

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ONTO SOLAR FLARE - UMA ONTOLOGIA DE  
DOMÍNIO/APLICAÇÃO SOBRE EXPLOSÕES  
SOLARES SOB ESCOPO DO MÉTODO ECID**

**NATÁLIA DE SOUZA FRANÇA**

**ORIENTADOR: PROF<sup>A</sup>. DR<sup>A</sup>. MARCELA XAVIER RIBEIRO**

São Carlos – SP

Agosto, 2020

**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**ONTO SOLAR FLARE - UMA ONTOLOGIA DE  
DOMÍNIO/APLICAÇÃO SOBRE EXPLOSÕES  
SOLARES SOB ESCOPO DO MÉTODO ECID**

**NATÁLIA DE SOUZA FRANÇA**

Defesa apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Banco de Dados.

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Marcela Xavier Ribeiro

São Carlos – SP

Agosto, 2020

*Dedico aos meus pais Vilma e Norival, madrinha Maria Aparecida (em memória) e irmão Daniel que sempre me apoiam em todas as situações.*

## AGRADECIMENTOS

---

---

Agradeço imensamente os meus pais Vilma Fernandes de Souza França e Norival França, o meu irmão Daniel de Souza França e a madrinha Maria Aparecida Fernandes Mariano (em memória) por estarem ao meu lado e acreditarem no meu potencial.

Aos meus orientadores oficiais e não oficiais Prof<sup>a</sup> Dr<sup>a</sup> Marcela Xavier Ribeiro, Prof<sup>a</sup> Dr<sup>a</sup> Marilde Terezinha Prado Santos, Dr. José Roberto Cecatto e Prof. Dr. Sérgio Discola Júnior por me apoiarem e orientarem com toda paciência e profissionalismo necessário. Muito obrigada pela confiança e por aceitarem esse desafio junto comigo.

Aos amigos do Departamento de Computação (DC) em especial Laís Villioni e Silva, Raquel Cristina de Fátima, Juliana Wolf, Eduardo Prado e Marco Túlio dos Santos Costa pelos momentos de estudos e confraternização.

Agradeço ao Instituto Nacional de Pesquisas Espaciais (INPE) pela parceria, em especial ao pesquisador Dr. José Roberto Cecatto, que sempre teve disposição em ensinar e ajudar. Sem o seu conhecimento a mim transmitido esse projeto não seria possível.

Agradeço a Universidade Federal de São Carlos em especial ao programa PPGCC, e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) por me acolher e viabilizar a minha pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Código de Financiamento 001”

Agradeço imensamente!

*"A educação tem raízes amargas, mas os seus frutos são doces."  
(Aristóteles)*

# RESUMO

Explosões solares são fenômenos capazes de causar diversos danos em sistemas de rádio, satélites de comunicação, sistemas de posicionamento global (GPS), dentre outros. Os possíveis danos podem variar de acordo com a intensidade da explosão, e estas explosões são classificadas como A, B, C, M e X. Prever esses fenômenos não é uma atividade fácil pois os dados são esparsos e não se sabe quais características deste fenômeno são relevantes. Neste trabalho buscou-se melhorar o processo de previsão das ocorrência das explosões solares a partir de conhecimentos sobre o domínio do ciclo solar estruturado em uma ontologia de domínio/aplicação. Aqui abstraiu-se a complexidade dos dados criando-se uma ontologia, a fim de melhorar o processo de predição. Para a aplicação da ontologia, optou-se por incorpora-la ao método de previsão de explosões solares ECID, que realiza previsão multi-class e multi-label com tratamento para o desbalanceamento de dados. Com o intuito de contribuir com as pesquisas sobre explosões solares, o ECID foi adaptado para englobar a ontologia Onto Solar Flare. A ontologia foi desenvolvida para enriquecer semanticamente o algoritmo ECID, a junção dos dois ferramentais gerou o ECID Onto. Os resultados mostram uma melhor distribuição de classes, facilitando a tomada de decisão final do especialista de domínio.

**Palavras-chaves:** Ontologia, Explosões Solares, ECID, Previsão.

# ABSTRACT

Solar flares are phenomena capable of causing several damages in radio systems, communication satellites, global positioning systems (GPS), among others. The possible damage can vary according to the intensity of the explosion and these explosions are classified as A, B, C, M and X. Predicting this phenomenon is not an easy task due the sparse data and the unknowledge of relevante characteristics of the phenomenon. In this work we aim to improve the process of prediction of solar flares building the knowledge about the domain of the solar cycle structured in a domain ontology. The complexity of the data was abstracted by creating a domain Ontology in order to improve the prediction process. To apply the ontology, it was decided to incorporate it to the ECID solar explosion prediction method, which performs multi-class and multi-label prediction with treatment for imbalanced data. In order to contribute to research on solar flares, ECID was adapted to encompass the Onto Solar Flare ontology. The ontology was developed to semantically enrich the ECID algorithm, the combination of the two tools generated the ECID Onto. The results show a better distribution of classes, facilitating the final decision of the domain specialist.

**Key-words:** Ontology, Solar Flares, ECID, Forecasting.

# ACRÔNIMOS

**Cmap** *Concept Map*

**ECID Onto** *Classifier Set for Unbalanced Data Sets With Ontologies*

**ECID** *Ensemble of Classifiers for Imbalanced Datasets*

**FN** False Negative

**GOES** *Geostationary Operational Environmental Satellite*

**GPS** Sistema de Posicionamento Global

**GR** Glossário de Referência

**HELIO** *Heliophysics Integrated Observatory*

**INPE** Instituto Nacional de Pesquisas Espaciais

**LA** Léxico de Aplicação

**LD** Léxico de Domínio

**NOAA** *National Oceanic and Atmospheric Administration*

**OIL** *Ontology Inference Layer*

**OWL** *Web Ontology Language*

**QC** Questão de Competência

**RA** Região Ativa

**RDF** *Resource Description Framework*

**SADT** *Structured Analysis & Design Technique*

**SDO/HMI** *Solar Dynamics Observatory/Helioseismic and Magnetic Imager*

**SPARQL** *SPARQL Query Language for RDF*

**SWRL** *Semantical Web Rule Language*

**TOVE** *Toronto Virtual Enterprise*

**TPR** *True Positive Rate*

**UML** *Unified Modeling Language*

**UPON** *Unified Process for Ontology*

**W3C** *World Wide Web Consortium*

**XML** *Extensible Markup Language*

## LISTA DE FIGURAS

---

---

Figura 1 – Tipos de ontologias. . . . .	18
Figura 2 – Ciclo de vida de uma ontologia. . . . .	20
Figura 3 – Visão geral da metodologia UPON. . . . .	21
Figura 4 – Visão geral do ECID. . . . .	22
Figura 5 – Distribuição dos dados na etapa 4.a do ECID. . . . .	23
Figura 6 – Método Agregador do ECID - etapa 4.c . . . . .	24
Figura 7 – Extrato da ontologia <i>Space Weather Ontology</i> . . . . .	26
Figura 8 – Extrato da ontologia <i>VSTO</i> . . . . .	27
Figura 9 – Diagrama SADT das etapas de um fluxo de trabalho proposto pela UPON . . . . .	31
Figura 10 – Representação gráfica das primeiras informações coletadas com o especialista de domínio sobre explosões solares . . . . .	35
Figura 11 – Extrato da ontologia <i>Onto Solar Flare</i> que representa um relacionamento is a . . . . .	37
Figura 12 – Extrato do código em OWL-DI da ontologia <i>Onto Solar Flare</i> . . . . .	38
Figura 13 – Ontologia <i>Onto Solar Flare</i> - Versão 1 . . . . .	40
Figura 14 – Ontologia <i>Onto Solar Flare</i> - Versão 2 . . . . .	41
Figura 15 – Arquivo da base de dados NOAA para coleta de dados da classificação de raio-x. Dados referente ao dia 2012 Mar 05 . . . . .	42
Figura 16 – Arquivo da base de dados NOAA para coleta de dados das classificações <i>McIntosh</i> e <i>Mount Wilson</i> . Dados referente ao dia 2012 Mar 05 . . . . .	43
Figura 17 – Visão geral do método ECID Onto - Versão 1 . . . . .	45
Figura 18 – Experimentos do ECID e ECID Onto - Versão 1 . . . . .	47
Figura 19 – Distribuição das classes previstas pelo ECID e pelo ECID Onto - Versão 1 . . . . .	48
Figura 20 – Visão geral do método ECID Onto - Versão 2 . . . . .	48
Figura 21 – Experimentos do ECID, ECID Onto - Versão 1 e Versão 2 . . . . .	50
Figura 22 – Distribuição das classes previstas pelo ECID Onto - Versão 2 . . . . .	50

## LISTA DE TABELAS

---

---

Tabela 1 – Intervalo de tempo entre as etapas do ciclo 24 . . . . .	40
Tabela 2 – Proporções das classes de explosões solares por etapa do ciclo, calculadas a partir dos dados de trinamento do ECID . . . . .	41

# SUMÁRIO

---

---

<b>CAPÍTULO 1–INTRODUÇÃO</b> . . . . .	<b>12</b>
1.1 Contexto e motivação . . . . .	12
1.2 Objetivos, hipótese e justificativa . . . . .	13
1.3 Síntese da metodologia e dos resultados . . . . .	13
1.4 Organização do trabalho . . . . .	14
<b>CAPÍTULO 2–REFERENCIAL TEÓRICO</b> . . . . .	<b>16</b>
2.1 Explosões solares . . . . .	16
2.1.1 Tipos de ontologias . . . . .	17
2.1.2 Representação de uma ontologia . . . . .	18
2.2 Engenharia de ontologias . . . . .	19
2.2.1 <i>Unified Process for Ontology (UPON)</i> . . . . .	21
2.3 O método ECID . . . . .	22
2.4 Considerações Finais . . . . .	24
<b>CAPÍTULO 3–TRABALHOS CORRELATOS</b> . . . . .	<b>25</b>
3.1 <i>Concept Maps for a Space Weather Ontology</i> . . . . .	25
3.2 <i>HELIO: Discovery and analysis of data in heliophysics</i> . . . . .	26
3.3 <i>Ontology-supported scientific data frameworks: The Virtual Solar-Terrestrial         Observatory experience</i> . . . . .	26
3.4 Considerações Finais . . . . .	27
<b>CAPÍTULO 4–<i>Onto Solar Flare</i></b> . . . . .	<b>29</b>
4.1 Questão de pesquisa . . . . .	29
4.2 Engenharia de ontologia . . . . .	30
4.3 Desenvolvimento da ontologia . . . . .	32
4.3.1 Levantamento de requisitos . . . . .	33
4.3.2 Construção dos glossários . . . . .	35
4.3.3 Organização das informações . . . . .	36
4.3.4 Implementação . . . . .	36
4.3.5 Testes . . . . .	37
4.4 Descrição da ontologia . . . . .	39
4.5 <i>Onto Solar Flare - Versão 1</i> . . . . .	39
4.6 <i>Onto Solar Flare - Versão 2</i> . . . . .	41
4.7 Considerações Finais . . . . .	43

<b>CAPÍTULO 5–ECID ONTO</b> .....	<b>44</b>
5.1 Visão Geral do ECID Onto .....	44
5.2 ECID Onto - Versão 1 .....	45
5.2.1 Desenvolvimento .....	45
5.2.2 Validação .....	46
5.3 ECID Onto - Versão 2 .....	48
5.3.1 Desenvolvimento .....	48
5.3.2 Validação .....	49
5.4 Considerações Finais .....	50
<b>CAPÍTULO 6–CONCLUSÃO</b> .....	<b>52</b>
<b>REFERÊNCIAS</b> .....	<b>54</b>

# Capítulo 1

## INTRODUÇÃO

---

---

*O presente capítulo visa introduzir o contexto e motivação deste trabalho de mestrado (seção 1.1), além de apresentar os objetivos, a hipótese e justificativa (seção 1.2), uma síntese dos resultados (seção 1.3) e a organização deste texto (seção 1.4).*

### 1.1 Contexto e motivação

Explosão solar é um fenômeno responsável pela liberação repentina de uma grande quantidade de energia ( $10^{24}$  -  $10^{32}$  erg) de um Região Ativa (RA) presente na atmosfera solar. Esse fenômeno geralmente se origina de um episódio de reconexão magnética com a consequente liberação de energia acumulada na atmosfera solar, mais especificamente na cromosfera/corona (HOLMAN, 2006). Dependendo da intensidade, o fenômeno pode causar efeitos diretos ou indiretos nas atividades humanas. Entre os efeitos, podemos citar distúrbios ou danos em sistemas de rádio, satélites de comunicação, Sistema de Posicionamento Global (GPS) e também, em casos extremos, *blackouts* de radiocomunicação e energia elétrica (BASU et al., 2010) (SCHRIJVER et al., 2015).

Com o intuito de minimizar os danos, diversas pesquisas multidisciplinares de caráter preditivos estão sendo desenvolvidas, porém, não se sabe muito a respeito do processo que causa a reconexão magnética e a consequente liberação de energia em uma RA (BENZ, 2017). Também trata-se de um domínio complexo com grande diversidade de dados e terminologias que não são triviais para todos os pesquisadores envolvidos. Por esses e outros motivos o processo de previsão é complexo. Segundo Noy e McGuinness (2001), dados estruturados em ontologia podem ser um ferramental capaz de auxiliar a padronização de termos auxiliando os pesquisadores multidisciplinares, além de facilitar a reutilização do conhecimento, realizar inferências sobre o domínio e tornar as informações estruturadas mais visíveis.

Motivado em contribuir com as pesquisas de previsão de explosões solares, a presente dissertação demonstra os resultados obtidos com o desenvolvimento de uma ontologia de domínio sobre explosões solares, denominado *Onto Solar Flare*. A ontologia foi desenvolvida com o intuito de extrair e estruturar os dados facilitando a compreensão dos pesquisadores.

Para a aplicação da ontologia, optou-se por incorporá-la ao método de predição de explosões *Ensemble of Classifiers for Imbalanced Datasets* (ECID), pois esse método é fruto de pesquisas realizadas no mesmo departamento em que este trabalho foi desenvolvido e durante o desenvolvimento do ECID se deparou com os problemas descritos anteriormente como dificuldade dos pesquisadores em padronizar algumas definições e identificar quais informações são relevantes para o processo de previsão, por isso surgiu a necessidade da ontologia. Trabalhar com pesquisadores do mesmo grupo auxiliou na troca de informações e uma consequente otimização do processo de desenvolvimento das ontologias. A junção da ontologia *Onto Solar Flare* ao método ECID gerou o método *Classifier Set for Unbalanced Data Sets With Ontologies* (ECID Onto).

## 1.2 Objetivos, hipótese e justificativa

O objetivo geral desse trabalho é estruturar e aplicar uma ontologia de domínio sobre explosões solares, a partir de informações coletadas na literatura e com especialistas de domínio.

Os objetivos específicos são:

- Criar uma base de conhecimentos sobre explosões solares a partir da literatura e especialistas de domínio;
- Criar uma base de dados para dar suporte a predição;
- Adaptar o método de predição ECID para incorporar a ontologia *Onto Solar Flare*.

A hipótese é de que é possível melhorar o processo de predição das explosões solares a partir de conhecimentos sobre o domínio do ciclo solar estruturado em uma ontologia de domínio. O trabalho justifica-se pela necessidade de abstrair a complexidade dos dados a fim de melhorar o processo de predição e ontologia é um ferramental capaz de suprir essa necessidade.

## 1.3 Síntese da metodologia e dos resultados

Para o desenvolvimento da ontologia *Onto Solar Flare*, utilizou-se a metodologia *Unified Process for Ontology* (UPON) descrita em Nicola et al. (2009). Ela propõe desenvolver a ontologia objetivando sua aplicação, então, este projeto conta com duas metodologias, a de desenvolvimento da ontologia e da sua incorporação ao método ECID.

A UPON simula um processo de desenvolvimento de software, que se baseia no modelo iterativo e incremental facilitando a exploração do domínio. O projeto foi dividido em quatro grandes iterações que geraram incrementos representativos. As duas primeiras iterações foram responsáveis pela exploração e coleta de informações sobre o domínio de explosões solares e sobre o domínio de aplicação (ECID), para seu desenvolvimento foram realizados levantamentos

bibliográficos e entrevistas com os especialistas de domínio. Nas duas últimas foram desenvolvidas duas versões da ontologia e por consequência duas versões do [ECID Onto](#), e para seu desenvolvimento, as informações coletadas nas entrevistas foram catalogadas, modeladas, implementadas e validadas seguindo a proposta da metodologia, em seguida foram implementadas no algoritmo de aplicação.

O processo de exploração do domínio sobre explosões solares e seleção das informações a serem representadas da ontologia foi um processo complexo, pois trata-se de um domínio com grande diversidade de dados e ao mesmo tempo não se sabe quais informações representam explicitamente a ocorrência futura do fenômeno. A parceria com o especialista sobre explosões solares e pesquisador do Instituto Nacional de Pesquisas Espaciais ([INPE](#)) foi de suma importância para a compreensão do domínio e identificar quais informações são relevantes para esse projeto.

A primeira versão da ontologia objetivou o estudo de viabilidade para o desenvolvimento do projeto. Nela foram representados os dados do último ciclo solar e classificação de raio-x. A ontologia foi incorporada ao método agregador do [ECID](#), onde é instanciada, retornando as proporções de ocorrência de raio-x. Como resultado final melhorou a distribuição das classes preditas.

Na segunda versão da ontologia foram representadas as demais informações propostas no escopo da ontologia (classificações *McIntosh* e *Mount Wilson*). Ela foi incorporada no mesmo método agregador da versão anterior, porém utilizou-se dados diferentes da primeira versão. Nessa versão optou-se por utilizar todos os dados das classificações de raio-x, *McIntosh* e *Mount Wilson* do último ciclo solar (ciclo 24).

## 1.4 Organização do trabalho

O restante desta dissertação está estruturada da seguinte forma:

**Capítulo 2:** apresenta os conceitos teóricos necessários para o desenvolvimento do projeto. São abordados assuntos como: explosões solares com ênfase nas informações que foram representadas na ontologia, definições e conceitos sobre ontologias, engenharia de ontologia (área que estuda teorias e técnicas para a construção de ontologias) e, por fim, método [ECID](#) definido como domínio de aplicação;

**Capítulo 3:** apresenta alguns trabalhos correlatos que contém ontologias relacionadas ao domínio solar.

**Capítulo 4:** apresenta a questão de pesquisa, a o desenvolvimento da ontologia *Onto Solar Flare* e resultados obtidos no desenvolvimento da mesma;

**Capítulo 5:** apresenta o desenvolvimento e resultados obtidos no método [ECID Onto](#). Por fim, uma comparação do [ECID](#) com o [ECID Onto](#);

**Capítulo 6:** detalha as contribuições geradas e os trabalhos futuros.

# Capítulo 2

## REFERENCIAL TEÓRICO

---

---

*O presente capítulo visa apresentar informações sobre o referencial teórico necessário para essa dissertação. Na seção 2.1 é apresentado as informações sobre explosões solares que foram representados na ontologia. A seção ?? contém conceitos importantes sobre ontologia, a seção 2.2 contem informações sobre engenharia de ontologia (área que estuda teorias e técnicas para a construção de ontologias) e, por fim, informações sobre o método ECID, que foi utilizado para validar a ontologia.*

### 2.1 Explosões solares

Informação, Engenharia de Software e Interação Humano-Computador. A seguir são apresentadas algumas definições encontradas na literatura.

Segundo [Isotani e Bittencourt \(2015\)](#), ontologia pode ser definida como um conjunto de conceitos e suas relações, que refletem como as pessoas interpretam um determinado domínio e permite a representação formal de tal interpretação, tornando-a compreensível por humanos e computadores.

[Gruber \(1995\)](#) afirma que ontologia é "uma especificação formal e explícita de uma conceitualização compartilhada". Para explicar essa afirmação, [Fensel \(2001\) apud \[Hernandes \\(2009\\)\]\(#\)](#) a descreveu em partes:

- **Formal:** a ontologia deve ser desenvolvida de forma a evitar ambiguidades e ser legível a computadores;
- **Explícita:** conceitos, propriedades, relações, restrições e axiomas devem ser bem claras/explicitas;
- **Conceitualização:** ontologias apresentam uma visão abstrata e simplificada de como as pessoas interpretam um domínio;
- **Compartilhada:** a ontologia visa captar um conhecimento consensual, pois ela pode ser utilizada por diferentes aplicações.

Segundo [Kiryakon \(2006\)](#), ontologias são compostas por quatro elementos básicos que podem ser representados pelo conjunto  $O = \{C, R, I, A\}$ , onde:

- C se refere ao conjunto de classes e subclasses que representam os conceitos.
- R se refere ao conjunto de relações entre as classes.
- I se refere ao conjunto de instâncias ou atributos que caracterizam as classes.
- A se refere ao conjunto de axiomas que modelam regras e restrições. As regras permitem validar a estrutura da ontologia e inferir novos conhecimentos a partir de mecanismos de inferência.

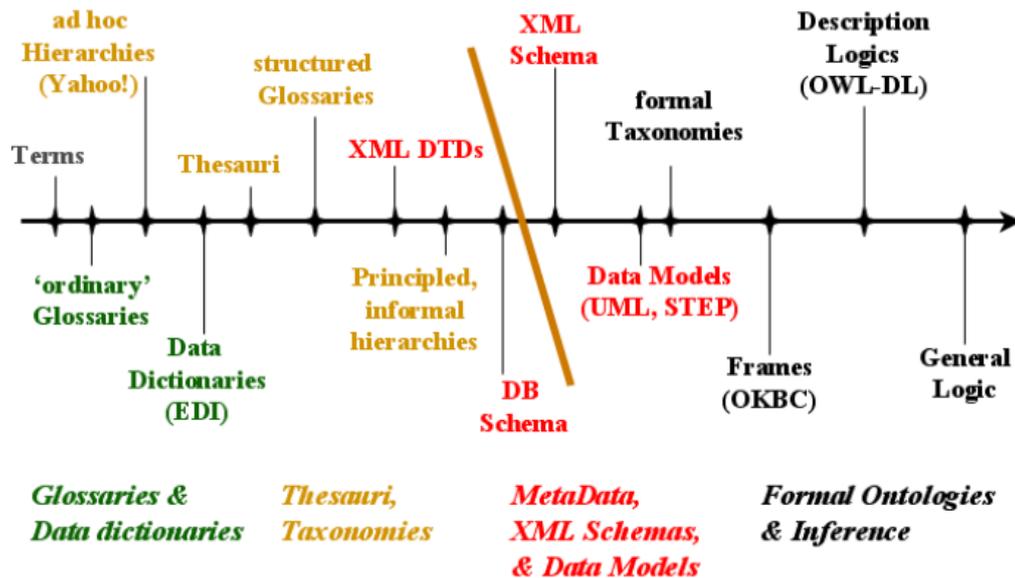
### 2.1.1 Tipos de ontologias

Tipos de ontologias Existem diferentes tipos de ontologias, elas se distinguem conforme a necessidade de especificar o significado de cada conteúdo representado. [Mizoguchi \(2004\)](#) e [Isotani e Bittencourt \(2015\)](#) as dividem em leves (*lightweight ontology*) ou pesadas (*heavyweight ontology*) e ontologias de domínio ou tarefa.

As *lightweight ontology* são uma taxonomia que representam as hierarquias entre os conteúdos. Não é necessário definir com detalhes cada conceito, essa ontologia geralmente é utilizada para categorizar um grande volume de dados ([BIZER, 2009](#)) *apud* ([ISOTANI; BITTENCOURT, 2015](#)). Já as *heavyweight ontology* definem de maneira detalhada um determinado domínio, fornecendo não apenas a taxonomia, mas também a semântica entre os conceitos. Seu desenvolvimento requer uma descrição de cada conceito e sua organização, além de uma definição formal da semântica entre os conceitos e suas relações ([GOMEZ et al., 2003](#)).

Na Figura 1 é apresentado um espectro evolutivo das ontologias, do lado esquerdo são as *lightweight ontology* e do lado direito as *heavyweight ontology*, ou seja, a medida que se avança para a esquerda no espectro, maior será o nível de detalhamento e formalismo da ontologia e menor o nível de abstração.

Figura 1 – Tipos de ontologias.



Fonte: (USCHOLD; GRÜNINGER, 2004).

Isotani e Bittencourt (2015) apresenta dois tipos de ontologias: de domínio e de tarefas. Ontologias de tarefas representam a capacidade de aplicar um conhecimento para resolver um determinado problema, ou seja, descreve os processos ou tarefas para atingir um fim. Ontologias de domínio representam um conhecimento de senso comum, de forma geral ela caracteriza o domínio que as tarefas são executadas. Neste trabalho foi desenvolvido uma ontologia de domínio para representar parte do conhecimento sobre explosões solares.

### 2.1.2 Representação de uma ontologia

Ontologias podem ser representadas de duas maneiras: formal e gráfica. A representação gráfica é utilizada para a interpretação humana, podendo ser visualizada por grafos, *Unified Modeling Language (UML)*, estruturas de árvores, entre outros. Já a representação formal é necessária para que os computadores interpretem as ontologias, utilizando algumas linguagens específicas baseadas em lógica de predicados, lógica descritiva ou linguagens baseadas em *frames* (ISOTANI; BITTENCOURT, 2015).

Horrocks et al. (2003) menciona duas linguagens desenvolvidas diretamente para a representação de ontologias, *Resource Description Framework (RDF)* e *Web Ontology Language (OWL)*, ambas são indicadas pela *World Wide Web Consortium (W3C)* para descrever metadados. Patel-Schneider (2005) afirma que o modelo *RDF* utiliza a estrutura de triplas para relacionar os dados possibilitando montar uma estrutura hierárquica (taxonomia), fornecendo assim os recursos básicos necessários para representar uma ontologia.

Triplas ou grafo *RDF* estabelecem relações ou interligações entre dados para diferentes propósitos. São compostas por:

*<Sujeito> <Predicado> <Objeto>*

em que *<Sujeito>* é um recurso no qual a semântica está se referindo. *<Predicado>* descreve uma característica, propriedade ou relacionamento utilizado para descrever o recurso. *<Objeto>* é o valor de uma determinada característica do recurso, podendo ser também um recurso (CYGANIAK et al., 2014).

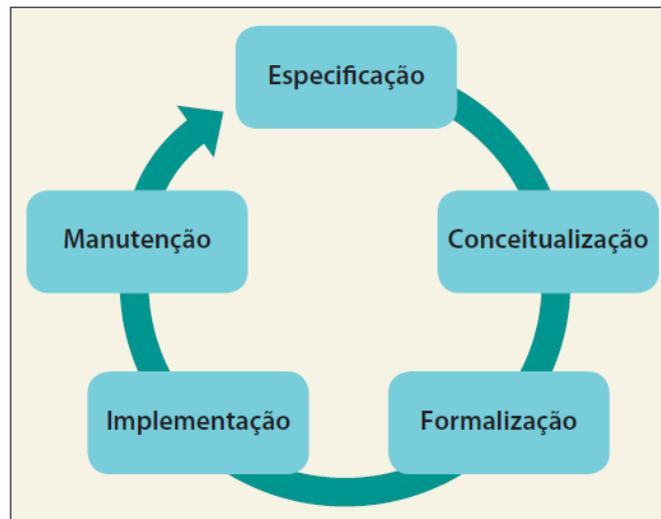
A linguagem **OWL** é considerada mais expressiva e complexa que o **RDF**, pois permite estabelecer igualdade, desigualdade e intersecção entre as classes, características e restrições de propriedades, além de determinar restrições de cardinalidades (MCGUINNESS; HARMELEN, 2004). Considerada a linguagem de ontologias da Web, **OWL** pode ser dividida em três sublinguagens que variam de acordo com a sua escalabilidade e expressividade (ISOTANI; BITTENCOURT, 2015), elas estão listadas abaixo. A ontologia desenvolvida nesse projeto é do tipo **OWL DL**.

- **OWL Lite:** é a sub-linguagem menos expressiva, sintaticamente simples destinada a usuários que necessitam apenas de uma classificação hierárquica com restrições;
- **OWL DL:** é a sub-linguagem mais expressiva que **OWL Lite**, baseada na lógica de descrição, é passível de raciocínio automático, ou seja, é possível realizar conclusões por sistemas computacionais em um tempo finito;
- **OWL Full:** é a sub-linguagem mais expressiva. Devido ao seu grande nível de detalhamento não foi possível desenvolver máquinas de inferências capazes de tratar todos os dados.

## 2.2 Engenharia de ontologias

Engenharia de Ontologias é uma evolução da área conhecida como Engenharia de Conhecimento, que estuda teorias e técnicas adequadas para desenvolver bases de conhecimentos utilizadas nos sistemas especialistas. Engenharia de Ontologias tem como objetivo auxiliar o processo de criação de ontologias adequadas para representar as suas informações (ISOTANI; BITTENCOURT, 2015). Esse processo envolve desafios arquiteturais e metodológicos. Para lidar com estes desafios, é necessário criar metodologias capazes de guiar o desenvolvimento e sua manutenção, porém, muitos projetos de pesquisa ainda realizam esse processo de maneira *ad-hoc*.

O processo de desenvolvimento de uma ontologia possui um ciclo de vida, as etapas desse ciclo estabelecem um conjunto de passos a serem seguidos, podendo variar de acordo com a metodologia proposta (CALERO et al., 2006). Isotani e Bittencourt (2015) estabelecem um ciclo de vida genérico, ilustrado na Figura 2 e os tópicos abaixo descrevem cada etapa.

**Figura 2 – Ciclo de vida de uma ontologia.**

Fonte: (ISOTANI; BITTENCOURT, 2015).

- **Especificação:** etapa de criação de documentos em linguagem natural, que contém o principal objetivo da ontologia e os níveis de granularidade dos dados;
- **Conceitualização:** essa etapa visa à organização do conhecimento não estruturado adquirido até o momento. Esse conhecimento é convertido em uma estrutura semi-formal, para que os especialistas de domínio e desenvolvedores da ontologia possam compreender;
- **Formalização:** etapa em que é identificado e estabelecido os componentes como classes, relações, axiomas e instâncias;
- **Implementação/Desenvolvimento:** fase de implementação da ontologia, requer ambientes que propiciam suporte às características das meta-ontologias. O resultado dessa etapa é uma parte da ontologia desenvolvida em alguma linguagem, como a [OWL](#);
- **Manutenção/Validação:** etapa destinada a verificar se o resultado desenvolvido está de acordo com o que foi proposto na fase de especificação.

[BREITMAN \(2005\)](#) afirma que existem diferentes metodologias capazes de auxiliar o processo de desenvolvimento de ontologias, porém não existe uma única metodologia indicada, pois varia de acordo com a necessidade do projeto. O mesmo autor lista algumas metodologias: Projeto Cyc, Proposta de Uschold, Projeto *Toronto Virtual Enterprise (TOVE)*, Projeto Kactus, Metodologia 101 e Methontology. Para o desenvolvimento deste projeto, utilizou-se a metodologia [UPON](#), descrita na seção [2.2.1](#).

[Alatrish \(2013\)](#) cita alguns editores de ontologia que auxiliam no processo de implementação, são eles: Apollo, Hozo, OntoStudio, Protégé, e TopBraid Composer. O Protégé é um editor *open source* que possui recursos como: suporte para extensão de funcionalidades por *plug-ins* possibilitando ao usuário adaptar a ferramenta conforme sua necessidade e importar/exportar

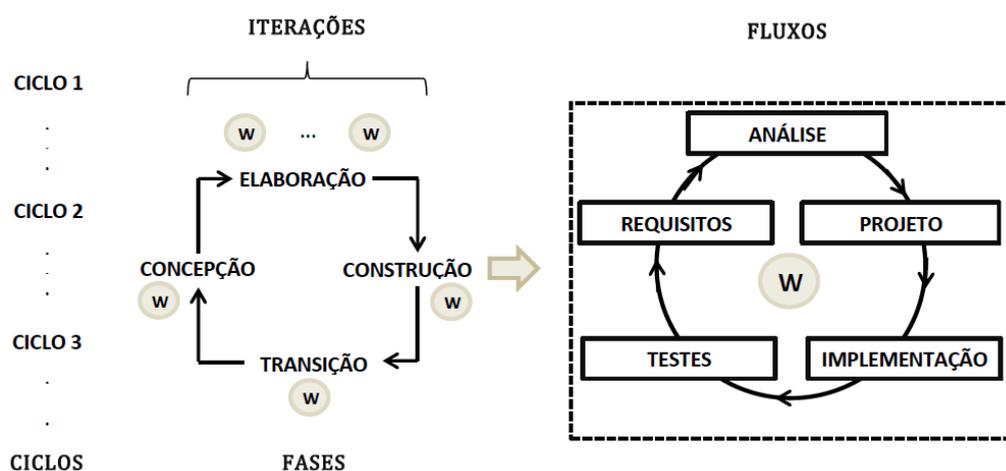
a ontologia para diversas linguagens como **OWL**, *Ontology Inference Layer (OIL)*, *Extensible Markup Language (XML)* e **RDF**. O Protégé<sup>1</sup> **OWL** disponibiliza a integração com raciocinadores externos, permitindo inferências sobre classes e indivíduos de uma ontologia e suporta o processo de alinhamento, técnica capaz de unir o conhecimento contido em diversas ontologias (ALATRISH, 2013).

### 2.2.1 Unified Process for Ontology (UPON)

A metodologia *Unified Process for ONtology (UPON)* foi proposta por Nicola et al. (2009). Sua construção foi baseada em um processo amplamente utilizado na engenharia de *software* intitulado Processo Unificado de Desenvolvimento de Software (ou Processo Unificado) aliado às técnicas de modelagem **UML**.

Trata-se de um método incremental com áreas de aplicação e objetivos bem definidos. A Figura 3 ilustra o método **UPON** que contém ciclos, fases, iterações e fluxos de trabalho. Em cada ciclo é realizado iterações, e cada interação possui quatro fases (concepção, elaboração, construção e transição). A fase é composta por um fluxo com cinco etapas: requerimentos, análise, projeto, implementação e testes. Note que na Figura 3, por questão de espaço, o fluxo de trabalho foi expandido apenas na fase de construção. Geralmente as fases iniciais são focadas em identificação do domínio de interesse, escopo da ontologia, etc. As fases finais aprimoram a versão final da ontologia produzida no ciclo (NICOLA et al., 2009).

**Figura 3 – Visão geral da metodologia UPON.**



Fonte: Nicola et al. (2009) adaptada por Baldárrago (2012).

Os objetivos do **UPON** são:

- Redução do tempo e custos na produção de grandes ontologias e orientações úteis para pequenas ontologias;

<sup>1</sup> Site oficial do Protégé - <https://protege.stanford.edu/community.php>

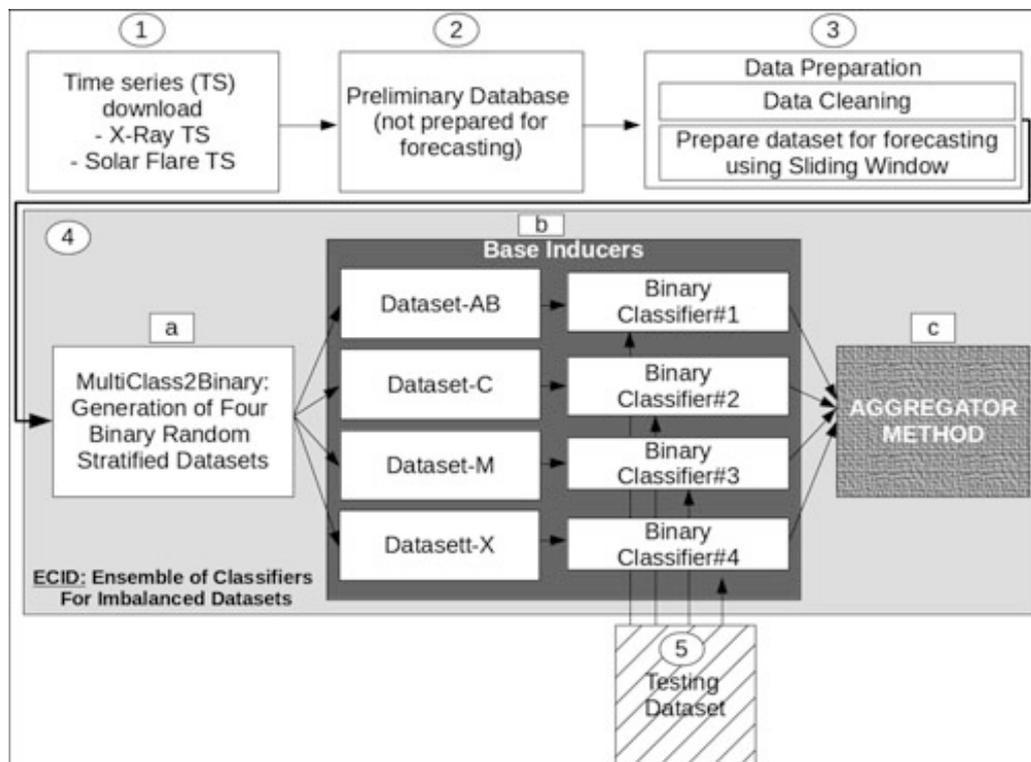
- Melhora na qualidade da ontologia, pois valida os resultados de forma progressiva, ou seja, aplica a validação em cada ciclo (iteração);
- Identificação clara da função de cada especialista;
- Disponibilização parcial da ontologia para os usuários.

## 2.3 O método ECID

O método **ECID** (*Ensemble of Classifiers for Imbalanced Datasets*) foi desenvolvido com o intuito de realizar previsões multiclasse e multilabel das ocorrências de explosões solares. A seguir, o método ECID é detalhado. Mais informações sobre o método ECID podem ser encontradas em [Junior et al. \(2019\)](#).

O ECID usa duas séries temporais de recursos solares: as séries temporais de intensidade de raios-x emitidas pelo Sol, com períodos de amostragem de 1 ou 5 minutos, e as séries temporais de recursos magnéticos fornecidos pelo instrumento *Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI)* com períodos de amostragem de 12 minuto. A Figura 4 demonstra uma visão geral do ECID.

Figura 4 – Visão geral do ECID.



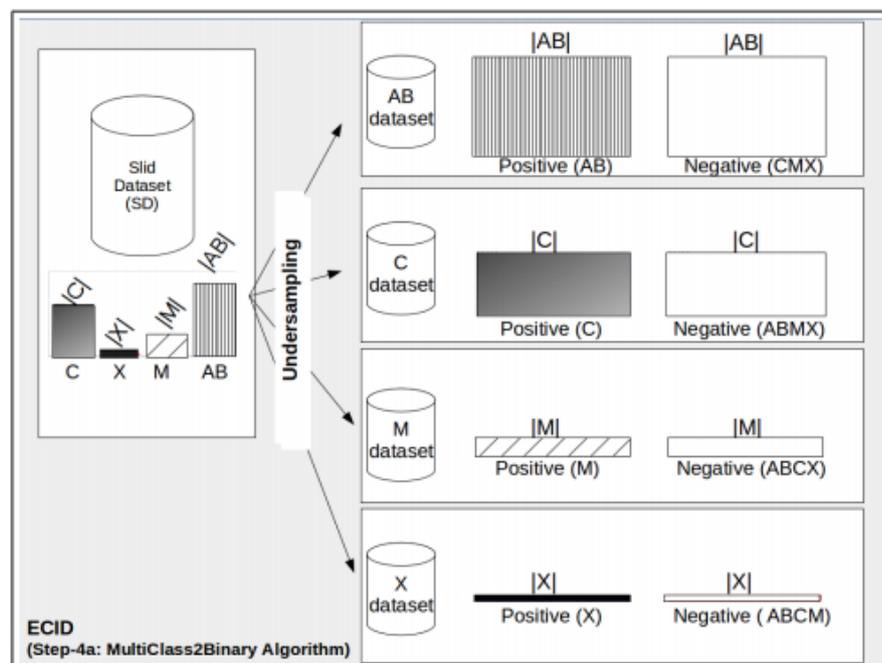
Fonte: ([JUNIOR et al., 2019](#)).

As etapas 1, 2 e 3 são as etapas de pré-processamento, nas quais é realizada a obtenção, limpeza e transformação dos dados. O método utiliza a estratégia de combinar medidas que

antecedem aos seus eventos. Isso permite gerar registros para utilizar classificadores tradicionais como método de previsão. Para gerar os registros é empregado o algoritmo *Sliding Window* proposto em [Junior et al. \(2018\)](#). Após o pré-processamento, os dados são divididos em 70% para treinamento e 30% para teste, na base de treinamento é aplicado o método ECID que gera quatro subamostras do conjunto de dados usando uma estratégia denominada *MultiClass2Binary* para a construção vários modelos de aprendizado (Etapa 4).

A estratégia *MultiClass2Binary* (etapa 4.a) gera quatro conjuntos de dados distintos balanceados para as classes das explosões solares (AB, C, M e X) usando subamostras estratificadas (ver Figura 5). O primeiro conjunto de dados é composto por registros de 50% das tuplas classes A ou B e 50% das demais, sendo as classes AB classificadas como positivas e as demais como negativas, o mesmo se repete para as classes C, M e X. Na etapa 4.b, em cada conjunto de dados é aplicado um método de classificação binária (ou indutor) do mesmo tipo, por exemplo árvore de decisão. Assim o classificador binário n.1 prediz a ocorrência da classe AB, o classificador binário n.2, n.3 e n.4 prediz a ocorrência das classes C, M e X respectivamente.

**Figura 5 – Distribuição dos dados na etapa 4.a do ECID.**

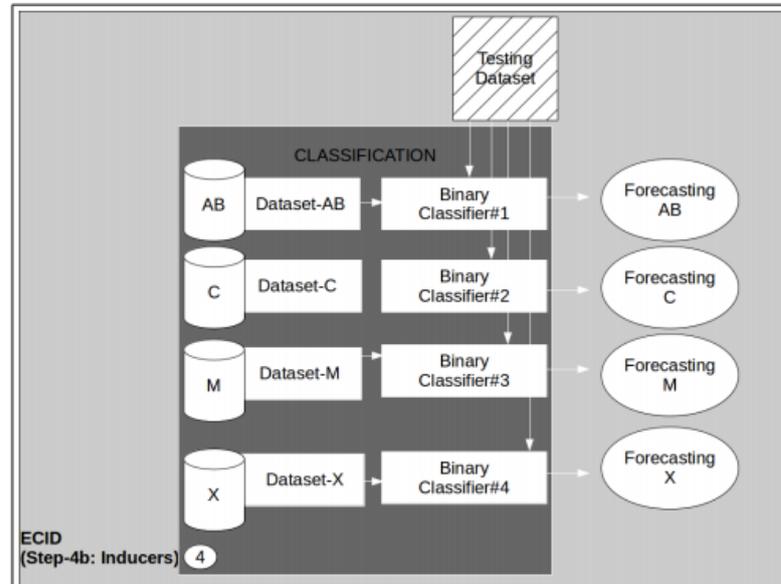


Fonte: ([JUNIOR et al., 2019](#)).

Como o maior período de amostragem dos dados é de 12 minutos, a etapa 4.b retorna 120 previsões por dia por classe, então o Método Agregador (etapa 4.c) agrega todas as previsões em uma final. A Figura 6 ilustra o Método Agregador, ele utiliza um sistema de votos contabilizando o número de previsões diárias de cada classificador binário para as classes. Por fim, é aplicado um sistema de parâmetros, onde até as 3 classes mais votadas podem ser retornadas como previsão final.

Neste projeto, a ontologia desenvolvida foi incorporada ao modelo agregador. No Capítulo 5 é descrito esse processo.

**Figura 6 – Método Agregador do ECID - etapa 4.c**



Fonte: (JUNIOR et al., 2019).

## 2.4 Considerações Finais

Neste capítulo foram apresentados conceitos básicos para o desenvolvimento do projeto. Inicialmente foi apresentada definições sobre explosões solares que foram representadas nas ontologia. São informações sobre ciclo solar, classificações magnéticas e classificações de raio-x. O ciclo solar tem duração média de 11 anos, ao longe desse período a quantidade visível de RA no globo solar varia de acordo com as etapas do ciclo. RA são manchas escuras visíveis no Sol que possuem campos magnéticos bipolares/multipolares. As explosões solares em sua maioria, são originadas em reconexões desses campos magnéticos. A complexidade da RA e de seus campos magnéticos está diretamente relacionada com a intensidade da explosão solar. Para parametrizar a complexidade das RAs são utilizadas as classificações magnéticas *McIntosh* e *Mount Wilson*, e para parametrizar a intensidade das explosões solares é utilizado a classificação de Raio-X.

Foi descrito o que é ontologia e engenharia de ontologias. Ontologias são conceitos organizados em uma hierarquia, com relacionamentos capazes de representar a interpretação de um conjunto de pessoas sobre um domínio. Já a engenharia de ontologias é uma área que estuda o desenvolvimento de ontologia e metodologias que facilitam esse desenvolvimento. Uma das metodologias é a UPON, utilizada para guiar parte do desenvolvimento deste projeto.

Por fim foi apresentado o método ECID, selecionado para este trabalho como forma de aplicação e validação da ontologia desenvolvida *Onto Solar Flare*.

# Capítulo 3

## TRABALHOS CORRELATOS

---

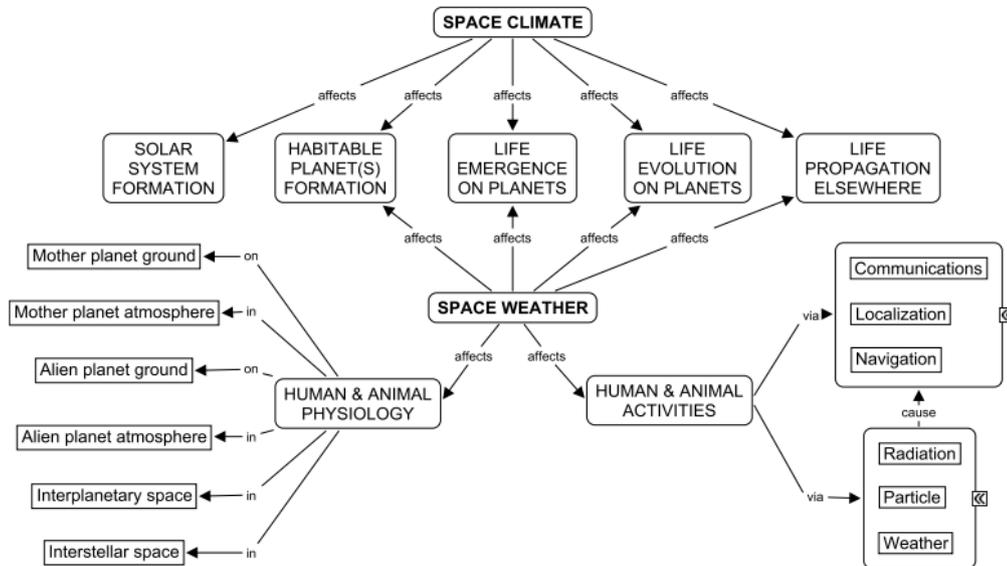
---

*No presente Capítulo, será apresentado alguns trabalhos correlatos a essa dissertação. Devido a dificuldade de encontrar trabalhos que relacionam diretamente explosões solares a ontologia, optou-se por pesquisar de forma ad-hoc, trabalhos que relacionam o domínio solar e ontologias.*

### **3.1 *Concept Maps for a Space Weather Ontology***

No artigo [Messerotti \(2007\)](#) é apresentado uma versão preliminar de uma ontologia sobre clima espacial. A motivação para o desenvolvimento dessa ontologia foi à necessidade de estabelecer padrões sobre terminologias do domínio, de forma à evitar ambiguidades.

A representação semântica é realizada por meio de *Concept Map (Cmap)*, sua implementação foi realizada na ferramenta *Cmap Tools Knowledge*. Como processo de desenvolvimento da ontologia, não foi identificado uma metodologia específica, porém os passos são descritos. Primeiro é identificado os conceitos elementares e realizado uma análise para a definição da semântica, em seguida é identificado os relacionamentos entre os conceitos, por fim são definidos as regras entre os conceitos e relações. No trabalho são apresentados alguns extratos da ontologia, porém não é listado as informações nela representada. A Figura 7 apresenta um dos extratos da ontologia *Space Weather Ontology*.

Figura 7 – Extrato da ontologia *Space Weather Ontology*

Fonte: (MESSEROTTI, 2007).

### 3.2 *HELIO: Discovery and analysis of data in heliophysics*

O *Heliophysics Integrated Observatory (HELIO)*, descrito em Bentley et al. (2013) é um sistema online que coleta e disponibiliza informações sobre *heliophysics*. No artigo é descrito dois problemas: fornecer o acesso ao dados na integra em todos os lugares do planeta, e coletar, organizar e armazenar as informações de forma a serem recuperadas e entendidas pelos pesquisadores.

A ontologia é utilizada por esse sistema para recuperar os arquivos com informações semântica igual ou similar ao que está sendo procurado. Foram encontrados outros trabalhos que apresentaram os mesmos objetivos desse trabalho correlato. Eles buscam coletar informações provenientes de diversos locais, centraliza-las e disponibiliza-las da melhor forma possível para o usuário. Isso não é uma tarefa fácil pois, enfrenta desafios como padronizar os dados coletados em granularidades diferentes e como disponibilizar os dados para os pesquisadores interdisciplinares. Todas também utilizam ontologias no processo de recuperação das informações. Os artigos similares são: Bentley et al. (2011) e King et al. (2010).

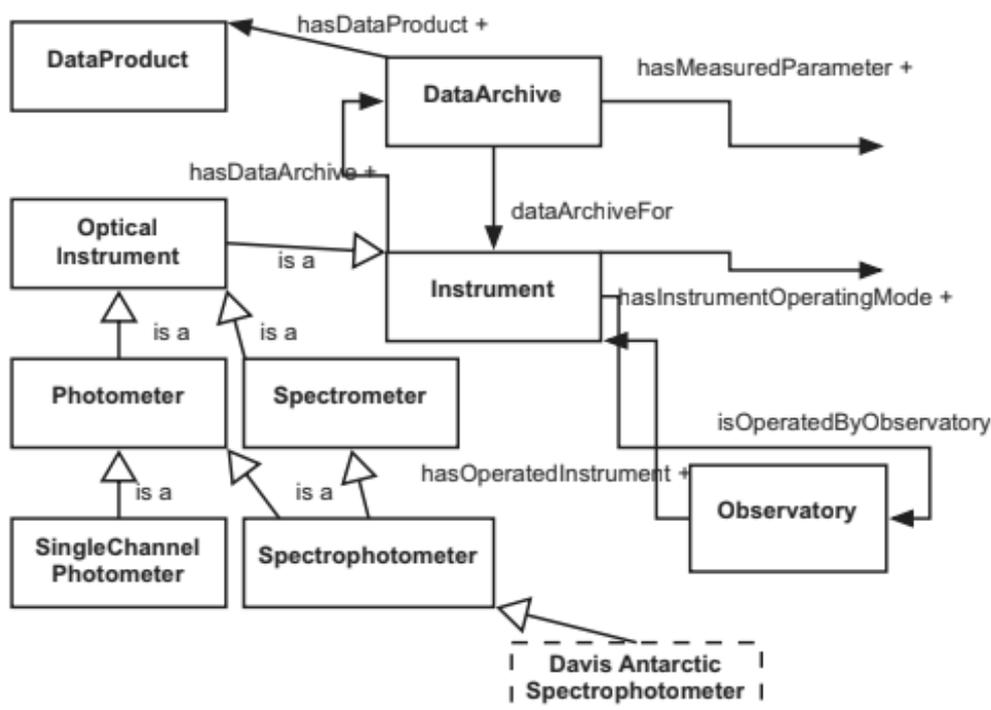
### 3.3 *Ontology-supported scientific data frameworks: The Virtual Solar-Terrestrial Observatory experience*

O artigo apresentado em Fox et al. (2009), cita que pesquisas uni disciplinares tem se transformando em pesquisas multidisciplinares, por isso surge a necessidade de encontrar formas de interligar as informações para que pesquisadores consigam recupera-las facilmente,

independente de sua área de pesquisa. O artigo trabalha com dados da física solar e da física solar-terrestre.

Foi desenvolvido um observatório virtual, que utiliza de técnicas de inteligência artificial e de semânticas para aperfeiçoar o acesso ao dado e a integração semântica deles. Na Figura 8, é apresentada algumas informações representadas na ontologia VSTO.

**Figura 8 – Extrato da ontologia VSTO**



Fonte: (FOX et al., 2009).

### 3.4 Considerações Finais

O levantamento dos trabalhos correlatos não foi um processo simples, inicialmente buscou-se por informações que relacionem explosões solares e ontologias, nenhuma informação foi encontrada. Então optou-se por buscar por assuntos análogos a explosão. Foram recuperadas algumas ontologias do domínio de clima espacial, *heliophysics* e física solar-terrestre.

Em todas as ontologias recuperadas foi possível analisar a forma que os dados sobre o domínio solar são modelos, porém nenhuma ontologia apresentada representa as mesmas informações que foram representadas na *Onto Solar Flare*. A ontologia que mais se assemelha é a referente ao clima espacial, porém ela aborda diversos assuntos que não são relevantes para essa dissertação.

As ontologias referentes aos domínios *heliophysics* e física solar-terrestre, estão voltadas a representar informações para catalogar arquivos com uma semântica embutida, visando melhorar à recomendação dos arquivos para o usuário final e padronizar os dados provenientes de

diversos lugares. Por mais que os objetivos de desenvolvimento dessas ontologias e a *Onto Solar Flare* sejam diferentes, ambas tiveram a mesma preocupação de modelar os termos de forma clara e que possa enriquecer semanticamente uma metodologia ou processo. Esses trabalhos correlatos são suporte à hipótese de que é possível melhorar processos (no caso dessa dissertação, o processo de predição da explosão solar) a partir de conhecimentos sobre o domínio estruturadas em uma ontologia.

# Capítulo 4

## *Onto Solar Flare*

---

---

*Este capítulo apresenta a ontologia desenvolvida. Também são retomadas as questões de pesquisa desta dissertação e a metodologia utilizada para o desenvolvimento da ontologia, bem como a estrutura da ontologia e sua validação.*

### 4.1 Questão de pesquisa

A atividade solar é composta por ciclos solares com duração média de 11 anos, nesse período ocorrem alterações no campo magnético do Sol ([LOSKUTOV et al., 2001](#)). As alterações estão relacionadas com as explosões solares, que são liberações de uma grande descarga de energia que ocorre numa RA da atmosfera solar. Ao atingir a Terra, a explosão pode prejudicar direta ou indiretamente: sistemas de comunicação e navegação, órbita de satélites, linhas de transmissão de energia elétrica, aplicações geofísicas e comunicações de rádio de alta frequência ([NASA SCIENCE, 2016](#)). Visando minimizar os possíveis prejuízos causados pelas explosões, diversos pesquisadores tem buscado formas de prever a ocorrência das explosões solares. Uma das técnicas empregadas é a mineração de dados, porém os pesquisadores tem se deparado com dificuldades de trabalhar com o domínio, pois é um domínio complexo com grande diversidade de dados e terminologias que não são triviais para todos os pesquisadores envolvidos.

Motivada em contribuir com as pesquisas sobre explosões solares fornecendo dados semânticos e apresentando uma nova forma de representar os dados, visando a melhor compreensão dos mesmos, esta dissertação de mestrado aborda as seguintes questões de pesquisa:

- É possível estruturar adequadamente os dados das explosões solares em uma ontologia para auxiliar o processo de predição de explosões solares?
- É possível usar a ontologia para melhorar a predição de explosões solares?

## 4.2 Engenharia de ontologia

Engenharia de ontologia incorpora decisões sobre como representar uma seleção de objetos e suas relações dentro de uma ordem lógica, desenvolvendo um modelo a nível ontológico. O processo de desenvolvimento de ontologias se assemelha ao desenvolvimento de um software, no sentido de ser necessário levantar os requisitos, determinar escopo, desenvolver e validar. Quando o domínio é complexo e desconhecido pelos desenvolvedores como o domínio de explosões solares, uma boa abordagem de se adotar é a metodologia iterativa e incremental. Nessa metodologia o projeto é subdividido em interações, nas quais são definidos o escopo, realizado testes e validações, e ao final é incrementado (adicionado) no projeto desenvolvido até o momento. Adotar essa abordagem auxilia o desenvolvedor a trabalhar com pequenos trechos facilitando a sua compreensão, propiciando a entrega de valor, a melhoria contínua entre outros.

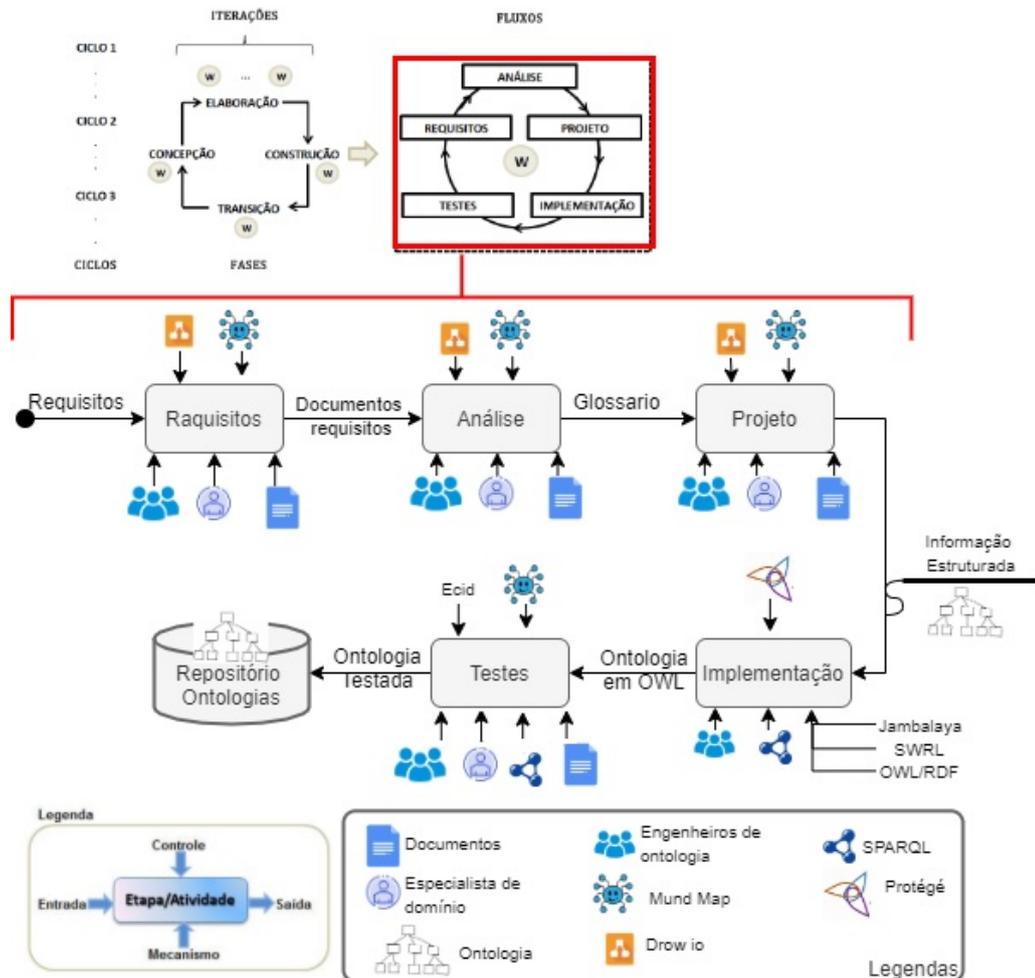
Para o desenvolvimento da ontologia *Onto Solar Flare* optou-se por utilizar a metodologia **UPON** descrita na seção 2.2.1, pois trata-se de uma metodologia iterativa que propicia as vantagens descritas no paragrafo anterior. Como podemos observar na Figura 3, as primeiras iterações (fase inicial) estão principalmente relacionadas à captura de requisitos e a levantamento parcial de alguns conceitos. Durante as iterações subsequentes, a análise é realizada e os conceitos fundamentais são identificados, estruturados e testados.

A cada iteração é enfatizada uma etapa do fluxo relevante para o desenvolvimento. Na primeira iteração para o desenvolvimento da ontologia *Onto Solar Flare* foram coletadas informações sobre as explosões solares em geral, visando entender a fenomenologia do evento e levantar informações sobre o domínio. Na segunda iteração foram levantadas informações sobre o domínio de aplicação (método **ECID**), como resultado final foi possível filtrar as informações coletadas na primeira iteração, que seriam relevantes para enriquecer semanticamente o método. Na terceira iteração foi desenvolvida a primeira versão da ontologia, e na quarta e última iteração foi desenvolvida a versão 2, ambas descritas na seção 4.4.

Por se tratar de um domínio complexo, com grande diversidade de dados e detalhes inerentes a representação do fenômeno, a participação do especialista de domínio do **INPE** foi de extrema importância. Esse pesquisador contribuiu para guiar e validar as informações encontradas na literatura, que foram representadas na ontologia. A seguir serão detalhados os fluxos de trabalho com suas etapas e resultados gerados a partir delas.

O diagrama *Structured Analysis & Design Technique* (**SADT**) da Figura 9 representa a visão geral das etapas do fluxo de trabalho proposto pelo **UPON**, em seguida contém detalhes sobre cada etapa. Nesse diagrama os retângulos representam as atividades, as setas que entram pelo lado esquerdo dos retângulos representam as entradas de dados e as do lado direito representam as saídas geradas em cada atividade. As setas do lado superior representam os controles que orientam a execução de cada atividade, já as setas do lado inferior representam os participantes, as ferramentas e os mecanismos que executam ou automatizam a execução das atividades.

**Figura 9 – Diagrama SADT das etapas de um fluxo de trabalho proposto pela UPON**



Fonte: Os autores.

### Etapa 1 - Requisitos ou Levantamento de requisitos

O principal propósito dessa etapa é identificar o objetivo da ontologia e levantar os requisitos necessários para a construção da mesma. São coletadas informações em entrevistas com os especialistas de domínio e informações coletadas na literatura. Como resultado final tem-se uma lista de termos relevantes, Questão de Competência (QC) (perguntas levantadas pelos pesquisadores, que serão respondidas pela ontologia no final do processo, como uma das forma de validação) e representações gráficas da informações coletadas.

### Etapa 2 - Análise ou Construção dos glossários

A principal proposta dessa etapa é refinar as informações adquiridas na etapa anterior e acrescentar novas informações pertinentes ao domínio. Recebe como parâmetro de entrada a lista de termos e requisitos. Como parâmetros de saída temos os termos detalhados e representações gráficas refinadas.

### Etapa 3 - Projeto ou Organização das informações

O principal objetivo dessa etapa é fornecer uma estrutura hierárquica ou estrutura onto-

lógica dos termos levantados anteriormente. Como parâmetro de entrega recebe-se as listas de termos e requisitos. Como produto final dessa etapa temos uma rede semântica com relacionamentos e possíveis axiomas.

#### **Etapa 4 - Implementação**

O objetivo desta etapa de trabalho é codificar a ontologia em uma linguagem formal, por essa razão o principal responsável por essa etapa é o engenheiro de dados. Inicialmente recebe como entrada uma rede semântica e como saída temos a ontologia implementada.

#### **Etapa 5 - Testes**

A etapa de testes e validação tem como objetivo avaliar a qualidade da ontologia resultante. A **UPON** baseia-se no processo de validação seguindo as quatro vertentes propostas por [Burton-Jones et al. \(2005\)](#):

1. **Validação Sintática:** avalia a estrutura formal da ontologia, ou seja, a forma que a ontologia está escrita.
2. **Validação Semântica:** avalia a consistência, verificando a ausência de conceitos contraditórios e ciclos na modelagem da hierarquia.
3. **Validação Pragmática:** avalia o conteúdo da ontologia e sua relevância para os usuários, independentemente de sua sintaxe e semântica. Está relacionada a três categorias:
  - I Fidelidade: medida pela confiabilidade das fontes utilizadas para adquirir informações no processo de construção dos léxicos.
  - II Relevância: verifica se a ontologia atende as necessidades dos especialistas.
  - III Completude: verifica se a ontologia satisfaz os requisitos e as restrições do problema a ser resolvido, ou seja, verifica se os objetivos, o escopo e as **QCs** definidos na primeira etapa foram alcançados.
4. **Validação Social:** avalia a popularidade da ontologia, ou seja, verifica quantas ontologias se baseiam na ontologia desenvolvida ou quantas vezes a ontologia foi visualizada após a sua publicação.

### **4.3 Desenvolvimento da ontologia**

O desenvolvimento da ontologia *Onto Solar Flare* foi dividido em duas versões. A primeira versão possibilitou analisar a viabilidade da implementação desse projeto e teve por objetivo representar parte das informações definidas no escopo, para contextualizar todos os pesquisadores de como a ontologia poderia contribuir com o processo de predição das explosões solares. A segunda versão teve por objetivo implementar as demais informações definidas no escopo. As etapas de levantamento de informações, construção dos léxicos e glossário são comuns

para as duas versões da ontologia, o que as diferencia são apenas as fases de implementação e definição do escopo.

A seguir é apresentado as saídas de cada etapa de todas as iterações.

### 4.3.1 Levantamento de requisitos

Inicialmente foram realizadas diversas reuniões com o especialista de domínio sobre explosões solares, com o objetivo de entender sobre a fenomenologia do evento e listar todas as informações que possam contribuir de alguma forma com a previsão da ocorrência de explosões solares. As reuniões foram presenciais, realizadas no [INPE](#) onde o pesquisador de explosões solares explicava sobre a evento e indicava alguns artigos a serem lidos referente ao tema abordado na reunião. Essas informações eram estruturadas em mapas mentais e apresentada ao pesquisador na reunião seguinte, e o ciclo de reuniões se repetia. O mapa mental foi escolhido porque é uma estrutura comum, que é utilizada independente da área de conhecimento e se assemelha a uma estrutura de grafo, converte-la em ontologia seria um processo mais fácil. Essas informações levantadas nas reuniões são chamadas de **Léxico de Domínio (LD)**.

Em seguida foram realizadas reuniões com os especialistas de aplicação, para entender o fluxo do método [ECID](#), e filtrar quais informações listadas anteriormente seriam relevantes para essa versão da ontologia, as informações serão chamadas de **Léxico de Aplicação (LA)**. Após as duas etapas exploratórias, foi possível estabelecer algumas definições:

1. **Domínio de Interesse:** o fragmento da realidade a ser modelada são alguns dados utilizados pelos especialistas de domínio para monitorar o fenômeno. Esses dados estão fortemente relacionados ao ciclo solar.
2. **Escopo:** Os dados que se destacaram durante as entrevistas com o especialista de domínio são classificações de raio-x e das **RAs** (*McIntosh* e *Mount Wilson*), ambas se relacionam com o ciclo solar. Para a primeira versão da ontologia foi considerado apenas informações sobre raio-x e ciclo solar. Na segunda versão foram representadas as demais informações propostas no escopo.
3. **Objetivo do Negócio:** viabilizar a visualização de informações que possam contribuir para representar a ocorrência de explosões solares, e aproveitar essas informações nos métodos de predição.
4. **Lista de Termos:** extrato da lista de termos coletadas nas reuniões com os especialistas:
  - **LD:** *Campo Magnético, Classes Raio-X, Classificação magnética McIntosh, Classificação magnética Mt. Wisson, Classificação magnética Zurick, Coroa, Cromosfera, Densidade do plasma, Detector EUV, Detector RX/R-gama, Ef. Doppler, Efeitos Sw Nível, Ejeções de Massa coronal (CMEs), Explosão solar, Fases do fenômeno*

(prê-flare, impulsiva, gradual, estendida), Filamentos, Fluxo de energia, Fotosfera, Girossíncrotron, H-alpha, Heliosismógrafo, Luz branca, Magnetômetro, Perturbações causadas pelas explosões solares, Raios-gama, Radiotelescópio, Região Ativa, RX duros, RX moles, Raios-X, Sensores EUV, Sensores Raios-gama, Sensores RX, Telescópio ótico, UV, etc.

- **LA:** Classes de Raio-X, Classificadores de series temporais, Classificadores, Classificação magnética McIntosh, Computar arquitetura de dispositivo unificado (CUDA), Dados desbalanceados, ECID, Ejeções de Massa coronal (CMEs), Explosão solar, GOES, Graphics Processing Unit (GPU), Mineração de dados, Predição, Região ativa, Solar Dynamics Observatory (SDO), SeMiner, Series temporais, etc.

5. **Questões de Competência (QC):** são perguntas que a ontologia deve ser capaz de responder como forma de validação. As informações foram retornadas de acordo com a base de dados utilizada em cada versão da ontologia. Esses dados estão descritos na seção 4.4.

- **QC 01:** Quais classes de explosões solares podem se manifestar na etapa do ciclo ascendente?

*Todas as classificações raio-x: A, B, C, M e X.*

- **QC 02:** Quais classes de explosões solares, podem se manifestar na etapa do ciclo máxima

*Todas as classificações raio-x: B, C, M e X.*

- **QC 03:** A qual ciclo e etapa do ciclo a data 07/05/2010 se refere?

*Ciclo 23, etapa ascendente.*

- **QC 04:** Quais classificações de explosões solares foram geradas no dia 09/06/2017?

*Classe B.*

- **QC 05:** Quais classificações McIntosh foram geradas no dia 09/06/2017?

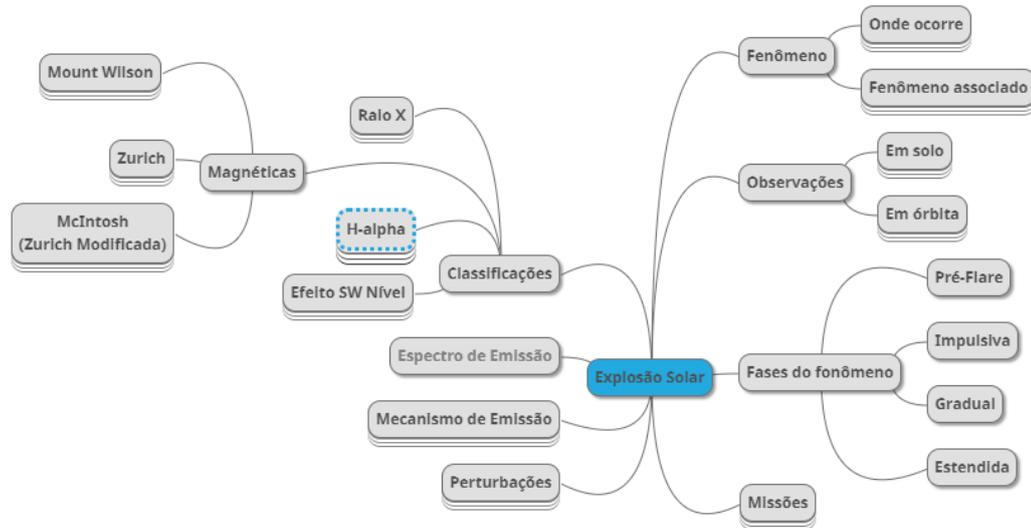
*Classe Axx*

- **QC 06:** Quais classificações Mount Wilson foram geradas no dia 09/06/2017?

*Classe Alpha*

6. **Representação Gráfica:** É recomendado pela UPON, para as representações gráficas, diagramas de casos de usos. No entanto, para o desenvolvimento deste projeto contou-se com um time de pesquisadores heterógeno, e esses diagramas não são frequentemente utilizados por todos, optou-se por representar as informações em forma de hierarquias similar denominado mapa mental. A Figura 10, ilustra um extrato das primeiras informações coletadas com o especialista de domínio sobre explosões solares.

**Figura 10 – Representação gráfica das primeiras informações coletadas com o especialista de domínio sobre explosões solares**



Fonte: Os autores.

### 4.3.2 Construção dos glossários

A construção dos glossários é a etapa responsável pelo refinamento das informações anteriormente identificadas.

1. **Glossário de Referência (GR):** De acordo com o método UPON, o LA é enriquecido com o LD, gerando o GR. Adicionalmente o GR incorpora as definições dos termos nele contido, dessa vez as informações foram coletadas a partir do referencial teórico indicado pelo especialista de domínio. A seguir é apresentada uma lista resumida dos termos utilizados nas duas versões da ontologia, ressalta-se que a maioria das informações foram descritas no Capítulo 2, especificamente na seção 2.1.

- **Ciclo solar:** Ciclo de atividade solar: intervalo de tempo, com duração média de 11 anos, de comportamento cíclico, durante o qual se registra um padrão sequencial de:
  - ausência de manchas solares na fotosfera e RA na atmosfera solar - fase de mínimo;
  - aumento gradual do número de manchas e grupos de manchas na fotosfera além de RA e explosões na atmosfera solar - fase ascendente;
  - maior número de manchas e grupos de manchas na fotosfera além de RA e explosões na atmosfera solar - fase de máximo;
  - finalmente diminuição gradual do número de manchas e grupos de manchas na fotosfera além de RA e explosões na atmosfera solar - fase de decaimento até o próximo mínimo. A partir de então o mesmo padrão se repete com um novo ciclo de atividade.

- **Classificações:** são classes agrupadas seguindo um método. No contexto da ontologia *Onto Solar Flare* é utilizado para representar as classificações que descrevem as **RA** e explosões solares.
  - **RA:** Grupos de manchas escuras visíveis na fotosfera, associadas a seus respectivos arcos magnéticos observados na atmosfera solar.
  - **Raio-x:** No contexto da ontologia *Onto Solar Flare* é um sensor que mede a intensidade da explosão solar na faixa de 1-8 Angstroms.
2. **Modelagem do Cenário da Aplicação:** Essa etapa tem como entrada a lista das informações que serão representadas na ontologia e que serão relevantes para o domínio de aplicação. Para a execução dessa etapa é necessário uma análise do domínio de aplicação, para identificar onde a ontologia pode contribuir para o processo de predição e elaborar um fluxo do novo método gerado. O Capítulo 5 contém informações sobre a incorporação da *Onto Solar Flare* ao método **ECID**.

### 4.3.3 Organização das informações

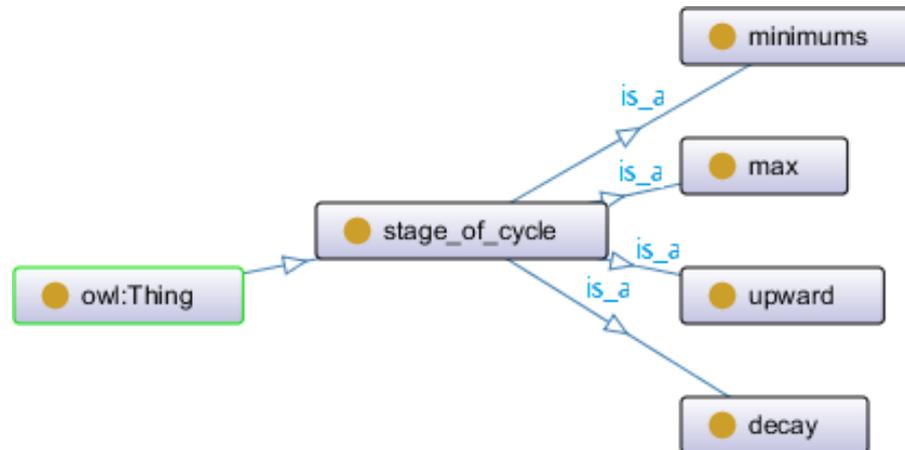
Nessa etapa deixa-se de considerar a dimensão linguística do desenvolvimento da ontologia e considera-se a dimensão conceitual, visando estabelecer uma estrutura ontológica dos termos contidos no **GR**. Isso é alcançado por meio da organização dos termos em hierarquias conceituais, e sua estruturação por meio de atributos e axiomas.

1. **Modelagem dos Conceitos:** Nessa etapa é identificado os principais conceitos do **GR**. Os termos definidos são os mesmos estabelecido no escopo do projeto.
2. **Modelar hierarquia de conceitos e relacionamentos específicos do domínio:** A organização dos conceitos em uma hierarquia traz a primeira versão gráfica da ontologia. Genericamente os conceitos são ligados por relacionamentos, os mais frequentes são 'é-um', agregação 'parte-de', e associação. A Figura 11 ilustra um extrato da ontologia com o relacionamento de "é-um" ou no inglês "is\_a".

### 4.3.4 Implementação

Nesta etapa, a codificação da ontologia foi realizada utilizando uma linguagem formal. A ontologia é caracterizada por um conjunto de conceitos e relações, além dos axiomas que, são restrições importantes para representar a realidade modelada. Os axiomas podem estar associados à restrições de cardinalidade, hierarquia (por exemplo, a disjunção de dois sub-conceitos) ou tipos de expressões como expressão booleana.

Como resultado final da etapa de desenvolvimento da *Onto Solar Flare*, obteve-se a ontologia codificada em **OWL**, mas especificamente a **OWL DL**. Para seu desenvolvimento foi

**Figura 11** – Extrato da ontologia *Onto Solar Flare* que representa um relacionamento *is a*

Fonte: Os autores.

utilizado: o Protegé 3.4.8; a extensão Jambalaya para visualização da base de conhecimento; *Semantical Web Rule Language (SWRL)* para a definição dos axiomas; *SPARQL Query Language for RDF (SPARQL)* para realizar consultas na ontologia.

Para a coleta dos dados utilizados na ontologia, foi necessário um processo de extração e limpeza dos dados descrito na seção 4.4.

A Figura 12 apresenta um extrato do código em *OWL-DL* gerado para a ontologia *Onto Solar Flare*, focando na parte de inserção de axiomas associados à hierarquia (disjunção de conceitos).

### 4.3.5 Testes

A etapa de teste é responsável por validar a qualidade semântica e pragmática da ontologia. De acordo com a metodologia de validação de ontologias, proposta por [Burton-Jones et al. \(2005\)](#) e recomendada pela *UPON*, as validações devem considerar os quatro pontos descrito a seguir:

1. **Qualidade sintática:** mede a qualidade da ontologia de acordo com sua formalidade (forma que está escrita).

*No caso da Onto Solar Flare, a estrutura das informações foram validadas nas duas primeiros etapas. A qualidade sintática foi verificada na etapa anterior e garantida durante a codificação da OWL DL, gerada automaticamente pela ferramenta Protégé.*

2. **Qualidade semântica:** segundo os autores da metodologia de validação, a principal preocupação é com a ausência de conceitos contraditórios e ciclos ou *loops* entre as informações.

Figura 12 – Extrato do código em OWL-DL da ontologia *Onto Solar Flare*

```
<?xml version="1.0"?>
<rdf:RDF
  xml:base="http://www.owl-ontologies.com/Ontology1594519724.owl">
  <owl:Ontology rdf:about=""/>
  <owl:Class>
    <owl:unionOf rdf:parseType="Collection">
      <owl:Class rdf:ID="Classes"/>
      <owl:Class rdf:ID="CycleStep"/>
    </owl:unionOf>
  </owl:Class>
  <owl:Class rdf:ID="Class_C">
    <owl:disjointWith>
      <owl:Class rdf:ID="Class_X"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Class_M"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Class_B"/>
    </owl:disjointWith>
    <owl:disjointWith>
      <owl:Class rdf:ID="Class_A"/>
    </owl:disjointWith>
    <rdfs:subClassOf>
      <owl:Class rdf:ID="X-ray"/>
    </rdfs:subClassOf>
  </owl:Class>
```

Fonte: Os autores.

A consistência da ontologia foi verificada pelo raciocinador Pellete na versão 1.5.2, a mesma que acompanha a versão utilizada do Protégé. Para evitar loops, as classes e subclasses são disjuntas, como pode-se ver no uso do comando *disjointWith* na Figura 12.

3. **Qualidade pragmática:** refere-se ao conteúdo da ontologia, independente de sua qualidade sintática e semântica. São considerados os três pontos a seguir:

- **Fidelidade:** verifica o quanto a ontologia é fidedigna ao domínio. Pode ser alcançada, por exemplo, verificando as fontes das informações coletadas.  
*O especialista de domínio sobre explosões solares auxiliou constantemente na checagem das referências utilizadas nos processos de construção de léxicos e do glossário.*
- **Relevância:** verificada junto com a integridade, é considerado a correta aplicação dos requisitos da ontologia.
- **Integralidade:** segundo os autores da metodologia, uma ontologia é completa quando satisfaz os requisitos e atende as restrições do problema a que se propõe resolver. Por essa razão esse requisito pode ser atendido considerando o escopo, objetivo de negócio e as QCs estabelecidos na etapa - levantamento de requisitos.

A ontologia *Onto Solar Flare* atendeu ao escopo estabelecido para cada versão, representando as informações que foram propostas. O objetivo de negócio foi alcançado com a visualização gráfica das informações na taxonomia gerada na ontologia, e a reutilização dessas informações na junção da ontologia com o método *ECID* (mais informações no Capítulo 5), e por fim, as *QC* foram respondidas corretamente.

4. **Qualidade social:** essa validação considera a popularidade da ontologia, como quantas vezes ela é utilizada por outras ontologias ou o número de vezes que ela é acessada. No caso, da *Onto Solar Flare*, essa validação será realizada apenas após a sua publicação.

## 4.4 Descrição da ontologia

Nessa seção é apresentado extratos da ontologia *Onto Solar Flare* versão 1 e 2, além do processo para a coleta dos dados nelas utilizados.

## 4.5 *Onto Solar Flare* - Versão 1

A Figura 13 ilustra a ontologia. Os retângulos são as classes e os relacionamentos são as linhas que as interligam.

A ontologia é instanciada com uma data na classes 'solar\_globe'. Com o relacionamento *has\_cycle* e *has\_step\_cycle* é possível inferir a qual ciclo e etapa do ciclo a ontologia referencia. Identificando a etapa do ciclo, é possível retornar quais foram as porcentagens de ocorrência de explosões solares por classe e por etapa do ciclo. Para exemplificar temos as seguinte ordem de triplas.

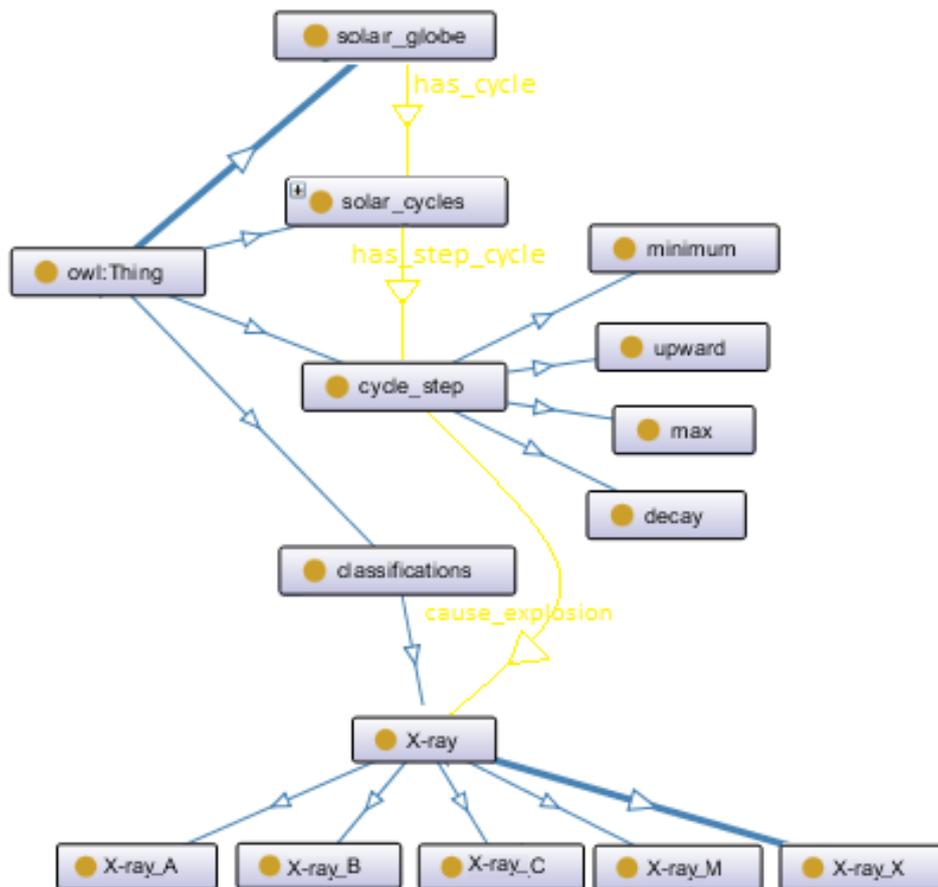
**Solar\_globe** *has\_cycle* **Solar\_cycle** *has\_step\_cycle* **Cycle\_step** *is\_a* **Max** *cause\_explosion*  
**X-ray\_X** *has\_proporcion* 2,89%

Para representar a ocorrência de explosões solares ao longo do ciclo, foi decidido calcular a proporção dos dados seguindo a Equação 4.1. A letra grega  $\gamma$  representa o número de ocorrências da uma classe de raio-x e em uma única etapa do ciclo. E a  $\delta$  representa o total de ocorrências de explosões na mesma etapa do ciclo.

$$\beta = \frac{\gamma}{\delta} \quad (4.1)$$

Os dados para realizar o cálculo são os mesmos utilizados pelo treinamento do método *ECID*. Eles foram coletados em dias aleatórios, entre os anos de 2010 a 2017, mantendo o desbalanceamento dos dados característico do domínio. Para delimitar as etapas do ciclo foram

**Figura 13 – Ontologia *Onto Solar Flare* - Versão 1**



Fonte: Os autores.

assumidos os períodos listados na Tabela 1. Essa informação foi fornecida pelo especialista de domínio sobre explosões solares que contribuiu com o desenvolvimento deste projeto.

**Tabela 1 – Intervalo de tempo entre as etapas do ciclo 24**

	Data Início	Data Fim
Miníma	dezembro 2006	dezembro 2008
Ascendente	janeiro 2009	agosto 2011
Máxima	setembro 2011	fevereiro 2015
Decaimento	março 2015	agosto 2019

Fonte: Os autores.

A Tabela 2 apresenta as porcentagens obtidas no cálculo das proporções. É importante notar que a etapa do ciclo mínima termina em 2008 e os dados de treinamento do ECID foram selecionados a partir de 2010, isso justifica a porcentagem zerada na etapa de mínima.

**Tabela 2 – Proporções das classes de explosões solares por etapa do ciclo, calculadas a partir dos dados de treinamento do ECID**

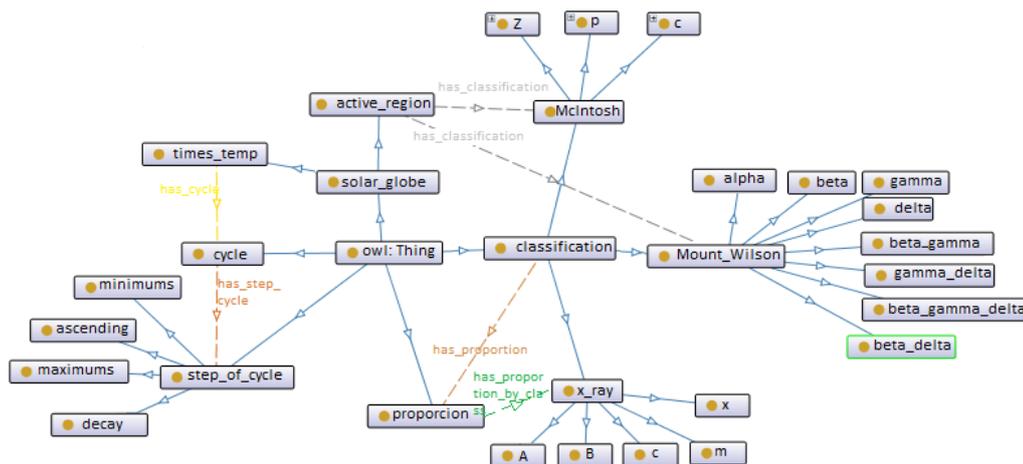
Classes / Fases Ciclo	Miníma	Ascendente	Máxima	Decaimento
Classe A	0%	21,82%	15,08%	48,14%
Classe B	0%	35,10%	16,77%	23,24%
Classe C	0%	37,68%	52,59%	24,54%
Classe M	0%	4,60%	12,64%	4,06%
Classe X	0%	0,78%	2,89%	0,0%

Fonte: Os autores.

## 4.6 Onto Solar Flare - Versão 2

A Figura 14 ilustra a ontologia *Onto Solar Flare - Versão 2*. Nela foram adicionado informações as classificações *McIntosh* e *Mount Wilson*.

**Figura 14 – Ontologia *Onto Solar Flare - Versão 2***



Fonte: Os autores.

A ontologia é instanciada com a data e as classes de raio-x que o *ECID* retorna de sua predição, e as configurações magnéticas de *McIntosh* e *Mount Wilson* das *RA* visíveis no dia que o método foi executado. É inferido qual ciclo e etapa do ciclo os dados estão referenciando e identificado as proporções para cada classificação. Com consultas em *SPARQL* é possível somar as proporções retornadas de para cada classificação e retornar seu resultado.

Para mensurar a ocorrência de cada rotulo das classificações de raio-x, *McIntosh* e *Mount Wilson*, utilizou-se a mesma equação da primeira versão da ontologia (Equação 4.1.). Entretanto os dados dessa versão são todos referentes ao ciclo solar 24, com início em dezembro de 2006 e fim em agosto de 2019. Para o calculo das proporções também foi utilizado os mesmo períodos das etapas do ciclo apresentados na Tabela 1.

Os dados da classificação raio-x foram coletados na base de dados da NOAA acessível em *NOAA (2020)*, a instituição disponibiliza um arquivo por dia com os dados dos eventos

produzidos naquele dia, a Figura 15 ilustra um exemplo de arquivo do dia 2012 Mar 05<sup>1</sup>. Para extrair as informações necessária, foi desenvolvido um *script* que seleciona a datas dos arquivos, classificação de raio-x caso tenha alguma ocorrência no dia e o número da RA. No caso da Figura 15, são recuperadas as informações: Data 2012 Mar 05, classificação de raio-x 'X' e número da RA 1429.

**Figura 15 – Arquivo da base de dados NOAA para coleta de dados da classificação de raio-x. Dados referente ao dia 2012 Mar 05**

```

:Product: 20120305events.txt
:Created: 2012 Mar 08 0332 UT
:Date: 2012 03 05
# Prepared by the U.S. Dept. of Commerce, NOAA, Space Weather Prediction Center
# Please send comments and suggestions to SWPC.Webmaster@noaa.gov
#
# Missing data: ////
# Updated every 30 minutes.
#
# Edited Events for 2012 Mar 05
#
#Event   Begin   Max     End  Obs  Q  Type  Loc/Frq  Particulars  Reg#
#-----
[...]|
1140     0230   0409    0443  G15  5  XRA  1-8A     X1.1  3.7E-01  1429
1270     0252   ////    0427  PAL  C  RSP  039-056  CTM/1
1140     0316   0348    0643  LEA  3  FLA  N17E52   2B    PRB      1429
1140     0333   0432    0606  LEA  G  RBR  4995     5300    1429
1170     0333   ////    0833  LEA  U  RSP  025-075  VI/2
1140     0333   0442    0615  LEA  G  RBR  410      480000  1429
1140     0334   0434    0601  LEA  G  RBR  2695     12000   1429

```

Fonte: Os autores.

FOs dados das classificações *McIntosh* e *Mount Wilson* estão disponíveis em ??), também são disponibilizados diariamente pela instituição *National Oceanic and Atmospheric Administration* (NOAA). A Figura 16 ilustra um exemplo de arquivo do dia 2012 Mar 05<sup>1</sup>. Para coletar as informações foi desenvolvido um segundo *script*, que seleciona a data, as classificações e o número da RA. No caso da Figura 16 podemos identificar as seguintes informações: data: 2012 Mar 05, *McIntosh*: Dkc, *Mount Wilson*: Beta-Gamma-Delta e número da RA: 1429. Ao analisar os dois arquivos nota-se que a RA numero 1429, de configuração magnética Dkc e Beta-Gamma-Delta produziu uma explosão solar classe X.

<sup>1</sup> Dados de raio-x - ftp://ftp.swpc.noaa.gov/pub/warehouse/2012/2012\_events.tar.gz

<sup>1</sup> Dados de magnetograma - ftp://ftp.swpc.noaa.gov/pub/warehouse/2011/2011\_SRS.tar.gz

**Figura 16 – Arquivo da base de dados NOAA para coleta de dados das classificações *McIntosh* e *Mount Wilson*. Dados referente ao dia 2012 Mar 05**

```

:Product: 0305SRS.txt
:Issued: 2012 Mar 05 0030 UTC
# Prepared jointly by the U.S. Dept. of Commerce, NOAA,
# Space Weather Prediction Center and the U.S. Air Force.
#
Joint USAF/NOAA Solar Region Summary
SRS Number 65 Issued at 0030Z on 05 Mar 2012
Report compiled from data received at SWO on 04 Mar
I. Regions with Sunspots. Locations Valid at 04/2400Z
Nmbr Location Lo Area Z LL NN Mag Type
[...]
1429 N18E55 300 0700 Dkc 08 08 Beta-Gamma-Delta
1430 N19E40 315 0020 Cro 03 03 Beta
IA. H-alpha Plages without Spots. Locations Valid at 04/2400Z Mar
Nmbr Location Lo
1426 N11W87 082
II. Regions Due to Return 05 Mar to 07 Mar
Nmbr Lat Lo
None

```

Fonte: Os autores.

## 4.7 Considerações Finais

Nesse capítulo foi apresentado a metodologia de desenvolvimento de ontologia *UPON*, e os resultados obtidos com a execução de cada etapa preposta pela metodologia, além de descrever as duas versões da ontologia desenvolvida.

A *UPON* conta com etapas bem estruturadas e definidas, possibilitando identificar facilmente quais informações são relevantes para cada etapa e qual o resultado final da mesma, tornando a metodologia bem clara e intuitiva. Todas as etapas estão interligadas, de uma forma que facilita o desenvolvimento da próxima, por exemplo as informações coletas já são estruturadas de uma forma que facilita a implementação e validação.

Nesse capítulo também foi desenvolvido dois objetivos específicos propostos para este projeto. O primeiro é a criação uma base de conhecimentos sobre explosões solares a partir da literatura e especialistas de domínio. Ele foi alcançado com o desenvolvimento das duas versões da ontologia *Onto Solar Flare*, . O segundo objetivo foi a criação de uma base de dados para dar suporte a predição, ele foi alcançado com a coleta e estruturação dos novos dados utilizados na ontologia, são eles *McIntosh* e *Mount Wilson* e ciclo solar.

# Capítulo 5

## ECID ONTO

---

---

*Nesse capítulo será descrito o método *Ensemble of Classifiers for Imbalanced Datasets with Ontology (ECID Onto)*, contendo informações do desenvolvimento e validação do mesmo. Esse método foi desenvolvido com o intuito usar as informações coletadas *Onto Solar Flare* para melhorar a predição de explosões solares. Também é apresentada uma comparação entre *ECID Onto* e seu predecessor *ECID*.*

### 5.1 Visão Geral do ECID Onto

Como proposto na metodologia de desenvolvimento de ontologias *UPON*, é recomendado estabelecer um domínio de aplicação para ajudar à delimitar o escopo do desenvolvimento. Neste trabalho, optou-se por aplicar à ontologia no método *ECID*, descrito na seção 2.3, pois o método foi desenvolvido na mesma instituição deste trabalho, facilitando a comunicação entre os pesquisadores. A junção do *ECID* e a *Onto Solar Flare* também pode ser considerada uma forma de validação da ontologia, caso se o método escolhido, apresentou melhores resultados com a ontologia, ela atendeu a necessidade para qual foi desenvolvida. Como resultado final obteve-se o *ECID Onto*.

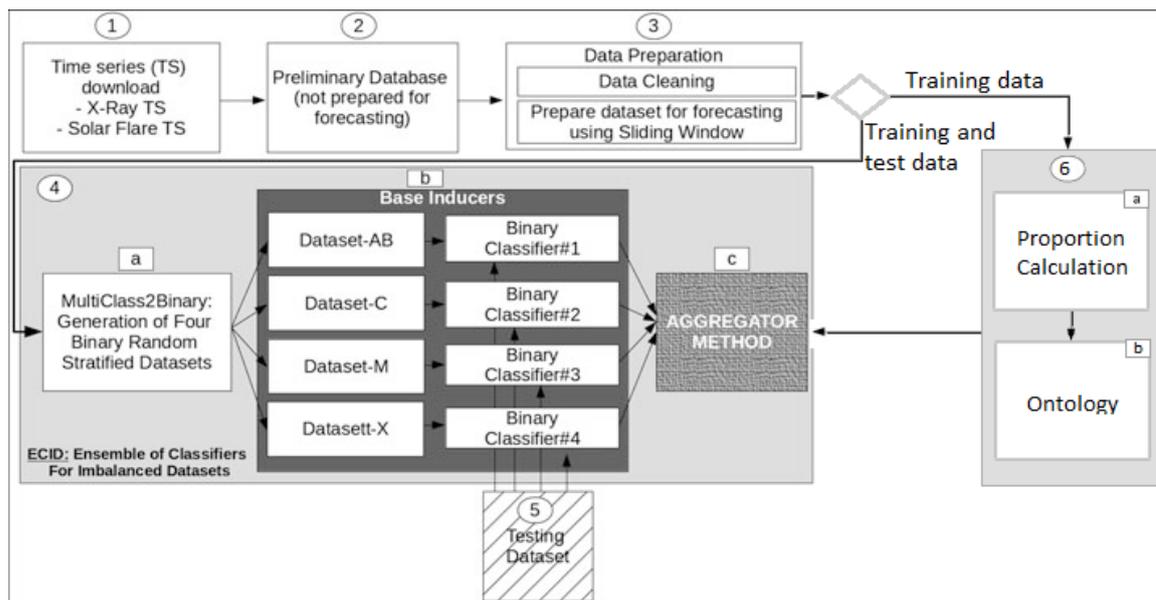
A ontologia pode enriquecer semanticamente o *ECID* em diversos pontos, por exemplo, para auxiliar nos pesos das classes ao realizar as classificações na etapa 4.b da Figura 4, e no modelo agregador na etapa 4.c da mesma figura. Como delimitação deste trabalho, optou-se por alterar o *ECID* no método agregador para incorporar a ontologia. Por a ontologia ter sido desenvolvida em duas versões, o *ECID Onto* também foi versionado. Nas seções 5.2 e 5.3 contém informações sobre cada versão.

## 5.2 ECID Onto - Versão 1

### 5.2.1 Desenvolvimento

A Figura 17 ilustra uma visão geral do método **ECID Onto** - versão 1. Essa versão foi importante para analisar a viabilidade de acoplar a ontologia ao método agregador do **ECID** (etapa 4.c), vide ilustração do método **ECID** na Figura 4.

Figura 17 – Visão geral do método **ECID Onto** - Versão 1



Fonte: Os autores.

A execução do **ECID Onto**, se inicia com os passos propostos pelo **ECID** para tratamento e limpeza dos dados. As etapas 4 e 6 são executadas em paralelo, ou seja, a partir do momento em que os dados estão prontos na etapa 3, uma cópia é consumida pela etapa 4, para gerar o modelo de predição e a outra cópia é utilizada para realizar o cálculo das frequências relativas de cada classe de explosão solar por ciclo solar (cálculo descrito na seção 4.4). Ressalta-se que os dados utilizados para o cálculo da frequência são imputados na ontologia, esses dados são os dados de treinamento obtidos antes de ser aplicado o algoritmo *Sliding Window*.

A etapa 6 é a responsável pelos passos da ontologia, como mencionado anteriormente, primeiro é calculado a frequência relativa de cada classe de explosão (etapa 6.a) e, em seguida, esses valores de frequência são imputados na ontologia (etapa 6.b). Por fim, tem-se a execução do novo método agregador (etapa 4.c). O resultado da agregação das classes da etapa 4.b é gerado e esse resultado é utilizado para instanciar a ontologia. No caso dessa versão, a ontologia é instanciada com a data referente a predição que está sendo realizada. A ontologia retorna a frequência relativa das classes de explosão solar (A, B, C, M e X), e o resultado retornado no primeiro passo do método agregador é comparado com a frequência relativa retornada pela ontologia, seguindo o seguinte o Algoritmo 1.

**Algorithm 1** Método ECID Onto - Versão 1

---

**Input:** Classes previstas pelo ECID  
**Result:** Classes previstas pelo ECID Onto

initialization;

*if Se etapa do ciclo for igual a etapa de menor frequência relativa da classe X then*  
  X then  
    Excluir classe X;

*if Se etapa do ciclo for igual a etapa de menor frequência relativa da classe M then*  
  M then  
    Excluir classe M;  
    Excluir classe X;

*if Se etapa do ciclo for igual a etapa de menor frequência relativa da classe C then*  
  C then  
    Excluir classe C;  
    Excluir classe M;  
    Excluir classe X;

*if Se etapa do ciclo for igual a etapa de menor frequência relativa da classe AB then*  
  AB then  
    Excluir classe AB;

---

Fonte: Os autores

---

Para exemplificar o algoritmo, supomos que o método agregador retorna as classes CMX, e a data da predição é referente a etapa do ciclo decaimento, etapa a qual apresentou menor proporção da classe M, seguindo os dados da Tabela 2. Então a classe M será desconsiderada e as suas subsequentes também. Considerando a escala de intensidade da explosão (A, B, C, M e X), tem-se como retorno final apenas a classe C. O mesmo processo ocorre para as demais classes. Se o resultado final não for nenhuma classe, então são retornadas as classes AB. Sobre o método, tem-se duas observações: as classes A e B são consideradas uma única classe; e, os dados utilizados originalmente pelo [ECID](#), são os mesmos utilizados no [ECID Onto](#). Esses dados não contemplam a etapa do ciclo *mínima*, por essa razão estão zeradas na Tabela 2 e, também, não foram considerados como menor proporção.

### 5.2.2 Validação

Para facilitar a comparação do método [ECID](#) com o [ECID Onto](#), optou-se por replicar as mesmas métricas apresentadas no artigo que descreve o [ECID](#) ([JUNIOR et al., 2019](#)). No artigo foram utilizadas três métricas para avaliar o desempenho de três indutores binários, aplicados separadamente na etapa 4.b (Figura 4). Os indutores são: IBK, SVM e J48, ainda segundo o mesmo artigo, o indutor que melhor apresentou resultados foi o J48, por isso, optou-se por comparar os resultados do [ECID Onto](#) apenas com o resultados obtidos com o J48.

As métricas utilizadas foram adaptadas para avaliar uma predição multi-classe e multi-label. Estão listadas a seguir, considera-se a variável **Actual** como o número de classes reais e **Forecast** o número de classes previstas.

- **Aggregated multi-label precision:**

$$\beta = \frac{|Actual \cap Forecast|}{|Forecast|} \quad (5.1)$$

- **True Positive Rate (TPR):**

$$\beta = \frac{|Actual \cap Forecast|}{|Actual|} \quad (5.2)$$

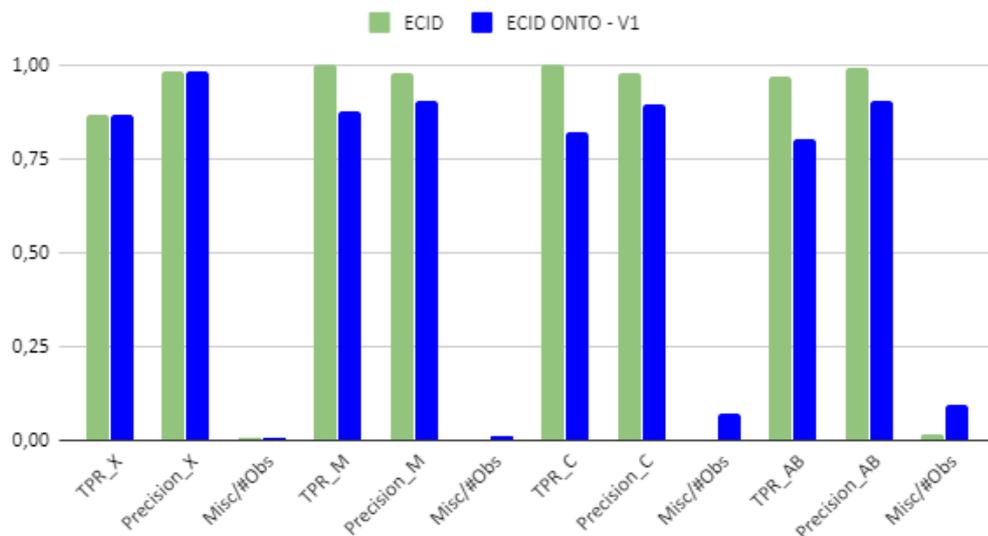
- **Misc/#Obs:**

$$\beta = \frac{FalseNegative(FN)}{AllData} \quad (5.3)$$

A Figura 18 demonstra os resultados dos experimentos com os métodos ECID e ECID Onto. Os valores obtidos pelas métricas do ECID Onto diminuíram em comparação com o ECID. O maior decréscimo foi de 17,08% na métrica TPR da classe C. Porém a classe X, considerada mais relevantes pelos especialistas, não variou. Mesmo com os decréscimos, os resultados finais são considerados satisfatórios pelo especialista.

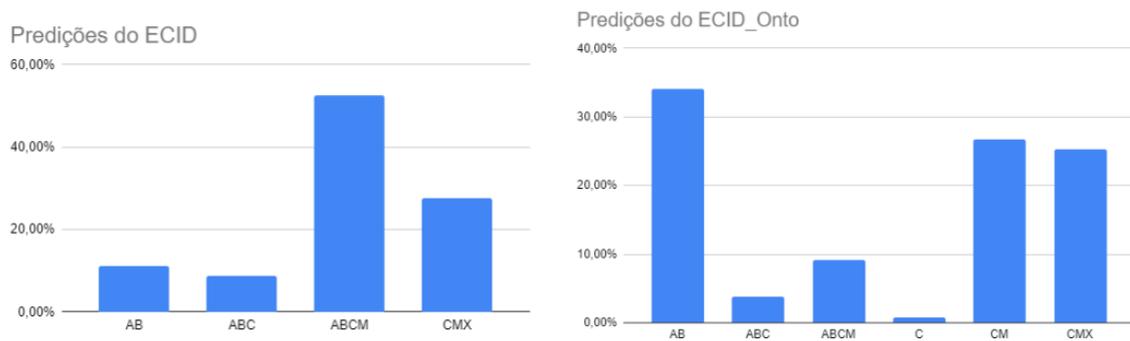
No entanto, houve uma melhora na separação das classes multi-label retornadas pelo método. Pode-se observar na Figura 19, a maior quantidade de predições obtidas pelo ECID é ABCM, o qual engloba muitas classes e o ECID Onto distribuiu melhor as predições entre as classes, retornando mais resultados C, CM e CMX.

**Figura 18 – Experimentos do ECID e ECID Onto - Versão 1**



Fonte: Os autores.

**Figura 19 – Distribuição das classes previstas pelo ECID e pelo ECID Onto - Versão 1**



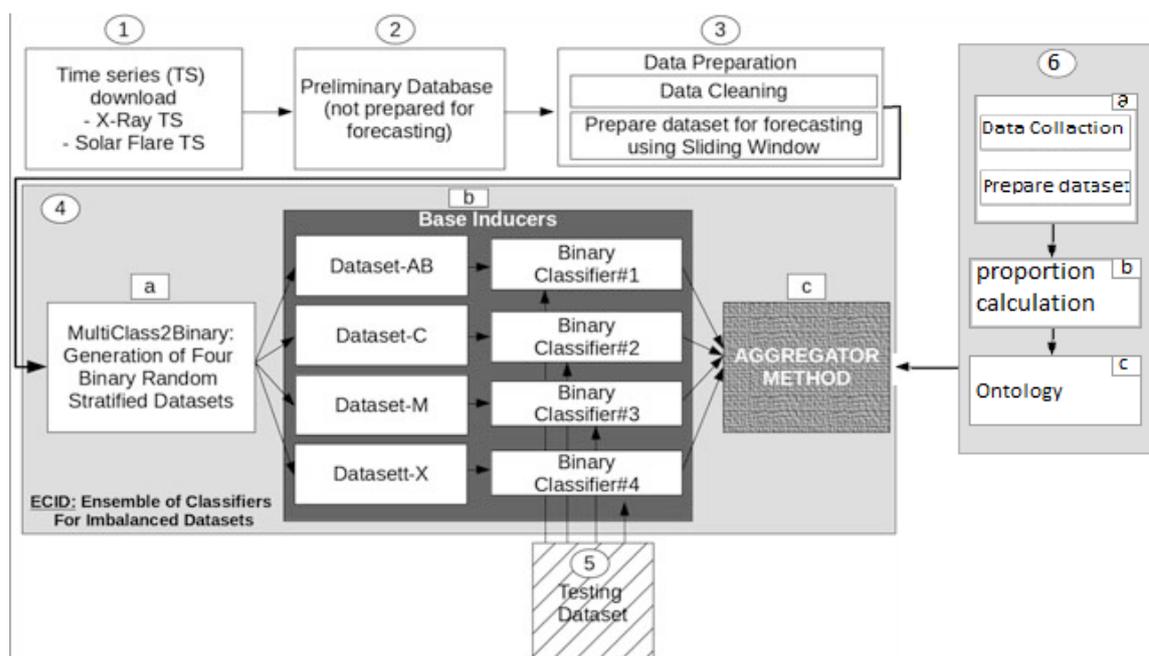
Fonte: Os autores.

### 5.3 ECID Onto - Versão 2

#### 5.3.1 Desenvolvimento

A segunda versão do **ECID Onto**, é composta pela *Onto Solar Flare* - versão 2 e pelo método **ECID**. A ontologia foi incorporada ao método agregador igual a primeira versão, porém com algumas variações na metodologia para a incorporação, pois sentiu-se a necessidade de trabalhar com um conjunto de dados que retrate os eventos de um ciclo solar completo. Esse novo conjunto conta com todos os dados coletados do ciclo solar 24, considerando o período de início em dezembro de 2006 e terminou em agosto de 2019. A Figura 17, ilustra o método **ECID Onto** - versão 2.

**Figura 20 – Visão geral do método ECID Onto - Versão 2**



Fonte: Os autores.

Nesse novo fluxo do [ECID Onto](#), as etapas 1, 2 e 3 responsáveis por coleta e tratamento dos dados para gerar o modelo de predição e a etapa 4 até o 4.b responsáveis por gerar o modelo são executadas em paralelo à etapa 6.

Na etapa 6.a são extraídos e tratados os dados que serão utilizados para calcular as frequências relativas de cada classe de explosão solar utilizadas pela etapa 6.b., esses dois passos estão descritos na seção 4.4. As frequências relativas são imputadas na ontologia e, em seguida, tem-se a execução do modelo agregador modificado (etapa 4.c). Para instanciar a ontologia na etapa 4.c, são utilizados os dados que o [ECID](#) retorna, como data da predição e classes preditas, e as classificações de *McIntosh*, *Mount Wilson* coletadas no banco de dados da [NOAA](#) ([NOAA, 2020](#)). A junção do todo método agregador e a ontologia segue o Algoritmo 2.

---

**Algorithm 2** Método ECID Onto - Versão 2

---

```
Input: Classes preditas pelo ECID  
Result: Classes preditas pelo ECID Onto  
initialization;  
if Se valor de X estiver abaixo da média then  
  | Excluir classe X;  
if Se valor de M estiver abaixo da média then  
  | Excluir classe M;  
if Se valor de C estiver abaixo da média then  
  | Excluir classe C;  
if Se valor de AB estiver abaixo da média then  
  | Excluir classe AB;
```

Fonte: Os autores

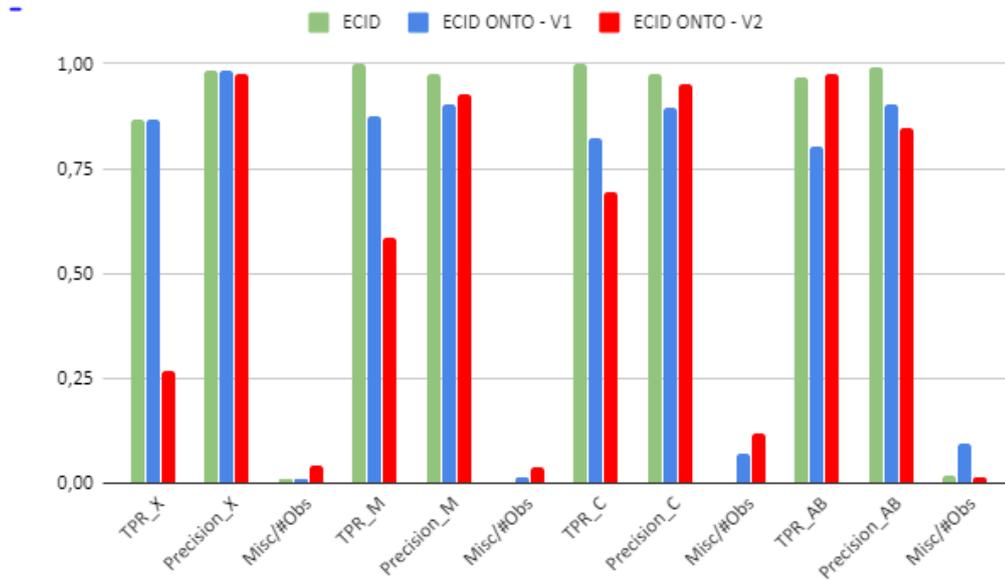
---

A ontologia retorna uma soma das três frequências relativas de cada classificação por classe predita pelo [ECID](#). Por exemplo, no dia 2010.05.29 foi predito as classes AB com classificação *McIntosh* igual a Hax e *Mount Wilson* igual a Beta, é somado as frequência relativas de cada classificação e retornado pela ontologia. A base original do [ECID](#) foi populada com essas somas e em seguida foi identificado a media de todas as classes de raio-x. Essa média é utilizada como critério de exclusão das classes seguindo o algoritmo 2. Para exemplificar o algoritmo, no dia 2010.05.07 foi predito as classes ABCM, e foi identificado que a soma retornada para a classe M nesse dia é melhor que a media estabelecida, então apenas a classe M será excluída.

### 5.3.2 Validação

Para a validação da ontologia versão dois, aplicou-se as mesmas métricas utilizadas pela primeira versão [ECID Onto](#). A Figura 21 apresenta os resultados obtidos nos experimentos das três versões.

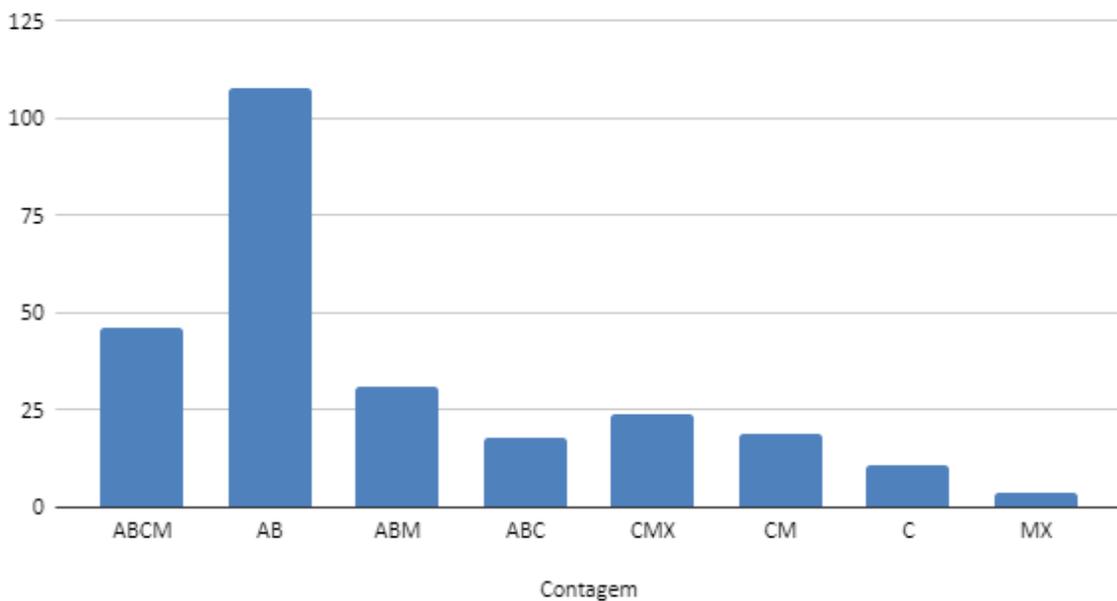
**Figura 21 – Experimentos do ECID, ECID *Onto* - Versão 1 e Versão 2**



Fonte: Os autores.

Como pode-se avaliar, ouve um decréscimo em todas as métricas comparando com o método ECID. A Figura 22 ilustra a distribuição das classes preditas pela nova versão apresentada.

**Figura 22 – Distribuição das classes preditas pelo ECID *Onto* - Versão 2**



Fonte: Os autores.

## 5.4 Considerações Finais

Nesse capítulo foi apresentado as duas metodologias para incorporar a ontologia *Onto Solar Flare* ao método ECID. Notou-se que os as métricas de TPR e precisão tiveram um

decréscimo em comparação com o **ECID**. Esse decaimento pode-se justificar pela melhor distribuição dos dados, o **ECID** prediz com grande frequência 3 das quatro possíveis classes a serem retornadas. A incorporação da ontologia, reduziu a quantidade de predições com três classes e as distribuiu entre duas ou uma única classe.

# Capítulo 6

## CONCLUSÃO

---

---

As pesquisas sobre predição da ocorrência de explosões solares, é um domínio de grande relevância, pois pode diminuir diversos prejuízos que a explosão pode causar. Porém é um domínio complexo de se explorar, pois os dados são desbalanceados, com terminologias que não são triviais para todos os pesquisadores envolvidos e não se sabe com precisão o que causa a reconexão magnética.

Motivado em contribuir com as pesquisas realizadas para prever a ocorrência de explosão, esse projeto de dissertação propôs fornecer dados semânticos organizados em uma ontologia de domínio denominada *Onto Solar Flare*. O foco da ontologia foi definir alguns termos para melhor comunicação entre os pesquisadores e aproveitar esses dados nos algoritmos de predição. Para o desenvolvimento da ontologia, optou-se por utilizar a metodologia [UPON](#), pois ela se baseia em uma metodologia iterativa e incremental que contribui com o processo de exploração do domínio. Essa metodologia propõe delimitar o escopo de desenvolvimento de acordo com o domínio de aplicação, ou seja, modelar e implementar apenas as informações que podem contribuir um o domínio de aplicação pré-definido. Foi definido como domínio de aplicação o método [ECID](#), pois ele foi desenvolvido no mesmo grupo de pesquisa desse projeto de pesquisa, e surgiu a necessidade do desenvolvimento da ontologia durante a concepção desse método.

Durante as reuniões com o especialista de domínio as informações que mais se destacaram foram as referentes a classificação de raio-x que mede a intensidade da explosão solar e informações sobre o ciclo solar, que está relacionado diretamente com as configurações magnéticas das [RA](#). Durante o levantamento das informações sobre o domínio de aplicação, constatou-se que essas informações poderiam ser úteis para enriquecer semanticamente o algoritmo [ECID](#).

O objetivo geral deste trabalho foi estruturar e aplicar uma ontologia de domínio sobre explosões solares, a partir de informações coletadas na literatura e com especialistas de domínio. Ele foi alcançado ao cumprir os seguintes objetivos específicos.

- Criar uma base de conhecimentos sobre explosões solares a partir da literatura e especialistas de domínio.

- Criar uma base de dados para dar suporte a predição.
- Adaptar o método de predição **ECID** para incorporar a ontologia *Onto Solar Flare*.

O primeiro objetivo específico foi alcançado no termino do desenvolvimento da ontologia de domínio *Onto Solar Flare*. Os dados citados no segundo objetivo são as informações imputadas na ontologia, por exemplo a frequências de ocorrências de explosões solares de acordo com a etapa do ciclo. E por fim foi incorporado a ontologia no método **ECID** gerando o **ECID Onto**. A ontologia contribuiu com uma melhor distribuição das classes preditas pelo **ECID** original.

Retomando as questões de pesquisa:

- É possível estruturar adequadamente os dados das explosões solares em uma ontologia para auxiliar o processo de predição de explosões solares?
- É possível usar a ontologia para melhorar a predição de explosões solares?

Conclui-se que é possível estruturar os dados em uma estrutura ontologia para auxiliar o processo de predição, porém a participação do especialista de domínio sobre explosões solares foi de suma importância, para nortear quais informações seriam relevantes. Conclui-se também que a ontologia pode gerar ganhos nos algoritmos de predição sobre explosões solares.

Os resultados obtidos na primeira versão da ontologia e do **ECID Onto** - versão 1, foram submetidos para o SBBD 2020.

Como trabalhos futuros, sugere-se melhor o processo de integração da ontologia *Onto Solar Flare* no método **ECID** a fim de encontrar melhores resultados. Sugere-se também incorporar a ontologia em outras etapas do **ECID** e/ou em outros algoritmos de predição. Sobre a ontologia, os próximos passos são, validar as informações com outros especialistas de domínio e incorporar mais informações que possam ser úteis.

## REFERÊNCIAS

---

- ALATRISH, E. Comparison some of ontology editors. *Management Information Systems*, v. 8, p. 18–24, 2013. Citado 2 vezes nas páginas 20 e 21.
- BALDÁRRAGO, A. E. U. *Abordagem de Recomendação Baseada em Conteúdo Utilizando Ontologia Fuzzy de Domínio e Ontologia Crisp de Preferência do Usuário*. Dissertação (Mestrado) — Universidade Federal de São Carlos, 9 2012. Citado na página 21.
- BASU, S.; BASU, S.; MACKENZIE, E.; BRIDGWOOD, C.; VALLADARES, C.; GROVES, K.; CARRANO, C. Specification of the occurrence of equatorial ionospheric scintillations during the main phase of large magnetic storms within solar cycle 23. *Radio Science*, v. 45, p. 5009–, 10 2010. Citado na página 12.
- BENTLEY, R.; BROOKE, J.; CSILLAGHY, A.; FELLOWS, D.; BLANC, A. L.; MESSEROTTI, M.; PEREZ-SUAREZ, D.; PIERANTONI, G.; SOLDATI, M. Helio: Discovery and analysis of data in heliophysics. *Future Generation Computer Systems*, 11 2013. Citado na página 26.
- BENTLEY, R.; CSILLAGHY, A.; ABOUDARHAM, J.; JACQUEY, C.; HAPGOOD, M.; BOCCHIALINI, K.; MESSEROTTI, M.; BROOKE, J.; GALLAGHER, P.; FOX, P.; HURLBURT, N.; ROBERTS, D.; DUARTE, L. S. Helio: The heliophysics integrated observatory. *Advances in Space Research*, v. 47, n. 12, p. 2235 – 2239, 2011. ISSN 0273-1177. Recent Advances in Space Weather Monitoring, Modelling, and Forecasting - 2. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0273117710001110>>. Citado na página 26.
- BENZ, A. Flare observations. *Living Reviews in Solar Physics*, v. 14, 12 2017. Citado na página 12.
- BIZER, C. The dbpedia data provision architecture. dbpedia. 2009. Disponível em: <<http://wiki.dbpedia.org/Architecture?v=14cg>>. Acesso em: 1 mar. 2019. Citado na página 17.
- BREITMAN, K. K. *WebSemântica: a internet do futuro*. [S.l.]: LTC, 2005. ISSN 978-8521614661. Citado na página 20.
- BURTON-JONES, A.; STOREY, V. C.; SUGUMARAN, V.; AHLUWALIA, P. A semiotic metrics suite for assessing the quality of ontologies. *Data Knowl. Eng.*, v. 55, p. 84–102, 2005. Citado 2 vezes nas páginas 32 e 37.
- CALERO, C.; RUIZ, F.; PIATTINI, M. *Ontologies for Software Engineering and Software Technology*. [S.l.]: Springer, 2006. 119 p. ISSN 9783540345183. Citado na página 19.
- CYGANIAK, R.; WOOD, D.; LANTHALER, M. Rdf 1.1 concepts and abstract syntax. *W3C Recommendation*, 2014. Disponível em: <<https://www.w3.org/TR/rdf11-concepts/>>. Acesso em: 18 jan. 2019. Citado na página 19.

FENSEL, D. *Ontologies: silver bullet for knowledge management and electronic commerce*. [S.l.]: Springer-Verlag Berlin Heidelberg, 2001. 136 p. ISBN 978-3-662-04396-7. Citado na página 16.

FOX, P.; MCGUINNESS, D.; CINQUINI, L.; WEST, P.; GARCIA, J.; BENEDICT, J.; MIDDLETON, D. Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience. *Computers Geosciences*, v. 35, p. 724–738, 04 2009. Citado 2 vezes nas páginas 26 e 27.

GOMEZ, A. P.; FERNANDEZ, M. L.; CORCHO, O. *Ontological engineering*. Springer-Verlag, London, 2003. Citado na página 17.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, v. 43, n. 5, p. 907 – 928, 1995. ISSN 1071-5819. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1071581985710816>>. Citado na página 16.

HERNANDES, E. C. M. *Um processo automatizado para tratamento de dados e conceitualização de ontologias com apoio de visualização*. Dissertação (Mestrado) — Universidade Federal de São Carlos, 8 2009. Citado na página 16.

HOLMAN, G. The mysterious origins of solar flares. *Scientific American*, v. 294, p. 38–45, 05 2006. Citado na página 12.

HORROCKS, I.; PATEL-SCHNEIDER, P. F.; HARMELEN, F. van. From shiq and rdf to owl: the making of a web ontology language. *Journal of Web Semantics*, v. 1, n. 1, p. 7–26, 2003. ISSN 1570-8268. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1570826803000027>>. Acesso em: 1 mar. 2019. Citado na página 18.

ISOTANI, S.; BITTENCOURT, I. I. *Dados Abertos Conectados*. São Paulo: Novatec, 2015. 176 p. ISBN 978-85-7522-449-6. Citado 5 vezes nas páginas 16, 17, 18, 19 e 20.

JUNIOR, S. D.; CECATTO, J.; FERNANDES, M.; RIBEIRO, M. Seminer: A flexible sequence miner method to forecast solar time series. *Information*, v. 9, p. 8, 01 2018. Citado na página 23.

JUNIOR, S. D.; CECATTO, J.; FERNANDES, M.; RIBEIRO, M. Handling imbalanced time series through ensemble of classifiers: A multi-class approach for solar flare forecasting. In: \_\_\_\_\_. [S.l.: s.n.], 2019. p. 209–214. ISBN 978-3-030-14069-4. Citado 4 vezes nas páginas 22, 23, 24 e 46.

KING, T.; THIEMAN, J.; ROBERTS, D. Spase 2.0: A standard data model for space physics. *Earth Science Informatics*, v. 3, p. 67–73, 06 2010. Citado na página 26.

KIRYAKON, A. Ontologies for knowledge management. In: JOHN MILEY. *Semantic Web Technologies: trends and research in ontology-based systems*. Chichester, 2006. p. 115–138. Citado na página 17.

LOSKUTOV, A.; ISTOMIN, I.; KUZANYAN, K.; KOTLYAROV, O. Testing and forecasting the time series of the solar activity by singular spectrum analysis. *arXiv preprint nlin/0010027*, 2001. Acesso em: 1 mar. 2019. Citado na página 29.

MCGUINNESS, D. L.; HARMELEN, F. Owl web ontology language overview. *W3C Recommendation*, p. 1–22, 2 2004. ISSN 15302180. Citado na página 19.

- MESSEROTTI, M. Solar activity: Exploration, understanding and prediction. *Solar Activity: Exploration, Understanding and Prediction*, 2007. Citado 2 vezes nas páginas 25 e 26.
- MIZOGUCHI, R. Part 3: Advanced course of ontological engineering. *New Generation Computing*, v. 22, n. 2, p. 193–220, Jun 2004. ISSN 1882-7055. Disponível em: <<https://doi.org/10.1007/BF03040960>>. Acesso em: 1 mar. 2019. Citado na página 17.
- NASA SCIENCE. *A New Kind of Solar Storm*. 2016. Disponível em: <[https://science.nasa.gov/science-news/science-at-nasa/2005/10jun\\_newstorm/](https://science.nasa.gov/science-news/science-at-nasa/2005/10jun_newstorm/)>. Acesso em: 1 mar. 2019. Citado na página 29.
- NICOLA, A. D.; MISSIKOFF, M.; NAVIGLI, R. A software engineering approach to ontology building. *Information Systems*, v. 34, n. 2, p. 258 – 275, 2009. ISSN 0306-4379. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306437908000628>>. Acesso em: 1 mar. 2019. Citado 2 vezes nas páginas 13 e 21.
- NOAA. *Índice de pub warehouse*. 2020. Disponível em: <<ftp://ftp.swpc.noaa.gov/pub/warehouse/>>. Acesso em: 04 agosto 2020. Citado 2 vezes nas páginas 41 e 49.
- NOY, N. F.; MCGUINNESS, D. L. Ontology development 101: A guide to creating y our first ontology. 2001. Disponível em: <[https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf)>. Acesso em: 17 jan. 2019. Citado na página 12.
- PATEL-SCHNEIDER, P. F. Building the semantic web tower from rdf straw. *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, p. 546–551, 1 2005. Acesso em: 1 mar. 2019. Citado na página 18.
- SCHRIJVER, C. J.; BAGENAL, F.; SOJKA, J. Heliophysics v. space weather and society early chapter collection v. january 5, 2015. 2015. Citado na página 12.
- USCHOLD, M.; GRÜNINGER, M. Ontologies and semantics for seamless connectivity. *SIGMOD Record*, v. 33, p. 58–64, 12 2004. Citado na página 18.