscópios eletrônicos dotados de detectores de raios X (REED, 1996), são diretamente comparados, sem conversão para valores de concentração. O problema central consiste em encontrar tanto descritores robustos dos espectros – que permitam a sua identificação mesmo quando adquiridos em diferentes resoluções e de amostras que apresentem variações de composição – como medidas de similaridade que atribuam a importância adequada a cada uma destas características. De maneira alguma é pretensão do autor considerar todos os minerais existentes na crosta terrestre. Trabalhando com um conjunto reduzido de dados, a presente proposta foi desenvolver um sistema que permitisse avaliar características espectrais e medidas de similaridade, de modo a lançar as bases para futuras pesquisas relacionadas ao tema.

## **1.2 Objetivos**

### **1.2.1 Objetivo geral**

O objetivo geral deste estudo é desenvolver e implementar um sistema de reconhecimento de espectros de raios X característicos de minerais que possibilite incorporar e testar tanto características espectrais como medidas de similaridade que auxiliem na identificação de minerais.

### 1.2.2 Objetivos específicos

De forma a atingir o objetivo geral deste trabalho, destacam-se os seguintes objetivos específicos:

- identificação de métodos clássicos e no estado da arte para reconhecimento dos espectros de raios X característicos de minerais;
- identificação de bancos de dados e *benchmarks* públicos para etapas de teste e validação;
- estudo e implementação de métodos para extração de características;
- estudo e implementação de algoritmos para determinação de similaridade entre dados;
- desenvolvimento de um sistema para reconhecimento dos espectros de raios X característicos de minerais;

### 1.3 Estrutura do texto

O restante deste trabalho está organizado como segue. O Capítulo 2 apresenta os conceitos básicos de espectroscopia e seu papel na identificação de minerais. O mesmo capítulo do texto apresenta também as técnicas de pré-processamento e as métricas de similaridade aqui estudadas. Os resultados e suas respectivas discussões são exibidos, em dois contextos distintos, no Capítulo 3. Por fim, as conclusões desse trabalho bem como estudos futuros são apresentados no Capítulo 4.



## 2 MATERIAIS E MÉTODOS

Este Capítulo apresenta os conceitos necessários para o entendimento da metodologia proposta neste trabalho. Primeiramente, na Seção 2.1, é apresentado o conceito de espectroscopia e seu papel na identificação de minerais. A Seção 2.2 apresenta duas abordagens de pré-processamento de espectros as quais são cruciais para posterior etapa de reconhecimento de minerais. Métricas de similaridade, adotadas neste trabalho como parte do sistema de apoio à decisão desenvolvido, são apresentadas na Seção 2.3. Finalmente, na Seção 2.4 a metodologia implementada é exposta na forma de fluxograma.

### 2.1 Espectroscopia e identificação de minerais

Espectroscopia é o nome dado às técnicas de caracterização da constituição da matéria por meio da análise da radiação eletromagnética emitida, absorvida, transmitida, refletida ou difratada. Um espectro é um conjunto de dados formado pelas medidas de intensidade da radiação correspondente a cada comprimento de onda, frequência ou energia dos fótons. Na Figura 2.1 está ilustrado o processo de aquisição do espectro de emissão visível do hidrogênio atômico.

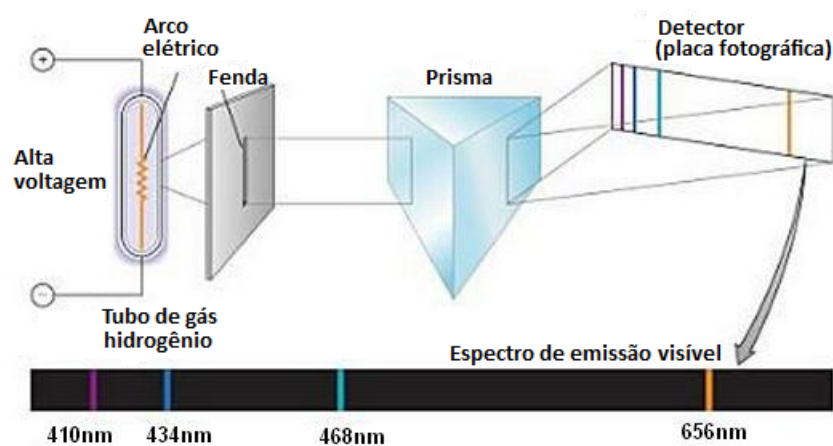


Figura 2.1: Espectro de emissão visível do hidrogênio atômico. Adaptada de TUTORVISTA (2015).

Quando o gás a baixa pressão é submetido a uma alta tensão, uma descarga elétrica induz a emissão de radiação. Parte desta radiação é visível e pode ser analisada utilizando-se uma fenda e um prisma. Ao atravessar o prisma, a luz é desviada (refratada) por ângulos diferentes que dependem de seu comprimento de onda. Projetando a luz que emerge do prisma

sobre uma superfície plana, obtém-se uma imagem formada por um conjunto de linhas (linhas espectrais), cujas posições dependem do comprimento de onda. Se a imagem for projetada sobre um detector, é possível medir a intensidade da luz correspondente a cada comprimento de onda, formando assim o espectro de emissão.

Cada elemento químico tem um espectro de emissão único, que lhe é característico, que o identifica, como uma assinatura ou uma impressão digital (OWEN, 2014). Na Figura 2.2 são apresentados os espectros de emissão visíveis de alguns elementos, evidenciando a clara diferença entre eles. Assim, no caso de amostras de elementos, basta saber a posição (comprimento de onda) das linhas de emissão para reconhecer o elemento correspondente.

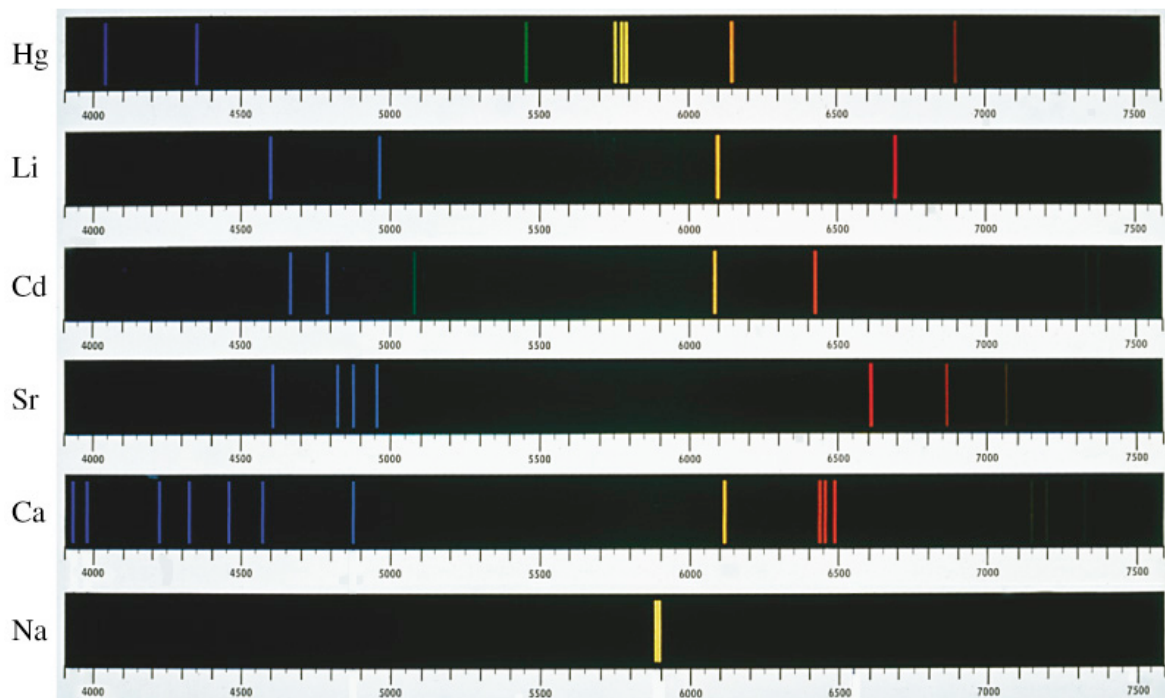


Figura 2.2: Espectros de emissão visíveis de diferentes elementos químicos. Adaptada de DIMAURO (2015).

Na Figura 2.3, o espectro de emissão de uma lâmpada fluorescente compacta é representado na forma de um gráfico da intensidade versus comprimento de onda. Cada pico neste gráfico corresponde a uma linha espectral, cuja altura é proporcional à respectiva intensidade. Este tipo de espectro é mais complexo, sendo formado pela superposição dos espectros dos diferentes elementos químicos presentes no tubo e no seu revestimento. As intensidades relativas das linhas de emissão dos diferentes elementos são determinadas em parte pelas concentrações, ou quantidades relativas, dos respectivos elementos. Assim, a análise deste tipo de espectro consiste em identificar, por seu comprimento de onda, as linhas espectrais dos elementos cons-

tituintes, e medir as intensidades relativas, de modo a determinar também suas quantidades relativas.

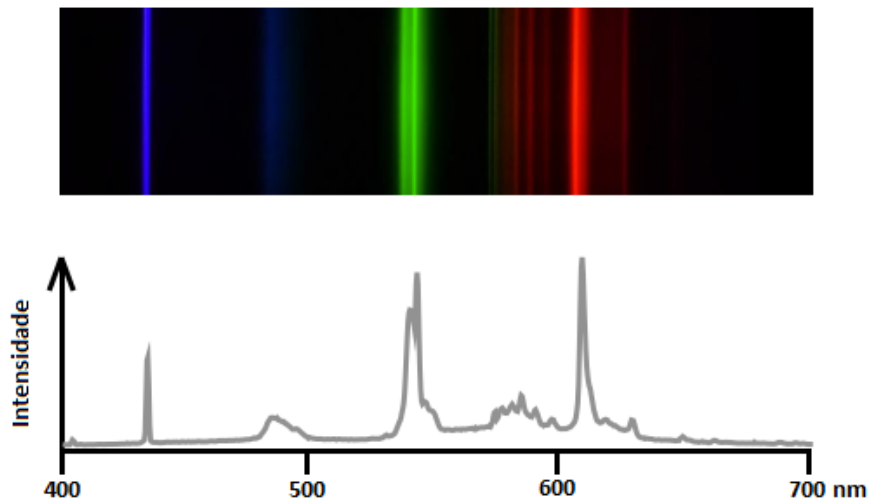


Figura 2.3: Espectro de emissão de uma lâmpada fluorescente compacta representado na forma de um gráfico de intensidade versus comprimento de onda. Adaptada de CORNER (2008) .

Modernamente, conversores analógico-digital (AD) são usados para registrar e armazenar os espectros na forma de arquivos digitais. A faixa de comprimentos de onda a ser analisada é dividida em pequenos intervalos e a intensidade da radiação com comprimento de onda dentro de cada uma destas faixas é registrada em múltiplos inteiros de uma quantidade mínima, conforme é ilustrado na Figura 2.4. Os dados assim obtidos podem ser organizados na forma de uma tabela ou registrados como um *array* unidimensional (1D).

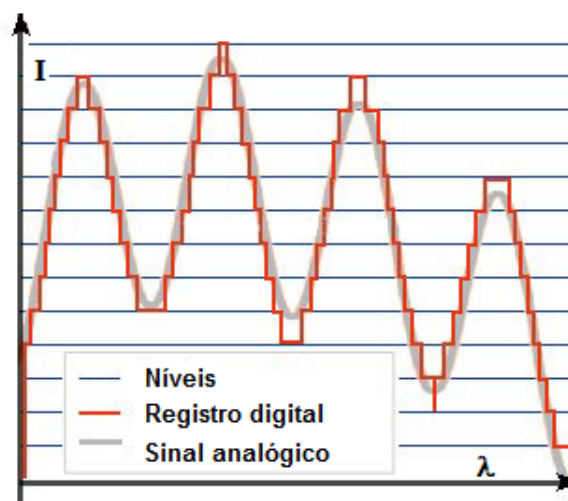


Figura 2.4: Conversão de um sinal analógico em um registro digital. Adaptada de PLATFORM (2015).

A resolução em comprimento de onda e em intensidade dos espectros digitais é variada e depende do sistema de aquisição. Assim, por exemplo, a conversão de um espectro visível, com comprimentos de onda entre 400 e 700nm (nanômetros), feita por um conversor AD com 1024 canais, permite distinguir a radiação detectada com diferença de cerca de 0,3nm ou 3Å (angstroms). Se cada um desses canais reserva 8 bits para o registro da intensidade, então tem-se 256 possíveis níveis de intensidade, o que nos permite observar somente diferenças relativas maiores do que cerca de 0,4%.

A variabilidade de resolução dos espectros obtidos por diferentes equipamentos ou armazenados em diferentes bases de dados, no entanto, não impede os especialistas de avaliar a composição química das amostras analisadas. Daí se conclui que existem características robustas, que se conservam em diferentes versões dos espectros das mesmas substâncias, e que permitem o seu reconhecimento.

Os mesmos conceitos introduzidos aqui para a análise de espectros óticos aplicam-se também para outras formas de espectroscopia. Neste trabalho, são considerados espectros de emissão de raios X característicos produzidos em microsondas eletrônicas (SEM) e em microscópios eletrônicos dotados de detectores de raios X (REED, 1996). Nestes equipamentos, um feixe de elétrons é acelerado em direção a um material sólido que se deseja analisar. Ao atingirem a superfície da amostra, os elétrons produzem uma série de efeitos físicos, entre os quais a emissão de raios X característicos dos elementos químicos que a constituem. Uma vez que o feixe de elétrons pode ser finamente focado sobre a superfície da amostra, os raios X emitidos podem ser atribuídos a uma posição determinada com precisão tipicamente entre 1 e 10  $\mu\text{m}$  (micrômetros). Essa resolução espacial em escala microscópica confere à técnica a denominação de *microanálise*. SMITH; LEIBOVITZ (1986) apontam para o fato de que tal resolução não pode ser obtida por métodos óticos ou de difração de raios X, de uso consagrado para grãos cristalinos grandes, mas que são incapazes de caracterizar crescimentos intergranulares de pequeno volume em amostras policristalinas.

Na Figura 2.5 são mostradas uma foto de uma microsonda eletrônica da marca Jeol e uma representação esquemática do seu funcionamento. No esquema estão representados o feixe de elétrons, a amostra, os raios X emitidos e o sistema de detecção que, neste caso, inclui um cristal que funciona como uma rede de difração que separa os fótons de raios X por seus comprimentos de onda. Este sistema é conhecido como WDS (*wavelength dispersive spectrometer*) (GOLDSTEIN et al., 1992).

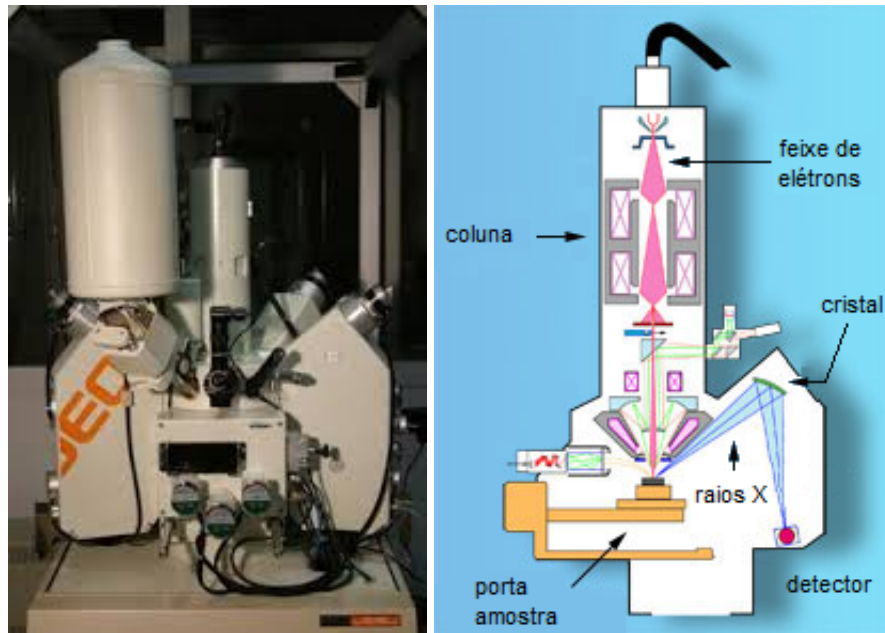


Figura 2.5: Foto de uma microsonda eletrônica da marca Jeol e sua representação esquemática. Adaptadas de MCSWIGGEN (2005).

Outra forma de aquisição de espectros de raios X separa a radiação emitida pela energia de seus fótons, sendo conhecido como EDS (*energy dispersive spectrometer*) (REED, 1996). Na Figura 2.6 é mostrado o espectro EDS do mineral conhecido como Anortita, cuja fórmula química é  $\text{CaAl}_2\text{Si}_2\text{O}_8$ .

Nesse tipo de espectro, o eixo das abcissas exibe as energias dos fótons de raios X – em elétrons-volt (eV) – no lugar do comprimento de onda. Como se vê, as linhas espectrais mais intensas são aquelas dos elementos que constam desta fórmula. Entretanto, as intensidades relativas não são diretamente proporcionais às concentrações atômicas, sendo determinadas também por outros fatores.

De fato, a conversão de intensidades relativas para concentrações atômicas a partir de primeiros princípios, levando em conta todos os fatores intervenientes, é uma tarefa complexa (REED, 1996). Assim, neste trabalho, no lugar de usar a estequiometria, ou “formulação química”, do mineral analisado, busca-se realizar a sua identificação pela comparação direta do espectro problema com os espectros de uma base de dados certificada, tomada como padrão de referência.

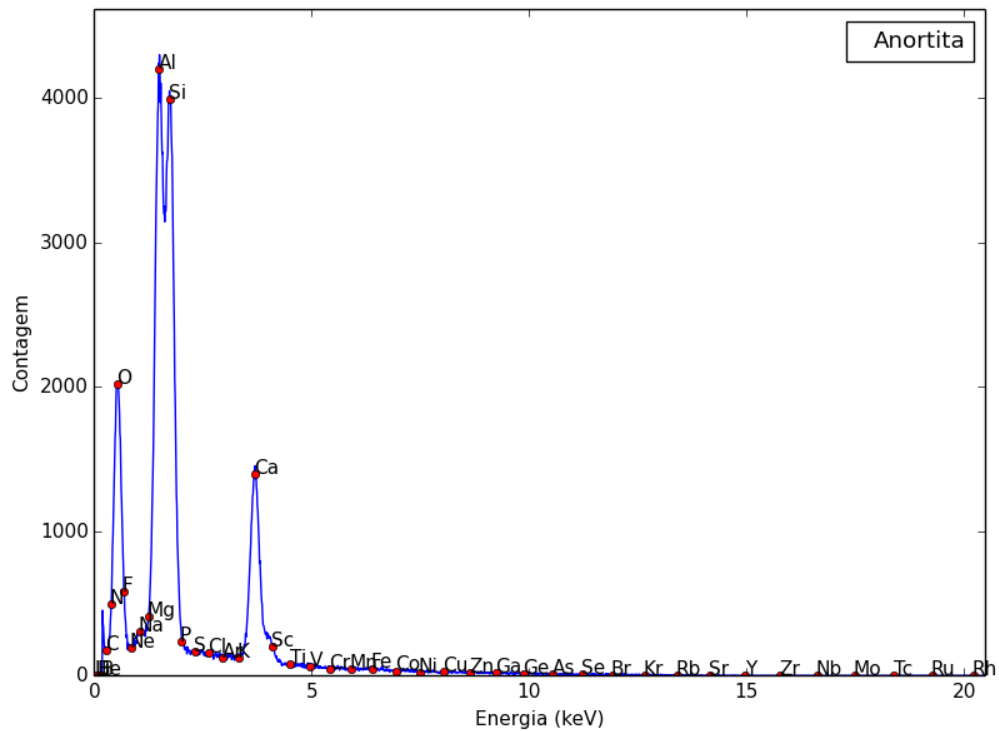


Figura 2.6: Espectro EDS do mineral Anortita ( $\text{CaAl}_2\text{Si}_2\text{O}_8$ ).

## 2.2 Pré-processamento

De acordo com o que foi exposto na Seção 2.1, o reconhecimento de um espectro EDS envolve (i) a identificação linhas espectrais (ocorrência de máximos de emissão de radiação em determinadas faixas de energia), ou picos no espectro, os quais indicam a presença dos elementos químicos a eles associados, e (ii) a determinação das magnitudes relativas destes picos (intensidades relativas de emissão), as quais são proporcionais às concentrações dos elementos correspondentes no mineral que gerou o espectro problema. A presente Seção explica como os procedimentos para resolver (i) e (ii) foram realizados neste estudo.

### 2.2.1 Detecção de linhas espectrais

A escolha de um valor universal para limiar (*threshold*), para decidir se a ocorrência de um máximo local no espectro de raios X deve ou não ser considerado uma linha espectral, ou pico, fica completamente proibida dadas as variabilidades de resolução de aquisição dos espectros e de concentração dos elementos químicos presentes em cada mineral.

A Figura 2.7 apresenta dois espectros de EDS de um mesmo mineral obtidos sob diferentes configurações de aquisição. Observe que os valores do eixo das ordenadas são diferentes

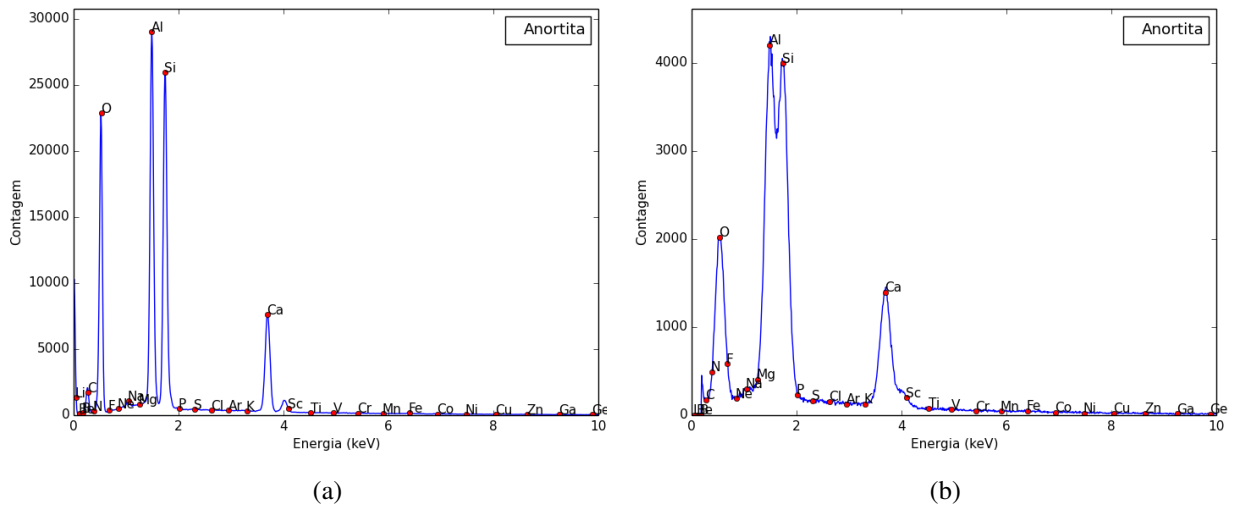


Figura 2.7: Dois espectros do mineral Anortita obtidos sob diferentes resoluções.

por uma relação de aproximadamente 10 vezes. Além disso, o espectro da Figura 2.7(b) apresenta sobreposição de linhas de emissão, como fica evidente na região de energias próximas aos 4keV.

Outra possibilidade para detecção de linhas espectrais é a utilização do conceito de derivada. Assim, um conjunto unidimensional de dados tem um máximo na posição  $i$  quando a derivada (diferença) dos elementos anteriores a  $i$  é crescente e após a  $i$  é decrescente (PRESS et al., 1992). Entretanto tal abordagem não costuma ser eficiente quando o conjunto de dados considerado é ruidoso.

Para contornar este tipo de problema, neste trabalho foi utilizado para detecção de picos o algoritmo provido em (BILLAUER, 2012) e ilustrado em Algoritmo 1. Este considera que um máximo local é um pico caso haja um mínimo local anterior e outro posterior ao “candidato a pico” cuja diferença seja maior ou igual a  $\delta$ . Aqui,  $\delta$  é tomado, adaptativamente, como sendo igual a 2.5% do maior valor no espectro sendo analisado. Este valor foi obtido empiricamente depois de uma série de testes.

### 2.2.2 Determinação das intensidades relativas

Esta última medida se faz contabilizando a área do espectro em faixas de energia centradas nas energias referentes às linhas de emissão de raios X dos elementos químicos presentes e com uma largura de 0.10keV (0.05keV para cada lado). O valor escolhido para esta largura acomoda adequadamente eventuais sobreposições de linhas espectrais de elementos diferentes.

---

**Algoritmo 1** Algoritmo para detecção de picos.
 

---

```

1: procedure PEAKDETECTOR( $X, Y, \delta$ )
2:    $resultSet \leftarrow \{\}$ 
3:    $maxPos \leftarrow 0$ 
4:    $looking4Max \leftarrow 1$ 
5:    $minValue \leftarrow \text{Infinity}$ 
6:    $maxValue \leftarrow -\text{Infinity}$ 
7:   for  $i \leftarrow 0$  to  $i < \text{length}(X)$  do
8:     if  $Y[i] > maxValue$  then
9:        $maxValue \leftarrow Y[i]$ 
10:       $maxPos \leftarrow X[i]$ 
11:     end if
12:     if  $Y[i] < minValue$  then
13:        $minValue \leftarrow Y[i]$ 
14:     end if
15:     if  $looking4Max = 1$  then
16:       if  $Y[i] < (maxValue - \delta)$  then
17:          $resultSet \leftarrow resultSet + \{(maxPos, maxValue)\}$ 
18:          $minValue \leftarrow Y[i]$ 
19:          $looking4Max \leftarrow 0$ 
20:       end if
21:     else
22:       if  $Y[i] > (minValue + \delta)$  then
23:          $maxValue \leftarrow Y[i]$ 
24:          $maxPos \leftarrow X[i]$ 
25:          $looking4Max \leftarrow 1$ 
26:       end if
27:     end if
28:   end for
29:   return ( $resultSet$ )
30: end procedure

```

---



No presente estudo, foram consideradas apenas as linhas de emissão  $K_{\alpha 1}$  na faixa de 1 a 10keV, uma vez que elementos cujos raios X emitidos têm energias abaixo de 1keV são frequentemente caracterizados como “contaminação da amostra” e linhas correspondentes energias acima de 10keV podem ser mais facilmente superpostas às de outros elementos. Além disso, os espectros usados neste trabalho são de minerais cujas principais linhas espectrais estão compreendidas nesta faixa de energia ([1,10]keV). Na Tabela 2.1 encontram-se as energias dos raios X emitidos pelos elementos químicos até cerca de 10 keV. Na segunda e terceira colunas, encontram-se os elementos químicos e as respectivas energias da linha  $K_{\alpha 1}$ .

Tabela 2.1: Tabela de energias de emissão de raios X com  $K_{\alpha 1}$  em [0,~10]KeV (EDAX, 2015).

Z	Elemento	$K_{\alpha 1}$ (eV)	$K_{\alpha 2}$ (eV)	$K_{\beta 1}$ (eV)	$L_{\alpha 1}$ (eV)	$L_{\alpha 2}$ (eV)	$L_{\beta 1}$ (eV)
3	Li	54,3					
4	Be	108,5					
5	B	183,3					
6	C	277					
7	N	392,4					
8	O	524,9					
9	F	676,8					
10	Ne	848,6	848,6				
11	Na	1040,98	1040,98	1071,1			
12	Mg	1253,6	1253,6	1302,2			
13	Al	1486,7	1486,27	1557,45			
14	Si	1739,98	1739,38	1835,94			
15	P	2013,7	2012,7	2139,1			
16	S	2307,84	2306,64	2464,04			
17	Cl	2622,39	2620,78	2815,6			
18	Ar	2957,7	2955,63	3190,5			
19	K	3313,8	3311,1	3589,6			
20	Ca	3691,68	3688,09	4012,7	341,3	341,3	344,9
21	Sc	4090,6	4086,1	4460,5	395,4	395,4	399,6
22	Ti	4510,84	4504,86	4931,81	452,2	452,2	458,4
23	V	4952,2	4944,64	5427,29	511,3	511,3	519,2
24	Cr	5414,72	5405,51	5946,71	572,8	572,8	582,8
25	Mn	5898,75	5887,65	6490,45	637,4	637,4	648,8
26	Fe	6403,84	6390,84	7057,98	705	705	718,5
27	Co	6930,32	6915,3	7649,43	776,2	776,2	791,4
28	Ni	7478,15	7460,89	8264,66	851,5	851,5	868,8
29	Cu	8047,78	8027,83	8905,29	929,7	929,7	949,8
30	Zn	8638,86	8615,78	9572	1011,7	1011,7	1034,7
31	Ga	9251,74	9224,82	10264,2	1097,92	1097,92	1124,8
32	Ge	9886,42	9855,32	10982,1	1188	1188	1218,5

Uma vez que o trabalho propõe a identificação de espectros de minerais potencialmente obtidos a diferentes resoluções, não se pode considerar intensidades (ou contagens) absolutas. Para isso, é necessário aplicar alguma forma de normalização.

Duas possibilidades para normalização das intensidades nas linhas de emissão referentes aos diferentes elementos químicos são:

- **Normalização pela área total do espectro.** Neste caso, as contagens de cada linha espectral (em faixas de 100 eV em torno do pico) são divididas pelo número total de contagens do espectro completo, ou seja a *área total* do espectro.
- **Normalização pela área do maior pico.** Aqui, dividem-se as contagens correspondentes a cada linha espectral pelas *contagens*, ou *área*, *do maior pico*.

A escolha do método de normalização é feita com base nos resultados obtidos nos testes.

### 2.3 Métricas de similaridade

Conforme OWEN (2014), métricas de similaridade podem ser utilizadas para distinção de espectros e posterior reconhecimento e classificação de minerais. Neste trabalho, utiliza-se como características relevantes para o reconhecimento de um espectro: (i) a presença ou ausência de determinadas linhas espectrais e (ii) as intensidades relativas dos picos correspondentes.

Vetores de características do tipo (i) são comparados por métricas de similaridade booleanas, e os do tipo (ii), por métricas não-booleanas. Estas métricas são apresentadas nas Seções 2.3.1 e 2.3.2.

#### 2.3.1 Métricas de similaridade booleanas

Sejam  $\mathbf{u}$  e  $\mathbf{v}$  vetores de tamanho  $N$  contendo apenas variáveis booleanas, cujos valores, um ou zero, representam, respectivamente, a ocorrência ou não de um dado evento associado à posição  $i$  do vetor, com  $i = 1, 2, \dots, N$ . No contexto deste trabalho, o valor 1 representa a ocorrência das linhas espectrais, isto é, a presença de um pico na respectiva posição do espectro, e 0 a ausência deste pico.

Para definir medidas de similaridade entre tais vetores, são introduzidas as contagens

comparativas

$$\alpha(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^N \mathbf{u}_i \cdot \mathbf{v}_i, \quad \text{e}$$

$$\gamma(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^N (\mathbf{u}_i - \mathbf{v}_i)^2,$$

$$\beta(\mathbf{u}, \mathbf{v}) = N - \gamma(\mathbf{u}, \mathbf{v}) - \alpha(\mathbf{u}, \mathbf{v})$$

como sendo, respectivamente, o número total de coocorrências de picos (coincidências ou casamentos do tipo 1-1), o número total de picos discordantes (não-casamentos dos tipos 0-1 e 1-0), o número total de ausências coincidentes (casamentos do tipo 0-0). Na linguagem da álgebra linear,  $\alpha(\mathbf{u}, \mathbf{v})$  e  $\gamma(\mathbf{u}, \mathbf{v})$  são, respectivamente, o produto interno e o quadrado da distância euclidiana entre  $\mathbf{u}$  e  $\mathbf{v}$ . De acordo com JOHNSON; WICHERN (2007), há várias formas de atribuir pesos a  $\alpha(\mathbf{u}, \mathbf{v})$ ,  $\gamma(\mathbf{u}, \mathbf{v})$  e  $\beta(\mathbf{u}, \mathbf{v})$ , de modo a encontrar a melhor medida de similaridade  $\mathbf{u}$  e  $\mathbf{v}$  para um dado problema. Algumas das várias formas de comparar  $\mathbf{u}$  e  $\mathbf{v}$  são apresentadas na Tabela 2.2, em que  $\omega$  é uma constante maior do que 1.

Tabela 2.2: Métricas de similaridade booleanas. Adaptação de JOHNSON; WICHERN (2007).

Equação	Explicação
$d_1(\mathbf{u}, \mathbf{v}) = 1 - \frac{\alpha(\mathbf{u}, \mathbf{v}) + \beta(\mathbf{u}, \mathbf{v})}{N}$	Distância de Hamming normalizada.
$d_2(\mathbf{u}, \mathbf{v}) = 1 - \frac{\omega(\alpha(\mathbf{u}, \mathbf{v}) + \beta(\mathbf{u}, \mathbf{v}))}{\omega(\alpha(\mathbf{u}, \mathbf{v}) + \beta(\mathbf{u}, \mathbf{v})) + \gamma(\mathbf{u}, \mathbf{v})}$	Peso $\omega$ para coocorrências (1-1) e coausências (0-0) de picos.
$d_3(\mathbf{u}, \mathbf{v}) = 1 - \frac{\alpha(\mathbf{u}, \mathbf{v}) + \beta(\mathbf{u}, \mathbf{v})}{\alpha(\mathbf{u}, \mathbf{v}) + \beta(\mathbf{u}, \mathbf{v}) + \omega\gamma(\mathbf{u}, \mathbf{v})}$	Peso $\omega$ para discordâncias de ocorrência de picos (1-0 e 0-1).
$d_4(\mathbf{u}, \mathbf{v}) = 1 - \frac{\alpha(\mathbf{u}, \mathbf{v})}{N}$	Apenas coocorrências de picos (1-1) são consideradas relevantes.
$d_5(\mathbf{u}, \mathbf{v}) = 1 - \frac{\alpha(\mathbf{u}, \mathbf{v})}{\alpha(\mathbf{u}, \mathbf{v}) + \gamma(\mathbf{u}, \mathbf{v})}$	Coausências de picos (0-0) são irrelevantes.
$d_6(\mathbf{u}, \mathbf{v}) = 1 - \frac{\omega\alpha(\mathbf{u}, \mathbf{v})}{\omega\alpha(\mathbf{u}, \mathbf{v}) + \gamma(\mathbf{u}, \mathbf{v})}$	Peso $\omega$ para coocorrências (1-1) e exclusão de coausências de picos (0-0).
$d_7(\mathbf{u}, \mathbf{v}) = 1 - \frac{\alpha(\mathbf{u}, \mathbf{v})}{\alpha(\mathbf{u}, \mathbf{v}) + \omega\gamma(\mathbf{u}, \mathbf{v})}$	Peso $\omega$ para discordâncias (1-0 e 0-1) e exclusão de coausências (0-0).

Para explicar o significado das diferenças entre as métricas JOHNSON; WICHERN (2007) exemplificam: “em uma multidão, o fato de que duas pessoas falam grego antigo é uma evidência mais forte de similaridade entre elas do que se não tivessem tal habilidade”, portanto deve-se ponderar mais fortemente o casamento 1-1 (“ambos falam grego antigo”) do que casamentos 0-0 (“nenhum dos dois fala grego antigo”).

Note que o valor obtido pela aplicação de qualquer uma das equações apresentadas na Tabela 2.2 é um número real no intervalo  $[0,1]$ . Um resultado igual a zero significa que  $\mathbf{u}$  e  $\mathbf{v}$  são “completamente iguais” e um resultado igual a um que estes são “completamente diferentes” sob o ponto de vista dos aspectos analisados em cada uma das equações.

### 2.3.2 Métricas de similaridade não-booleanas

Diferentemente das métricas booleanas, nesta seção são apresentadas métricas de similaridade que consideram vetores de características cujos valores são definidos no domínio dos números reais. Para tal problema, uma infinidade de formulações teoricamente embasadas podem ser utilizadas, dentre as quais destacam-se as distâncias Manhattan, Euclideana, de Minkowski (PRESS et al., 1992), de Bray-Curtis (BRAY; CURTIS, 1957), de Chebyshev (KAUR, 2014), Canberra (EMRAN; YE, 2002), do cosseno (GILLER, 2012) e multiplicativa (MANSOURI; KHADEMI, 2014). Tais medidas de similaridade são frequentemente aplicadas em problemas relacionados ao reconhecimento de padrões (CHEN; MA; ZHANG, 2009), em especial sistemas de busca por conteúdo (TORRES; FALCÃO, 2006) e de apoio a decisão (XUE et al., 2014).

A distância de Minkowski entre dois vetores  $\mathbf{u}$  e  $\mathbf{v}$  de tamanho  $N$  é dada por

$$d_p(\mathbf{u}, \mathbf{v}) = \sqrt[p]{\sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i|^p} \quad ,$$

em que  $p \in \mathbb{R}$  é um parâmetro de controle. As distâncias Manhattan e Euclideana são casos particulares da distância de Minkowski em que  $p = 1$  e  $p = 2$ , respectivamente. A distância de Bray-Curtis entre  $\mathbf{u}$  e  $\mathbf{v}$  é dada por

$$d_9(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i|}{\sum_{i=1}^N |\mathbf{u}_i + \mathbf{v}_i|} \quad .$$

Note que esta formulação não tem um resultado definido caso todos os valores dos vetores  $\mathbf{u}$  e  $\mathbf{v}$  forem iguais a zero.

A distância de Chebyshev entre os vetores  $\mathbf{u}$  e  $\mathbf{v}$  considera apenas a componente que mais difere, isto é,

$$d_{10}(\mathbf{u}, \mathbf{v}) = \max_i |\mathbf{u}_i - \mathbf{v}_i| ,$$

em que  $\max(\cdot)$  é a função de maximização de seu argumento em função de  $i$ .

A distância Canberra de  $\mathbf{u}$  a  $\mathbf{v}$ , dada por

$$d_{11}(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i|}{\sum_{i=1}^N |\mathbf{u}_i| + |\mathbf{v}_i|} ,$$

é definida de forma similar à de Bray-Curtis.

Já a distância do cosseno entre os mesmos vetores é obtida através de

$$d_{12}(\mathbf{u}, \mathbf{v}) = 1 - \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} ,$$

em que o ponto representa o produto escalar entre os vetores  $\mathbf{u}$  e  $\mathbf{v}$ , e  $\|\cdot\|$  é a norma Euclideana ou módulo dos vetores  $\mathbf{u}$  e  $\mathbf{v}$  (GOLUB; LOAN, 1985). A recentemente proposta distância multiplicativa entre  $\mathbf{u}$  e  $\mathbf{v}$  é computada por

$$d_{13}(\mathbf{u}, \mathbf{v}) = \prod_{i=1}^N (1 + |\mathbf{u}_i - \mathbf{v}_i|)^{p_i} - 1,$$

em que  $p_i$  é um parâmetro de ponderação dependente de  $i$ . Como ilustrado em MANSOURI; KHADEMI (2014),  $p_i$  pode ser previamente fixado, sem variar de acordo com  $i$ , o que dá lugar a um parâmetro  $p$ . Nos experimentos realizados neste estudo fixou-se  $p_i = p = 1$ , uma vez que quando há grandes valores diferença entre  $\mathbf{u}_i$  e  $\mathbf{v}_i$ , com  $i \in 1, 2, \dots, N$ , o expoente  $p > 1$  tende a estourar o limite de representação de números em ponto flutuante da linguagem de programação utilizada.

Da mesma forma com o que ocorre com as métricas de similaridade booleanas, quanto menor for o resultado da aplicação das métricas não-booleanas, maior a similaridade entre os dois vetores argumento.

## 2.4 Visão geral da metodologia proposta

A Figura 2.8 apresenta, em alto nível, o fluxo de execução do método desenvolvido no presente estudo, o que engloba os tópicos discutidos nas Seções 2.1, 2.2 e 2.3.

Primeiramente, é importada uma tabela de energias de emissão de raios X característicos de cada elemento químico. Inicia-se então a composição da base de informações de minerais certificados. Cada espectro EDS de um dado mineral, armazenado em um arquivo no formato EMSA (ISO/TC 202, 2012), é importado. Após isto, esse espectro é decomposto e suas características são extraídas. Isso é repetido até que todos os minerais que devem compor a base tenham sido processados. Após, o mesmo procedimento é feito com o espectro do mineral problema. Neste ponto, métricas de similaridade entre o vetor de características do mineral problema com os da base de dados são computadas. Ao fim, os resultados ordenados por similaridade são apresentados ao usuário.

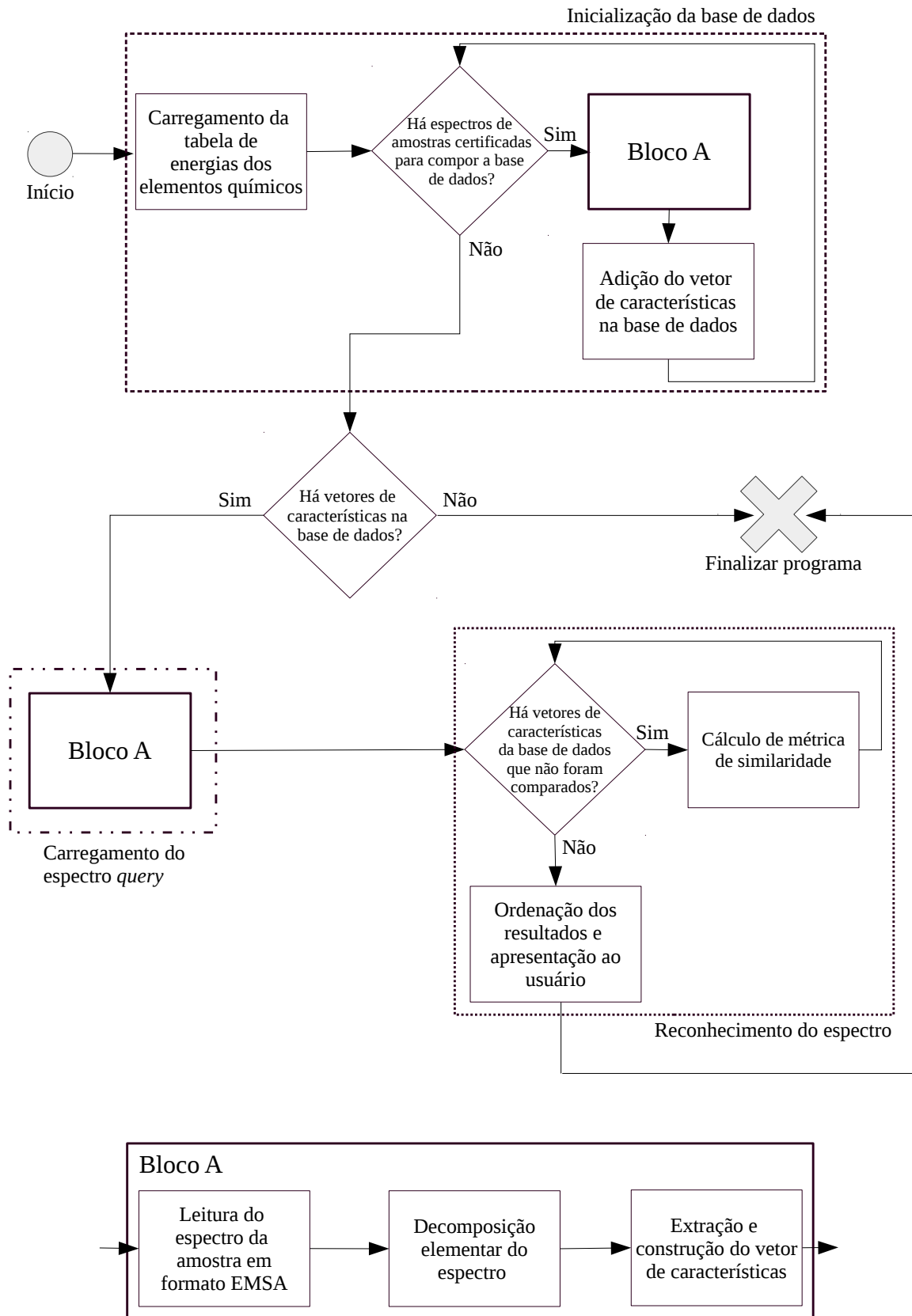


Figura 2.8: Fluxograma da metodologia proposta.

### 3 RESULTADOS E DISCUSSÕES

Conforme mencionado no Capítulo 1, uma vez que não temos espectros de várias amostras de um mesmo mineral, não é possível aplicar algoritmos de aprendizado de máquina, como aqueles utilizados em DING; COLPAN (2006); AKKAŞ et al. (2015). Ainda assim, a grande maioria de algoritmos de classificação presentes na ferramenta *Waikato Environment for Knowledge Analysis* (WEKA) (HALL et al., 2009), bem como seus parâmetros, foram testados. Acurácias máximas ficaram próximas aos 50%.

Visando obter melhores resultados, as métricas de similaridade discutidas na Seção 2.3 foram investigadas. A próxima seção explica as formas de avaliação de resultados utilizadas neste estudo. As outras duas seções deste capítulo apresentam os resultados considerando dois diferentes contextos.

#### 3.1 Métricas de avaliação

Neste estudo são utilizadas duas métricas de avaliação do desempenho do sistema: acurácia (WITTEN; FRANK; HALL, 2011) e o que denominamos “chance do usuário”. Métricas de avaliação específicas por classe não foram consideradas importantes uma vez que se tem apenas um exemplar de cada mineral em cada base de dados.

Suponha um vetor de características  $\mathbf{a}_k$  de um mineral de nome  $X$ . Considere também uma base de dados  $B$  de tamanho  $L$  que contém um vetor de características  $\mathbf{b}_k$  do mesmo mineral  $X$  – potencialmente adquirido de outras amostras, com outras resoluções, em equipamentos diferentes, etc. Quer-se submeter  $\mathbf{a}_k$  ao sistema de reconhecimento para teste. O mineral  $X$  é corretamente classificado caso

$$d_*(\mathbf{a}_k, \mathbf{b}_k) = \min_i d_*(\mathbf{a}_k, \mathbf{b}_i),$$

em que  $d_*(.)$  é a métrica de similaridade escolhida (ver Seções 2.3.1 e 2.3.2). Em palavras, o vetor de características  $\mathbf{a}_k$  de  $X$  é reconhecido com sucesso caso sua “distância” em relação a  $\mathbf{b}_k$  for a menor (mas potencialmente igual a outras) quando comparada com a de todos os vetores de características  $\mathbf{b}_i, i \neq k, i = 1, 2, \dots, L$  em  $B$ .

Note que  $\min_i d_*(\mathbf{a}_k, \mathbf{b}_i)$  pode ter sido satisfeito por mais de um vetor de características  $\mathbf{b}_i$ . Denomina-se  $\mathcal{C}$  o conjunto de minerais associados aos vetores de características que satisfizeram a operação de minimização.



Neste trabalho, a acurácia de um método baseado em  $d_*(.)$  o qual deve reconhecer um conjunto  $A$  de  $M$  queries (conjunto de teste) a partir de uma base dados  $B$  é definida por

$$\text{Acurácia}(A, B) = \frac{\sum_{j=1}^M \text{Contagem}(\mathbf{a}_j, B)}{M}$$

em que

$$\text{Contagem}(\mathbf{a}_j, B) = \begin{cases} 1, & \text{se o mineral associado a } \mathbf{a}_j \text{ está em } \mathcal{C} \\ 0, & \text{caso contrário} \end{cases}.$$

Por outro lado, a “chance do usuário”, considera não apenas o fato de o mineral a ser reconhecido estar entre aqueles de maior similaridade (conjunto  $\mathcal{C}$ ). Neste caso, há um “desconto na acurácia do método” caso haja mais de um mineral dentre as recomendações do sistema (aqueles em  $\mathcal{C}$ ). Interpreta-se esta métrica como a chance de o usuário identificar corretamente um mineral caso ele siga as indicações do sistema de reconhecimento. Calcula-se a “chance do usuário” através de

$$\text{Chance}(A, B) = \frac{\sum_{j=1}^M \text{Proporção}(\mathbf{a}_j, B)}{M}$$

em que

$$\text{Proporção}(\mathbf{a}_j, B) = \begin{cases} 1/\#\mathcal{C}, & \text{se o mineral associado a } \mathbf{a}_j \text{ está em } \mathcal{C} \\ 0, & \text{caso contrário} \end{cases}$$

e  $\#$  é o operador de cardinalidade de conjuntos.

### 3.2 Experimentos com dados certificados

Para avaliação, neste estudo, são considerados duas bases de dados, os quais foram obtidas no *Laboratorio de Microscopía Electrónica y Análisis por Rayos X (LAMARX)* da *Universidad Nacional de Córdoba*, Argentina. Estes conjuntos de dados contém espectros de diversos minerais, cujas amostras são certificadas, que foram capturados por diferentes dispositivos equipados com detectores EDS.

Os minerais das bases utilizadas, nomeadas de SEM e EPMA como referência aos instrumentos utilizados para captura, uma para treino e outra para teste, são apresentados nas Tabelas 3.1 e 3.2. O conjunto de minerais apresentado na Tabela 3.1 foi obtido em um equipamento que permite maior resolução de amostragem.

Nos resultados de acurácia e “chance do usuário”, são considerados apenas os minerais que são comuns às duas bases, embora o conjunto de dados utilizado como base seja considerado por inteiro. Os minerais comuns a ambas as bases estão destacados nas Tabelas 3.1 e 3.2 em itálico.

Tabela 3.1: Base de dados SEM.

Mineral	Fórmula química
<i>Anortita</i>	$\text{CaAl}_2\text{Si}_2\text{O}_8$
<i>Anortoclasa</i>	$(\text{Na,K})\text{AlSi}_3\text{O}_8$
<i>Apatita</i>	$\text{Ca}_5(\text{PO}_4)_3(\text{OH, F, Cl})$ .
<i>Calcita</i>	$\text{CaCO}_3$
<i>Cromita</i>	$\text{FeCr}_2\text{O}_4$
<i>Diopsida</i>	$\text{CaMgSi}_2\text{O}_6$
<i>Faialita</i>	$\text{Fe}_2\text{SiO}_4$
<i>Flogopita</i>	$\text{KMg}_3(\text{AlSi}_3\text{O}_{10})(\text{F,OH})_2$
<i>F-Topázio</i>	$\text{Al}_2(\text{F,OH})_2\text{SiO}_4$
<i>Horneblenda</i>	$\text{Ca}_2(\text{Mg, Fe, Al})_5(\text{Al, Si})_8\text{O}_{22}(\text{OH})_2$
<i>Ilmenita</i>	$\text{FeTiO}_3$
<i>Labradorita</i>	$(\text{Ca,Na})(\text{Al,Si})_4\text{O}_8$
<i>Magnetita</i>	$\text{Fe}_3\text{O}_4$
<i>Olivina</i>	$(\text{Mg,Fe})_2\text{SiO}_4$
<i>Ortoclasa</i>	$\text{KAlSi}_3\text{O}_8$
<i>Pirofilita</i>	$\text{Al}_2\text{Si}_4\text{O}_{10}(\text{OH})_2$
<i>Piropo</i>	$\text{Mg}_3\text{Al}_2(\text{SiO}_4)_3$
<i>Riolita</i>	$69\%^+ \text{SiO}_2$
<i>Rodocrosita</i>	$\text{MnCO}_3$
<i>Rodonita</i>	$(\text{Mn,Ca,Fe})\text{SiO}_3$
<i>Titanita</i>	$\text{CaTiOSiO}_4$
<i>Wollastonita</i>	$\text{CaSiO}_3$
<i>Zoisita</i>	$\text{Ca}_2\text{Al}_3\text{Si}_3\text{O}_{12}(\text{OH})$

Tabela 3.2: Base de dados EPMA.

Mineral	Fórmula química
Albita	$\text{NaAlSi}_3\text{O}_8$
Anortita	$\text{CaAl}_2\text{Si}_2\text{O}_8$
Anortoclasa	$(\text{Na},\text{K})\text{AlSi}_3\text{O}_8$
Apatita	$\text{Ca}_5(\text{PO}_4)_3(\text{OH}, \text{F}, \text{Cl})$ .
Atacamita	$\text{Cu}_2\text{Cl}(\text{OH})_3$
Calcita	$\text{CaCO}_3$
Cromita	$\text{FeCr}_2\text{O}_4$
Crocoíta	$\text{PbCrO}_4$
Cuprita	$\text{CuO}_2$
Danburita	$\text{CaB}_2(\text{Si}_2\text{O}_8)$
Diopsida	$\text{CaMgSi}_2\text{O}_6$
Dioptasa	$\text{CuSiO}_2(\text{OH})_2$
Flogopita	$\text{KMg}_3(\text{AlSi}_3\text{O}_{10})(\text{F},\text{OH})_2$
Fluorita	$\text{CaF}_2$
F-Topázio	$\text{Al}_2(\text{F},\text{OH})_2\text{SiO}_4$
Hematita	$\text{Fe}_2\text{O}_3$
Horneblenda	$\text{Ca}_2(\text{Mg}, \text{Fe}, \text{Al})_5(\text{Al}, \text{Si})_8\text{O}_{22}(\text{OH})_2$
Ilmenita	$\text{FeTiO}_3$
Labradorita	$(\text{Ca},\text{Na})(\text{Al},\text{Si})_4\text{O}_8$
Magnetita	$\text{Fe}_3\text{O}_4$
Niquelina	$\text{NiAs}$
Ortoclasa	$\text{KAlSi}_3\text{O}_8$
Petalita	$\text{LiAlSi}_4\text{O}_{10}$
Pirofilita	$\text{Al}_2\text{Si}_4\text{O}_{10}(\text{OH})_2$
Piropo	$\text{Mg}_3\text{Al}_2(\text{SiO}_4)_3$
Quartzo	$\text{SiO}_2$
Riolita	$69\%^+ \text{SiO}_2$
Rodocrosita	$\text{MnCO}_3$
Rodonita	$(\text{Mn},\text{Ca},\text{Fe})\text{SiO}_3$
Soldalita	$\text{Na}_4\text{Al}_3(\text{SiO}_4)_3\text{Cl}$
Titanita	$\text{CaTiOSiO}_4$
Wollastonita	$\text{CaSiO}_3$
Zoisita	$\text{Ca}_2\text{Al}_3\text{Si}_3\text{O}_{12}(\text{OH})$

As duas possibilidades para teste foram avaliadas: utilizando SEM como base e EPMA como *query* (Cenário I) e vice-versa (Cenário II). Os resultados, utilizando métricas de similaridade booleanas e não-booleanas, para ambos os cenários de avaliação são exibidos, respectivamente, nas Tabelas 3.3 e 3.4.

Tabela 3.3: Comparação de métricas de similaridade booleanas para os Cenários I e II.

Métrica de Similaridade	Cenário I		Cenário II	
	Acurácia	Chance do Usuário	Acurácia	Chance do Usuário
$d_1(\cdot)$	90,48%	68,57%	85,71%	63,25%
$d_2(\cdot)$ com $\omega = 2$	90,48%	70%	85,71%	64,68%
$d_3(\cdot)$ com $\omega = 2$	90,48%	68,57%	85,71%	64,68%
$d_4(\cdot)$	85,71%	37,16%	66,66%	37,37%
$d_5(\cdot)$	90,48%	72,22%	90,48%	71,19%
$d_6(\cdot)$ com $\omega = 2$	90,48%	72,22%	90,48%	71,19%
$d_7(\cdot)$ com $\omega = 2$	90,48%	72,22%	90,48%	72,62%

Tabela 3.4: Comparação de métricas de similaridade não-booleanas para os Cenários I e II.

Tipo de normalização	Métrica de similaridade	Cenário I		Cenário II	
		Acurácia	Chance do Usuário	Acurácia	Chance do Usuário
Área total	$d_8(\cdot)$ com $p = 1$	90,48%	90,48%	85,71%	85,71%
	$d_8(\cdot)$ com $p = 2$	80,95%	80,95%	80,95%	78,57%
	$d_8(\cdot)$ com $p = 3$	80,95%	80,95%	76,19%	76,19%
	$d_9(\cdot)$	90,48%	90,48%	80,95%	80,95%
	$d_{10}(\cdot)$	66,66%	66,66%	90,48%	90,48%
	$d_{11}(\cdot)$	61,90%	61,90%	85,71%	85,71%
	$d_{12}(\cdot)$	100%	100%	90,48%	90,48%
	$d_{13}(\cdot)$ com $p = 1$	90,48%	90,48%	85,71%	85,71%
Área do pico maior	$d_8(\cdot)$ com $p = 1$	100%	100%	90,48%	90,48%
	$d_8(\cdot)$ com $p = 2$	100%	100%	90,48%	90,48%
	$d_8(\cdot)$ com $p = 3$	100%	100%	90,48%	90,48%
	$d_9(\cdot)$	100%	100%	90,48%	90,48%
	$d_{10}(\cdot)$	100%	100%	90,48%	90,48%
	$d_{11}(\cdot)$	61,90%	61,90%	33,33%	33,33%
	$d_{12}(\cdot)$	100%	100%	90,48%	90,48%
	$d_{13}(\cdot)$ com $p = 1$	100%	100%	90,48%	90,48%

Note que medidas de similaridade booleanas apresentaram “chances do usuário” bastante baixas. Isso, pela definição, dá-se uma vez que o sistema apresenta várias opções como resposta. Neste contexto, o Cenário I – que apresenta a base de dados de melhor resolução – mostrou-se como preferível em todas menos a última métricas de similaridade avaliadas.

Em relação às métricas de similaridade não-booleanas, observa-se a clara vantagem de

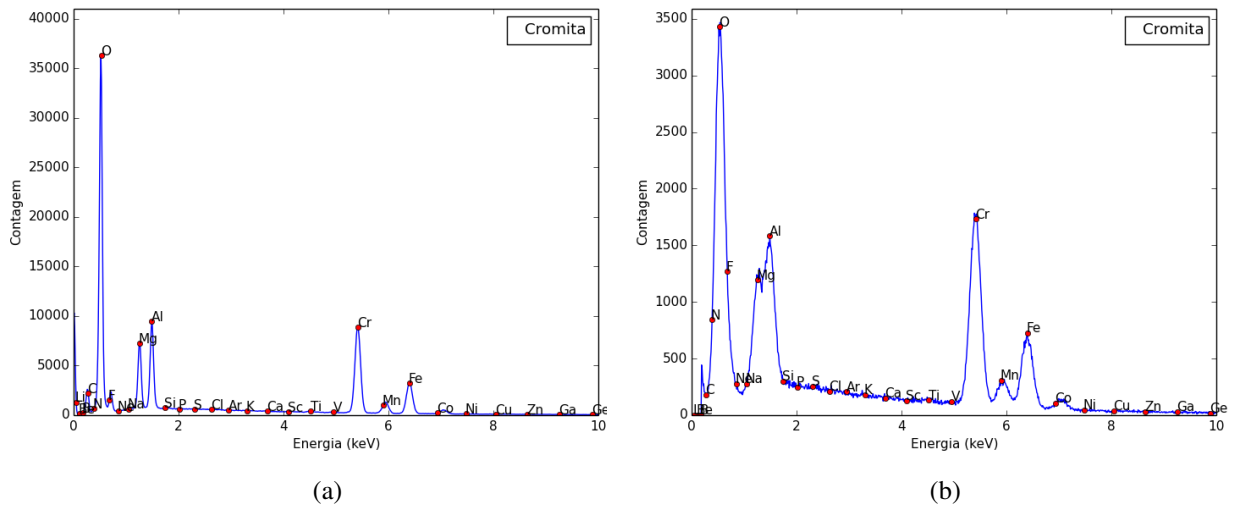


Figura 3.1: Espectros do mineral Cromita das bases de dados (a) SEM e (b) EPMA.

se utilizar a normalização pela área do pico maior. É evidente o comportamento estável deste tipo de normalização, quase que independentemente da métrica de similaridade escolhida. Por outro lado, nota-se o péssimo desempenho da distância Canberra para o problema em questão quando comparada com as demais – em especial no Cenário I com normalização pela área total e nos Cenários I e II com a normalização feita pela área do pico maior.

Fica claro também que ambas as métricas de avaliação de desempenho do sistema apresentaram mesmos resultados quando os vetores de características dos espectros não são booleanos. Isso é bastante óbvio, uma vez que há uma infinidade de valores para cada característica do espectro, tornando quase impossível a obtenção de um mesmo índice de similaridade para dois espectros diferentes da base de dados.

A Figura 3.1 apresenta os espectros do mineral Cromita das bases SEM e EPMA. Note que a Cromita foi, com sucesso, classificada (ver Tabela 3.4) mesmo com as grandes diferenças em resolução – possível falha do *software* QEMSCAM – dificultando a separação de várias linhas espectrais e apresentando valores absolutos com um fator multiplicativo de mais de 11 vezes. Ao mesmo tempo, tais espectros ilustram o bom funcionamento dos métodos testados mesmo quando ocorre a inversão de intensidades de alguns picos (Al e Cr) – possível falha do *software* MLA.

Tabela 3.5: Bases de dados em MINERAL SPECTRA (2014).

Base(s) de dados	Mineral	Fórmula química
<i>Thin window</i> e <i>Be window</i>	Anortita	$\text{CaAl}_2\text{Si}_2\text{O}_8$
<i>Thin window</i> e <i>Be window</i>	Apatita	$\text{Ca}_5(\text{PO}_4)_3(\text{OH}, \text{F}, \text{Cl})$ .
<i>Thin window</i>	Cromita	$\text{FeCr}_2\text{O}_4$
<i>Thin window</i> e <i>Be window</i>	Diopsida	$\text{CaMgSi}_2\text{O}_6$
<i>Thin window</i> e <i>Be window</i>	Flogopita	$\text{KMg}_3(\text{AlSi}_3\text{O}_{10})(\text{F}, \text{OH})_2$
<i>Thin window</i> e <i>Be window</i>	F-Topázio	$\text{Al}_2(\text{F}, \text{OH})_2\text{SiO}_4$
<i>Thin window</i> e <i>Be window</i>	Horneblenda	$\text{Ca}_2(\text{Mg}, \text{Fe}, \text{Al})_5(\text{Al}, \text{Si})_8\text{O}_{22}(\text{OH})_2$
<i>Thin window</i>	Ilmenita	$\text{FeTiO}_3$
<i>Thin window</i>	Rodocrosita	$\text{MnCO}_3$
<i>Thin window</i>	Titanita	$\text{CaTiOSiO}_4$

### 3.3 Experimentos com dados pós-processados

Além das bases de dados utilizadas nos experimentos da seção anterior, aqui são considerados espectros reconstruídos<sup>1</sup> através de imagens disponíveis em MINERAL SPECTRA (2014). Tal situação explora a robustez do método implementado, mesmo quando há grande perda de informações dos espectros originais, uma vez que as conversões de dados numéricos para imagens e vice-versa introduzem erros.

Os espectros originais em MINERAL SPECTRA (2014) foram obtidos sob duas tecnologias de aquisição: *thin window* – que, na prática, capta energias de elementos leves (valores para  $K_{\alpha 1}$  menores que 1keV); e *Be window* – que desconsidera tais energias. Nos experimentos aqui realizados, apenas espectros daqueles minerais que constam na base SEM, tomada como base de dados referência, são utilizados. A Tabela 3.5 apresenta os minerais considerados em ambos os subconjuntos de dados, aqui denotados de acordo com a tecnologia de aquisição empregada, *thin window* e *Be window*.

Dois experimentos foram realizados considerando SEM como base de dados certificados e tomando como *query* os espectros de *thin window* (Cenário III) e os de *Be window* (Cenário IV). Os resultados, seguindo a mesma notação utilizada na Seção 3.2, são exibidos nas Tabelas 3.6 e 3.7.

Note que, como há apenas 6 espectros na base *Be window* e 10 na base *thin window*, o não reconhecimento de um mineral contribui fortemente para redução do valores de acurácia e “chance do usuário”. Grandes variações nos resultados referentes ao Cenário III, onde valores

<sup>1</sup> Dados processados por Carolina Santos dos Santos, integrante do grupo de pesquisas relacionado a esse trabalho, através de manipulação de imagens.

Tabela 3.6: Comparação de métricas de similaridade booleanas para os Cenários III e IV.

Métrica de Similaridade	Cenário III		Cenário IV	
	Acurácia	Chance do Usuário	Acurácia	Chance do Usuário
$d_1(\cdot)$	80%	65%	83,33%	66,66%
$d_2(\cdot)$ com $\omega = 2$	80%	65%	83,33%	66,66%
$d_3(\cdot)$ com $\omega = 2$	80%	65%	83,33%	66,66%
$d_4(\cdot)$	80%	64,42%	83,33%	48,61%
$d_5(\cdot)$	90%	85%	83,33%	66,66%
$d_6(\cdot)$ com $\omega = 2$	90%	80%	83,33%	66,66%
$d_7(\cdot)$ com $\omega = 2$	90%	80%	83,33%	66,66%

Tabela 3.7: Comparação de métricas de similaridade não-booleanas para os Cenários III e IV.

Tipo de normalização	Métrica de similaridade	Cenário III		Cenário IV	
		Acurácia	Chance do Usuário	Acurácia	Chance do Usuário
Área total	$d_8(\cdot)$ com $p = 1$	70%	70%	83,33%	83,33%
	$d_8(\cdot)$ com $p = 2$	70%	70%	83,33%	83,33%
	$d_8(\cdot)$ com $p = 3$	70%	70%	83,33%	83,33%
	$d_9(\cdot)$	70%	70%	83,33%	83,33%
	$d_{10}(\cdot)$	70%	70%	83,33%	83,33%
	$d_{11}(\cdot)$	70%	70%	83,33%	83,33%
	$d_{12}(\cdot)$	80%	80%	83,33%	83,33%
	$d_{13}(\cdot)$ com $p = 1$	70%	70%	83,33%	83,33%
Área do pico maior	$d_8(\cdot)$ com $p = 1$	60%	60%	83,33%	83,33%
	$d_8(\cdot)$ com $p = 2$	80%	80%	83,33%	83,33%
	$d_8(\cdot)$ com $p = 3$	80%	80%	83,33%	83,33%
	$d_9(\cdot)$	70%	70%	83,33%	83,33%
	$d_{10}(\cdot)$	80%	80%	83,33%	83,33%
	$d_{11}(\cdot)$	60%	60%	83,33%	83,33%
	$d_{12}(\cdot)$	80%	80%	83,33%	83,33%
	$d_{13}(\cdot)$ com $p = 1$	60%	60%	83,33%	83,33%

mais acurados são obtidos para métricas de similaridade booleanas, são observadas. Nota-se inclusive que a “chance do usuário” para  $d_5(\cdot)$  é maior que a acurácia de todas as métricas não-booleanas aplicadas do mesmo cenário, em ambos os tipo de normalização.

Independentemente da métrica de similaridade escolhida, a acurácia mantém-se constante para os minerais da base Be *window*, onde apenas o mineral Anortita não é corretamente reconhecido. Quando este mineral é submetido ao processo de reconhecimento, na maioria das vezes, ele é confundido com o mineral Zoisita. A Figura 3.2 compara os espectros desses minerais: Anortita e Zoisita. Note que tais minerais diferem apenas pelas concentrações de seus elementos (ver Tabela 3.1 ou Tabela 3.2) o que justifica os erros no reconhecimento. Note

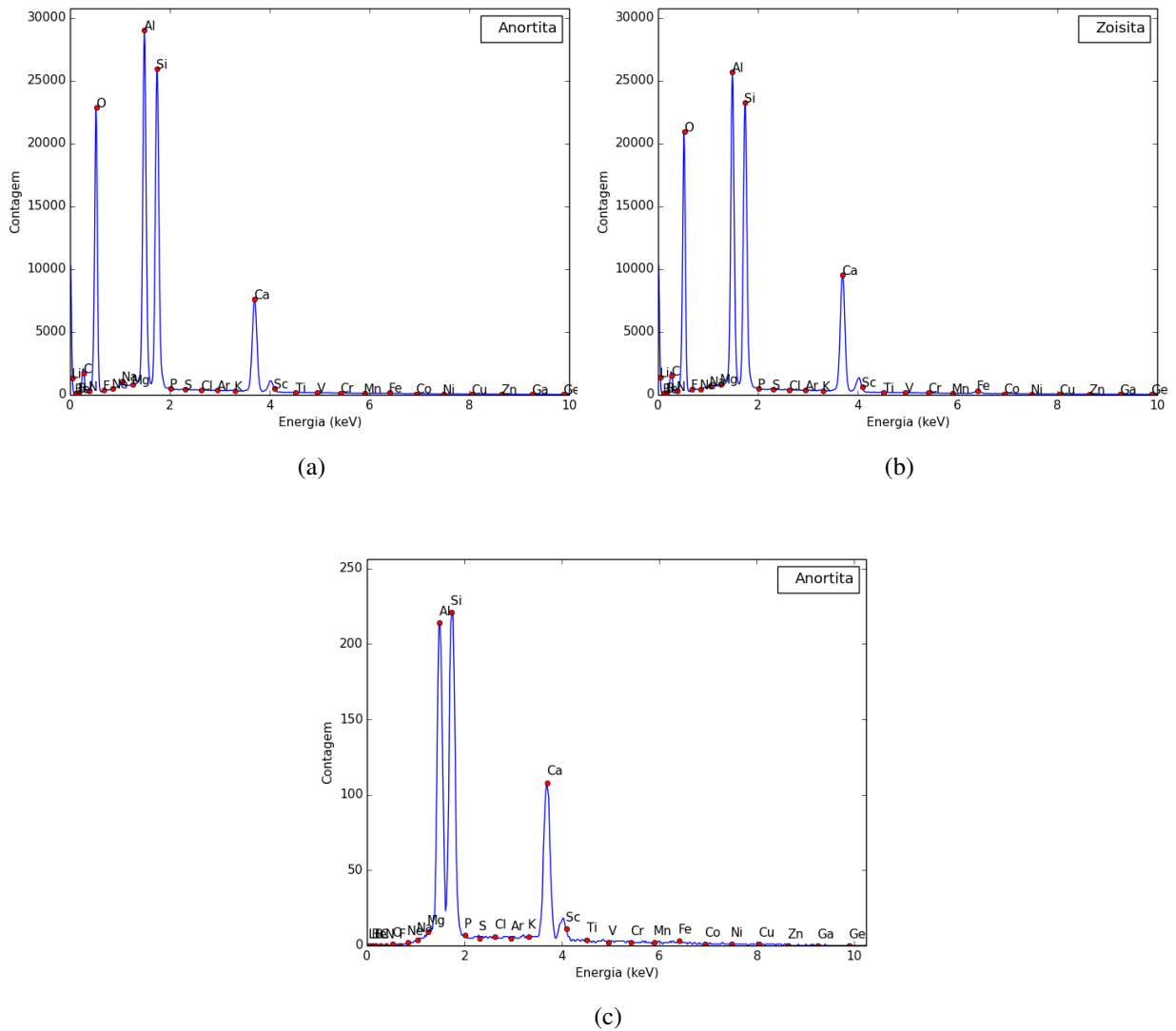


Figura 3.2: Espectros dos minerais (a) Anortita e (b) Zoisita da base de dados SEM e (c) Anortita da base de dados Be *window*.

também que, apesar disso, esses minerais foram corretamente identificados no Cenário I (ver Tabela 3.4).



## 4 CONCLUSÕES

O presente trabalho estudou técnicas para proposta de um sistema de apoio à decisão para reconhecimento de minerais através de seus espectros EDS de raios X característicos. Neste sentido, foram apresentadas várias combinações entre técnicas de pré-processamento para extração de características booleanas e não-booleanas e métricas de similaridade de mesmo escopo. A metodologia foi avaliada em dois contextos distintos: um com dados certificados, obtidos sob diferentes configurações e equipamentos de aquisição; e outro com dados reconstruídos, o que dá um caráter forte de avaliação de robustez do sistema, uma vez que se tem distorções intrínsecas ao processo de conversão dos dados (sequência numérica para imagem, e vice-versa).

Duas métricas de avaliação foram consideradas: a acurácia e a “chance do usuário” – particularmente útil para avaliação dos resultados obtidos através de métricas de similaridade booleanas. Os resultados mostraram que, quando considerados apenas dados certificados, métricas de similaridade não-booleanas reagem melhor. Ademais, a normalização pela área do maior pico no espectro, tende a trazer melhores resultados. Quando o método considera, para teste, dados pós-processados, melhores resultados são obtidos através da aplicação de métricas de similaridade booleanas. Este fato nos leva a crer que novos estudos, tomando o presente como base, devem ser realizados para entender esse comportamento.

Como trabalhos futuros pretende-se seguir buscando por bases de dados públicas para avaliação de técnicas que estendam as aqui desenvolvidas. A combinação de métricas de similaridade booleanas, aplicadas como uma espécie de pré-processamento, juntamente com as não-booleanas pode trazer bons resultados, principalmente a medida que a base de dados cresce. Tais resultados não são testados no presente estudo uma vez que, com o uso de métricas puramente não-booleanas, já é possível ter a taxa máxima de acertos nos Cenários I e II.

## REFERÊNCIAS

- AKKAŞ, E. et al. Application of Decision Tree Algorithm for classification and identification of natural minerals using SEM-EDS. **Computers & Geosciences**, [S.l.], v.80, p.38 – 48, 2015.
- BENTZ, D. et al. SEM/X-ray imaging of cement based materials. In: EUROSEMINAR ON MICROSCOPY APPLIED TO BUILDING MATERIALS, 7. **Proceedings...** [S.l.: s.n.], 1999. p.223–232.
- BILLAUER, E. **peakdet**: peak detection using matlab. Disponível em <http://www.billauer.co.il/peakdet.html>. Acessado em 12/10/2015.
- BRAY, J. R.; CURTIS, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. **Ecological Monographs**, [S.l.], v.27, n.4, p.325, Oct 1957.
- CHEN, S.; MA, B.; ZHANG, K. On the similarity metric and the distance metric. **Theoretical Computer Science**, [S.l.], v.410, n.24-25, p.2365–2376, May 2009.
- CORNER, B. **Spectrum of Fluorescent Light**. Em <http://bealecorner.org/best/measure/cf-spectrum/index.html>. Acessado em 12/10/2015.
- DIMAURO, T. **Physics Education Research**. Em [http://sdsu-physics.org/physics180/physics180B/Topics/light/images\\_light/line\\_spectra.jpg](http://sdsu-physics.org/physics180/physics180B/Topics/light/images_light/line_spectra.jpg). Acessado em 12/10/2015.
- DING, Q.; COLPAN, M. Decision tree induction on hyper-spectral cement images. **International Journal of Computational Intelligence**, [S.l.], v.2, n.3, 2006.
- EDAX. **EDAX Peak Identification Chart**. Disponível em <http://www.edax.com/>. Acessado em 07/10/2015.
- EMRAN, S. M.; YE, N. Robustness of Chi-square and Canberra distance metrics for computer intrusion detection. **Quality and Reliability Engineering International**, [S.l.], v.18, n.1, p.19–28, Jan 2002.
- FEI. **Quantitative Evaluation of Minerals by Scanning Electron Microscopy (QEMSCAN)**. Informações técnicas do *software* em <http://www.fei.com/products/sem/qemscan/>. Acessado em 07/10/2015.

FEI. **Mineral Liberation Analyzer (MLA)**. Informações técnicas do *software* em <http://www.fei.com/products/sem/mla/>. Acessado em 07/10/2015.

GILLER, G. L. The Statistical Properties of Random Bitstreams and the Sampling Distribution of Cosine Similarity. **SSRN Journal**, [S.l.], 2012.

GOLDSTEIN, J. et al. **Scanning Electron Microscopy and X-ray Microanalysis**. 2.ed. [S.l.]: Plenum Press, 1992.

GOLUB, G. H.; LOAN, C. F. V. **Matrix Computations**. [S.l.]: Johns Hopkins University Press, 1985.

HALL, M. et al. The WEKA Data Mining Software: an update. **SIGKDD Explorations**, [S.l.], v.11, n.1, 2009.

HURLBUT, C. S. **Dana's Manual of Mineralogy**. [S.l.]: John Wiley and Sons, 1963.

ISO/TC 202. **ISO 22029:2012** / microbeam analysis - emsa/mas standard file format for spectral-data exchange. Disponível em [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56211](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56211). Acessado em 12/10/2015.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 6.ed. [S.l.]: Pearson, 2007.

KAUR, D. A Comparative Study of Various Distance Measures for Software fault prediction. In: INTERNATIONAL JOURNAL OF COMPUTER TRENDS AND TECHNOLOGY. **Anais...** [S.l.: s.n.], 2014. (3, v.17).

LAVINE, B.; DAVIDSON, C.; MOORES, A. Genetic algorithms for spectral pattern recognition. **Vibrational Spectroscopy**, [S.l.], v.28, n.1, p.83–95, Feb 2002.

LISBOA, J. C. F. Reconhecimento de Minerais. **Revista de Ensino de Ciências**, [S.l.], n.7, p.36–39, 1982.

MANSOURI, J.; KHADEMI, M. Multiplicative distance: a method to alleviate distance instability for high-dimensional data. **Knowledge and Information Systems**, [S.l.], p.1–23, 2014.

MCSWIGGEN, P. **Analytical Services**. Em <http://www.mcswiggen.com/Services.htm>. Acessado em 12/10/2015.

- MINERAL SPECTRA. **Mineral Energy Dispersive Spectra (EDS)**. <http://www.sfu.ca/~marshall/sem/spectra.html>. Acessado em 11/10/2015.
- NEWBURY, D. E. Mistakes encountered during automatic peak identification of minor and trace constituents in electron-excited energy dispersive X-ray microanalysis. **Scanning**, [S.l.], v.31, n.3, p.91, 2009.
- OWEN, M. **Mineral identification using mineral definitions having compositional ranges**. WO Patent App. PCT/US2013/062,648.
- PLATFORM, R. **Analog to Digital Conversion**. Em [http://www.robotplatform.com/knowledge/ADC/adc\\_tutorial.html](http://www.robotplatform.com/knowledge/ADC/adc_tutorial.html). Acessado em 12/10/2015.
- PRESS, W. H. et al. **Numerical Recipes in C - The Art of Scientific Programming**. 2.ed. [S.l.]: Cambridge University Press, 1992.
- REED, S. J. B. **Electron Microprobe Analysis and Scanning Electron Microscopy in Geology**. [S.l.]: Cambridge University Press, 1996.
- SMITH, D. G. W.; LEIBOVITZ, D. P. MinIdent: a data base for minerals and a computer program for their identification. **Canadian Mineralogist**, [S.l.], v.24, p.695–708, 1986.
- TORRES, R. D. S.; FALCÃO, A. X. Content-Based Image Retrieval: theory and applications. **Revista de Informática Teórica e Aplicada**, [S.l.], v.13, p.161–185, 2006.
- TUTORVISTA. **Atomic Spectra**. Em <http://chemistry.tutorvista.com/analytical-chemistry/atomic-spectra.html>. Acessado em 12/10/2015.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: practical machine learning tools and techniques**. 3.ed. [S.l.]: Morgan Kaufmann/Elsevier, 2011.
- XUE, T. et al. The application of text similarity computing in the Clinical Decision Support System. **2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems**, [S.l.], Nov 2014.