

# UMA ARQUITETURA PARA EXTRAÇÃO DE UNIDADES DE PRODUTOS EM LOJAS VIRTUAIS BRASILEIRAS

Marcio Kolling<sup>1</sup>

Gilberto Irajá Müller<sup>2</sup>

**CONTEXTO:** O e-commerce cresce a cada ano pela facilidade que oferece ao e-consumidor em consultar as melhores ofertas e efetuar suas compras on-line. Para possibilitar essa tarefa, há diversos processos nos sites de busca utilizados pelos e-consumidores, permitindo a comparação de produtos similares. **PROBLEMA:** Essa tarefa de comparação de produtos similares se torna um desafio devido à falta de padronização das informações cadastradas na WEB. Na maioria dos casos, as informações sobre atributos de produtos são descritas de forma livre, sem formatação e diretamente na descrição do produto, dificultando o estabelecimento de um padrão para comparação. Outro fator importante é o crescimento do e-commerce brasileiro e, conseqüentemente, a necessidade em simplificar o processo de análise do e-consumidor. **OBJETIVO:** Devido aos problemas apresentados, este estudo apresenta uma arquitetura para extração automatizada de unidades de produtos e dois protótipos, sendo o primeiro de um motor de extração de atributos de produtos e, o segundo, de um protótipo WEB para apresentação e comparação de preços de produtos por loja virtual. **MÉTODO PROPOSTO:** Para atingir o objetivo deste estudo, optou-se pelo método de pesquisa do tipo experimentação, onde foi comparado o protótipo de extração com o estudo correlato de Horch, Kett e Weisbecker (2015). Ainda, foi utilizada a prototipação para o desenvolvimento do protótipo WEB para mostrar a aplicação do resultado da extração de atributos de produtos. Para a validação do protótipo WEB, foi aplicado o modelo de aceitação TAM (Technology Acceptance Model). **CONCLUSÃO:** Por fim, apresentam-se a análise e discussão dos resultados, além da aceitação do protótipo, mostrando que a arquitetura é viável e útil para o e-commerce.

Palavras-chave: WEB Semântica. Ontologia. OWL (Web Ontology Language). E-commerce. TAM (Technology Acceptance Model). Extração de Unidades.

## 1 INTRODUÇÃO

Existem, basicamente, dois principais tipos de fonte de pesquisa de produtos on-line, sendo a primeira, sites de busca como a Google e a Yahoo, e a segunda, sites específicos para comparação de preços como Buscapé e Bondfaro. Ambos dependem dos dados fornecidos por sua fonte e que em alguns casos podem ser estruturadas, mas dificilmente terão um padrão, fazendo uso de sua própria estrutura de dados. (BALOG, 2011).

Balog (2011) afirma que diversas lojas on-line apresentam o mesmo produto com diferentes estruturas de formatação de suas características, sendo que algumas apresentam toda informação em um campo específico, outras podem colocar a mesma informação em atributos específicos como: fabricante, marca, cor, entre outros, além de registros com estruturas de formatação diferentes. Essa falta de padronização pode não fazer diferença para um e-consumidor

---

<sup>1</sup>Aluno do curso de Sistemas de Informação. Email: kolling.marcio@gmail.com

<sup>2</sup>Orientador, professor Mestre da Unisinos, Mestre em Computação Aplicada pela mesma Universidade (2006). Email: gimuller@unisinos.br

que está analisando um produto, mas para os processos automatizados (computadorizados) de extração de dados é um grande empecilho, dificultando a comparação de mesmos produtos com diferentes identificações.

Com o advento da informatização, a necessidade de integrar e relacionar dados de diferentes fontes está cada vez maior, e devido à falta de padronização essa é uma tarefa complexa. Para melhorar o registro e interpretação desses dados facilitando seu relacionamento, é fundamental o uso de ontologia como apoio a esse processo. (RIJGERSBERG; ASSEM; TOP, 2013)

Guarido, Oberle e Staab (2009) afirmam que as ontologias computacionais são usadas para modelar formalmente estruturas (Ex.: um produto e suas características). Devido às suas características, as ontologias têm papel fundamental para a representação padronizada de um produto que originalmente está disposto sem uma formatação comum.

Do ponto de vista de crescimento do e-commerce e, conseqüentemente, aumento dos dados no Brasil, segundo a E-BIT (2019), o e-commerce brasileiro voltou a crescer dois dígitos em 2018, alcançando 12% de crescimento em relação a 2017, com um faturamento de R\$ 53,2 bilhões e 58 milhões de compradores que fizeram pelo menos uma compra on-line em 2018. Com o contínuo crescimento do e-commerce, disponibilidade de lojas e ofertas de produtos on-line e a falta de padronização da apresentação de registros de produtos, é grande o desafio em disponibilizar um resultado satisfatório na identificação de atributos de produtos similares por meio dos dados disponibilizados na WEB com a finalidade de comparação de preços.

## 1.1 Contexto desta Pesquisa

Este estudo é guiado pela seguinte questão desta pesquisa: **qual a relevância de uma arquitetura de extração e comparação de produtos em lojas on-line?** Para responder essa questão os seguintes objetivos foram realizados: (i) desenvolvimento de um protótipo de um motor de extração de atributos dos produtos; (ii) estudo comparativo do cenário brasileiro com o cenário apresentado em Horch, Kett e Weisbecker (2015), e (iii) desenvolvimento de um protótipo WEB para apresentação e comparação de preços de produtos por loja virtual para mostrar a aplicação da arquitetura.

Este estudo se justifica pela consolidação do e-commerce no mercado brasileiro e aumento da oferta de produtos em todas as categorias, gerando concorrência entre lojistas. Além disso, tem-se o e-consumidor cada vez mais inserido no mundo digital utilizando ferramentas para encontrar seus produtos e, principalmente, encontrá-los com o menor preço entre as inúmeras ofertas disponíveis em diversas lojas on-line.

Salienta-se que neste estudo não foi abordado o mecanismo de leitura das informações de produtos das lojas virtuais e a persistência dos mesmos em banco de dados.

## 1.2 Organização deste Estudo

Este artigo está organizado em seis seções. Na Seção 2, é apresentada a Fundamentação Teórica, onde aborda-se os conceitos relevantes para esse estudo, sendo eles: WEB semântica, ontologias e métodos de extração de dados. A Seção 3 apresenta os Trabalhos Relacionados, com principal enfoque ao estudo europeu de Horch, Kett e Weisbecker (2015), que serviu como base para esse estudo. Na Seção 4, é apresentada a metodologia para validação deste estudo. A Seção 5 apresenta o Protótipo e Tecnologias Utilizadas como a arquitetura do protótipo, descrição das tecnologias e processo de funcionamento. A Seção 6 aborda a apresentação, análise e discussão de resultados. Por fim, na Seção 7, apresenta-se as conclusões deste estudo e sugestões de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 WEB SEMÂNTICA

Conforme Berners-Lee, Hendler e Lassila (2001), "A Web Semântica não é uma Web separada, mas uma extensão da atual, em que é dado significado bem definido a informação. Adicionar semântica na Web envolve permitir que documentos tenham informações em formas legíveis [...]".

A WEB semântica é um projeto patrocinado pelo W3C<sup>3</sup> e tem como objetivo atribuir inteligência e contexto nos códigos XML embutidos nas páginas WEB, para melhorar a interação dos programas com páginas WEB tornando mais intuitivo o uso pelos usuários. (SOUZA; ALVARENGA, 2004).

Segundo Freitas (2003), a WEB provocou uma vasta diversidade de aplicativos e informações para as pessoas, possibilitando até o comércio eletrônico entre empresas e clientes, conhecido como B2C (Business-to-Consumer). Entretanto, mesmo os robôs (como motor de busca) dotados de inteligência se deparam com um ambiente hostil para a realização de suas tarefas, tudo devido as páginas possuírem informações sem padronização.

Uma das aplicações mais populares da WEB são máquinas de busca; esses motores enfrentam dificuldades para a extração de informações na WEB, pois tem uma compreensão limitada das informações contidas nas diversas páginas da WEB. Como existem diversos dados distribuídos de formatos diferentes em páginas distintas, são necessários algoritmos que tratem essa informação, originalmente apresentada para uma comunicação com seres humanos. (LAUFER, 2015).

Desse modo, surgiu o desafio de como identificar todos esses dados sem uma identificação específica para máquinas, assim como formatar informações na WEB de forma que possam

---

<sup>3</sup>W3C (World Wide Web Consortium) é um consórcio internacional que tem a finalidade de estabelecer padrões para a criação e a interpretação de conteúdos para a Web. Site: <<https://www.w3.org/>>

ser úteis para máquinas, além de distribuir de novas formas esses dados. Para alcançar esses objetivos, surgiu a WEB semântica, alcançando a ideia de uma WEB de Dados, entendida por serem humanos e por máquinas. Sites de comparação de preços são um exemplo da aplicação da WEB Semântica, pois utilizam programas específicos chamados scrapers, que extraem os dados de produtos estruturados dentro do HTML das páginas WEB, identificando assim, padrões sintáticos. (LAUFER, 2015).

Isotani et al. (2009) afirma que "a Web Semântica propõe tecnologias que dão mais "poder" aos computadores através da criação de padrões, protocolos e linguagens formais que facilitam o processamento da informação na Web de forma automática e inteligente".

A ideia da WEB semântica é de que seja possível a identificação de cada um dos dados agrupados por páginas WEB de forma individual e legível por máquinas. Portanto, é necessária a inclusão de metadados, que são informações extras sobre os dados dentro do código HTML, para que sejam consumidas por máquinas. (LAUFER, 2015).

### 2.1.1 Metadados

Souza e Alvarenga (2004) explicam que para atender o propósito da WEB semântica, é necessário que as tecnologias de linguagens e metadados descritivos sejam padronizadas, fazendo com que todos os usuários da WEB sigam certas regras comuns, orientando dessa forma, o armazenamento e a descrição da informação.

Laufer (2015) explica que metadados são dados sobre dados, que fornecem informações adicionais sobre os mesmos, auxiliando desenvolvedores de aplicações e usuários a compreenderem o significado dos dados publicados, o seu conteúdo e sua estrutura. Os metadados auxiliam no processo de comunicação entre quem publica o dado e quem o consome, para que seja possível a compreensão de todas as características pertinentes à utilização desses dados. Os metadados podem ser mesclados dentro do código HTML de uma página WEB contendo informações destinadas aos usuários e outra parte ao consumo por máquinas. Metadados também podem ser armazenados em catálogos contendo informações de dados publicados na WEB.

As informações hoje apresentadas nos metadados fazem parte de uma arquitetura onde os dados são publicados de diferentes formas e com diferentes aplicações, sendo WEB Services e WEB APIs que manipulam esses dados para gerar os resultados a serem apresentados para o consumidor. Porém, não existe uma padronização quanto à estrutura e semântica desses dados e a interpretação da estrutura e semântica ficam interiorizadas nas aplicações que consumirão os dados. A inclusão de semântica a esses dados publicados em diversos formatos e com diferentes protocolos de acesso tem como objetivo facilitar a interpretação e comunicação transparente entre os dados desse universo de informações desconexas. (LAUFER, 2015).

Segundo Souza e Alvarenga (2004), o primeiro passo para atender o objetivo da WEB Semântica é a criação de padrões de descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados. A seguir apresentam-se algumas dessas

linguagens e padrões.

### 2.1.2 Linguagens e Padrões

A linguagem mais utilizada é o HTML (HyperText Markup Language), sendo sua estrutura composta por tags que são marcações sintáticas para descrever os dados e comandos que manipulam o documento. Essa linguagem foi criada para atender à necessidade da construção de documentos para sua exibição na WEB, por isso, tem o foco em tratar o formato que os dados presentes no documento serão exibidos no navegador. (SOUZA; ALVARENGA, 2004). Na Figura 1 é apresentado um exemplo de código HTML.

Figura 1: HTML

```
<!DOCTYPE html>
<html>
<head>
  <meta charset="UTF-8"/>
  <title>Document</title>
</head>
<body>
  <!-- Conteúdo -->
</body>
</html>
```

Fonte: Elaborada pelo Autor, com base em Devmidia (2010).

Segundo Freitas (2003), o HTML tem como limitação a falta de recursos na linguagem para anotação semântica, pois seus recursos são apenas para formatação dos dados a serem apresentados. Para descrever algum tipo de semântica, o HTML traz tags opcionais como título, descrição, sumário e palavra-chave e, devido a essa limitação, para atender melhor aos objetivos para anotação semântica, o HTML foi abstraído para XML (eXtensible Markup Language).

Souza e Alvarenga (2004) explicam que o XML é recomendado pelo W3C e tem algumas semelhanças com o HTML, sendo ambos derivados do SGML (Standard Generalized Markup Language), que é uma metalinguagem para descrever outras linguagens. Enquanto o HTML tem como objetivo controlar a forma de apresentação dos dados, o XML foca na descrição dos dados que o documento contém, sendo flexível quando adiciona-se novas tags de acordo com a necessidade, desde que estejam descritas em um documento padrão de validação, como o DTD (Document Type Definition) ou o XSD (XML Schema Definition).

Apesar do XML ter a flexibilidade na criação de novas tags, ocorre problemas de identificação semântica como ambiguidades. Segundo Souza e Alvarenga (2004), para tratar esses problemas de identificação semântica foram criados no projeto da WEB semântica, padrões de metadados e de construção de XML, como o Dublin Core, que é uma iniciativa para a criação de um vocabulário controlado para uso na WEB, que se baseia na presunção de que a busca por informações não deve depender do meio em que estão armazenadas. Na Figura 2 é apresentado um exemplo de um texto e sua correspondência em código XML.

Figura 2: Texto e seu correspondente em XML

Catálogo de endereços	<?xml version="1.0"?>
João Silva	<catálogo de endereços>
Rua Carijós, 135	<entrada>
Belo Horizonte, MG 30.000	<nome> João Silva </nome>
Brasil	<endereço>
31 3335-5556 (preferido)	<rua> Carijós, 135</rua>
31 3549-4446	<estado> MG </estado>
<a href="mailto:joaosilva@net.com.br">joaosilva@net.com.br</a>	<cep> 30.000 </cep>
José Almeida	<pais> Brasil </pais>
<a href="mailto:jalmeida@net.com.br">jalmeida@net.com.br</a>	<endereço>
	<telefone preferido="true">31 3335-4456</telefone>
	<telefone> 31 3594-4446 </telefone>
	<email> <a href="mailto:joaosilva@net.com.br">joaosilva@net.com.br</a> </email>
	</entrada>
	<entrada>
	<nome><prim>José</prim>
	<sobren>Almeida</sobren>
	<email> <a href="mailto:jalmeida@net.com.br">jalmeida@net.com.br</a> </email>
	</entrada>
	</catálogo de endereço>

Fonte: Almeida (2002).

Para fornecer suporte à semântica, o RDF (Resource Description Framework) estabelece um padrão de metadados a ser embutido na codificação de um XML, pois sua ideia é descrever dados e metadados através de um esquema de “triplas” de recurso, propriedade e valor. O RDF possui uma forma racional de acesso aos padrões de metadados publicados na WEB, como, por exemplo, o Dublin Core ou qualquer outro namespace compartilhado. (SOUZA; ALVARENGA, 2004).

Fensel (2000) afirma que o RDF traz mais semântica a um documento, pois não necessita se referir a sua estrutura. RDF é uma recomendação do W3C para que seja implementada na construção de páginas da WEB semântica.

Souza e Alvarenga (2004) apresentam na Figura 3 um exemplo de código XML que apresenta um padrão de metadados RDF para um cartão de visita.

Figura 3: RDF

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdisyntax.ns#"
  xmlns:dc="http://purl.ocic.org/DC#"
  xmlns:v="http://www.w3.org/2001/vcard-rdf/3.0#"
  <rdf:Description about=" http://www.ucla.edu/einstein" />
    <dc:Creator>
      <rdf:Description about=" http://www.ucla.edu/staff/einstein"/>
        <v:Name> Isaac Einstein</v:Name>
        <v:Email="einstein@ucla.edu"/>
        <v:Orgname>UCLA</v:Orgname>
        <v:Orgunit>Department of Physics</v:Orgunit>
      </rdf:Description>
    </dc:Creator>
  </rdf:Description>
</rdf:RDF>
```

Fonte: Elaborada pelo Autor, com base em Souza e Alvarenga (2004).

Nas primeiras linhas do XML da Figura 3 são declarados os namespaces utilizados no documento XML, sendo eles o padrão RDF, Dublin Core e o namespace utilizado para a espe-

cificação do cartão de visita, que faz a padronização das informações usadas para representar tal recurso. Com os namespaces definidos, os descritores são utilizados de acordo com sua definição, evitando assim, ambiguidades.

Com o padrão RDF, as ontologias e os namespaces de qualquer publicação de informações em páginas WEB são feitas garantindo que máquinas tenham a capacidade de interpretar as informações de forma semântica garantindo o aproveitamento inteligente da informação.

## 2.2 ONTOLOGIAS

Roque (2003) define ontologia como “uma descrição compartilhada de conceitos e relações de um domínio, expressas em uma linguagem legível por computador”.

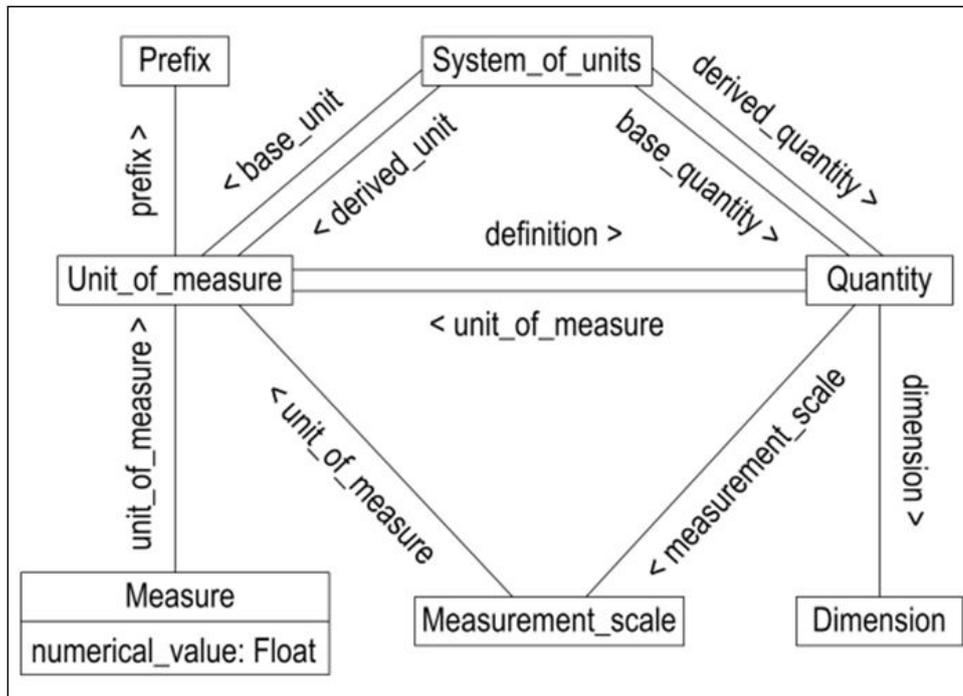
A aplicação de ontologias se adapta à falta de soluções para captar a semântica das páginas da WEB, sendo essa tecnologia uma forma de estruturação de informações soltas na rede. A ontologia pode ser vista como uma etapa importante na evolução da descrição do conhecimento. (FREITAS, 2003).

Pickler (2007) explica que as ontologias são usadas para a atribuição de sentido e significado dentro de seus contextos, com o objetivo de atribuir semântica ao conteúdo dos documentos, atuando como um meio de representação do conhecimento. Souza e Alvarenga (2004) complementam afirmando que o objetivo da construção de ontologias se dá pela necessidade do compartilhamento de um vocabulário, a fim de possibilitar a troca de informações entre os agentes da rede WEB, sejam eles máquinas ou humanos.

Roque (2003) explica que para a definição de ontologias existe uma variedade de linguagens representativas. A linguagem formal que interessa para esse estudo é a OWL (Web Ontology Language), sendo essa uma linguagem orientada à ontologia. Roque (2003) explica que a linguagem foi projetada pelo W3C Web - Ontology Working Group - como uma revisão do DAML (DARPA Agent Markup Language) e OIL (Ontology Inference Layer). A OWL foi projetada para se tornar uma linguagem de ontologia padrão para a WEB, que é uma extensão do RDF e RDFS (RDF Schema).

Além dos fundamentos sobre OWL, é necessário compreender o trabalho realizado por Rijgersberg, Assem e Top (2013), que desenvolveram uma ontologia de unidades de medida (OM) utilizando a linguagem OWL. A OM desenvolvida possibilita uma melhor compreensão de dados quantitativos, viabilizando sua integração e comparação. Na Figura 4 é apresentado um diagrama de classes simplificado de uma ontologia de unidades de medida.

Figura 4: Diagrama de Classes simplificado de uma OM

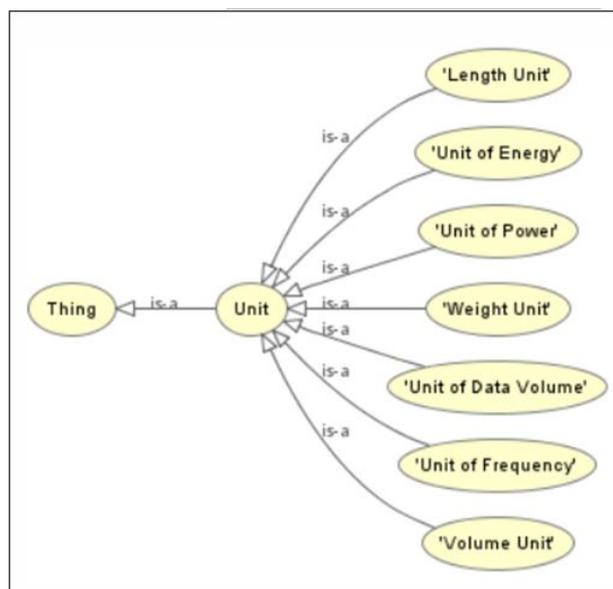


Fonte: Rijgersberg, Assem e Top (2013).

Segundo Horch, Kett e Weisbecker (2015), a OM ilustrada na Figura 4 possui todas as ferramentas necessárias para o processo de extração de atributos de produtos, bem como para a conversão de unidades de medida para uma unidade em comum. Nesse processo, a ontologia é utilizada para identificar a relação entre as unidades de medida, realizando a conversão dessas unidades para uma unidade em comum, possibilitando assim, a identificação de produtos correspondentes, ou seja, com as mesmas características.

Esse estudo utilizou como base a ontologia desenvolvida por Horch, Kett e Weisbecker (2015), sendo que esta possui sete classes para diferentes tipos de unidades conforme apresentado na Figura 5.

Figura 5: Diagrama da ontologia de unidades



Fonte: Horch, Kett e Weisbecker (2015).

Cada classe apresentada na Figura 5 possui indivíduos que representam diferentes unidades de medida dentro de cada classe, como, por exemplo, na classe LengthUnit: Millimeter, Centimeter, Inch, Meter, Foot, Yard, Kilometer e Mile.

### 3 TRABALHOS RELACIONADOS

Diversas pesquisas sobre o tema de extração de dados utilizando ontologias foram realizadas. Entre elas, pode-se citar o trabalho realizado por Holzinger, Krüpl e Herzog (2006)<sup>4</sup>, onde foi desenvolvida uma abordagem para a extração de características de produtos a partir de páginas da WEB, com o uso de duas ontologias, sendo a primeira para a especificação da estrutura das tabelas e atributos do produto, e a segunda, específica para o domínio.

O estudo de Holzinger, Krüpl e Herzog (2006) utilizou uma abordagem visual dos dados da página WEB para identificar a tabela de produtos e transferir seus dados de descrição explícita para uma ontologia. Essa descrição explícita é registrada com tipos e palavras-chaves na ontologia de domínio. E por último, um algoritmo analisa e interpreta os dados de ambas ontologias, gerando como resultado uma ontologia OWL, com a informação extraída do produto.

Também, pode-se citar o estudo de Su, Wang e Lochovsky (2009) onde foi desenvolvido um método para extração de dados de páginas da WEB, assistido por uma ontologia de domínio (ODE). A ontologia é montada automaticamente com a combinação de interfaces de consulta e diferentes páginas de resultado de um domínio específico. Para a extração dos resultados,

<sup>4</sup>HOLZINGER, Wolfgang, KRÜPL, Bernhard, HERZOG, Marcus. Using ontologies for extracting product features from web pages. in Interational Semantic Web Conference, ser. Lecture Notes in Computer Science, I. F. Cruz, vol. 4273. Springer, 2006, pp. 286-299.

a ODE procura por subárvores dentro da árvore de elementos HTML, que tenham a melhor correspondência com a ontologia de domínio.

Este estudo tem como base o artigo “Extracting product unit attributes from product offers by using an ontology” desenvolvido por Horch, Kett e Weisbecker (2015), portanto, a seguir apresenta-se um resumo sobre o trabalho.

### 3.1 Extracting Product Unit Attributes from Product Offers by Using an Ontology

Horch, Kett e Weisbecker (2015) apresentam dados estatísticos do crescimento do mercado e-commerce europeu no ano de 2014. Também abordam o crescimento da disponibilidade de diversos produtos on-line, criando uma necessidade por mecanismos de pesquisa e comparação de produtos. Afirmam que o desafio para a comparação automatizada de preços de produtos é a identificação do produto, que se dá em função de diferentes ofertas formatadas em diversos modos e que representam o mesmo produto. Para a correta identificação de um produto, é necessária a extração dos atributos a partir de sua descrição, tarefa desafiadora, devido à descrição livre das informações dos produtos.

O estudo aponta que a extração dos atributos de produtos a partir de sua descrição ainda não é suficiente para o processo de comparação, pois diferentes anúncios podem ter diferentes unidades de medida informadas para seus produtos. Para possibilitar a correta comparação de produtos, após a extração de seus atributos, é necessário a conversão das unidades de medida, a fim de chegar a uma medida em comum.

O objetivo do estudo de Horch, Kett e Weisbecker (2015) é uma abordagem para a extração e conversão de atributos de produtos do e-commerce por meio do desenvolvimento de uma ontologia de unidades de medida, sendo que o resultado foi a extração de atributos em que a comparação de produtos fosse a melhor forma possível, na mesma unidade de medida.

Em relação aos materiais e métodos, foi desenvolvido um motor de extração e conversão de atributos de produtos utilizando como base a ontologia de unidades de medidas. A abordagem proposta foi validada através de um experimento. Esse experimento foi realizado com base em uma lista de 300 produtos previamente extraídos de três sites de e-commerce europeus. Os sites selecionados são divididos em três categorias, sendo elas: eletrônicos, cosméticos e alimentos.

O experimento teve como objetivo medir a correta extração de todas as unidades de medida e seus valores da lista de produtos da amostra, considerando somente as unidades de medida que estavam previstas na ontologia desenvolvida.

Na análise, as autoras explicam que, dentre a amostra de 300 produtos, 221 produtos continham ao menos uma unidade de medida, sendo que o total de atributos desses produtos foi de 276; desse total foram extraídos 213 atributos corretamente (positivos verdadeiros), equivalente a uma taxa de precisão de 99,07% ( $\frac{\text{positivos verdadeiros (213)}}{\text{positivos verdadeiros (213) + falsos positivos (2)}}$ ), sendo que 2 atributos foram extraídos de forma incorreta (falso positivo). O experimento teve um recall de 77,74% ( $\frac{\text{positivos verdadeiros (213)}}{\text{positivos verdadeiros (213) + falsos negativos (61)}}$ ) e uma taxa de extração correta de 77,17%

( $\frac{\text{positivos verdadeiros (213)}}{\text{total de atributos (276)}}$ ), que por sua vez apresentou uma taxa de não extração ou extração incorreta de 22,83% (63 atributos).

Horch, Kett e Weisbecker (2015) concluíram que o resultado do experimento apresentou uma boa adequação à proposta da nova abordagem para a extração de unidades de atributos de produtos. A partir dos resultados do experimento, as autoras propuseram melhorias na abordagem com a evolução das expressões regulares utilizadas para a extração de unidades dentro da ontologia, reduzindo assim a taxa de falsos negativos na identificação das unidades de medida.

#### 4 MATERIAIS E MÉTODOS

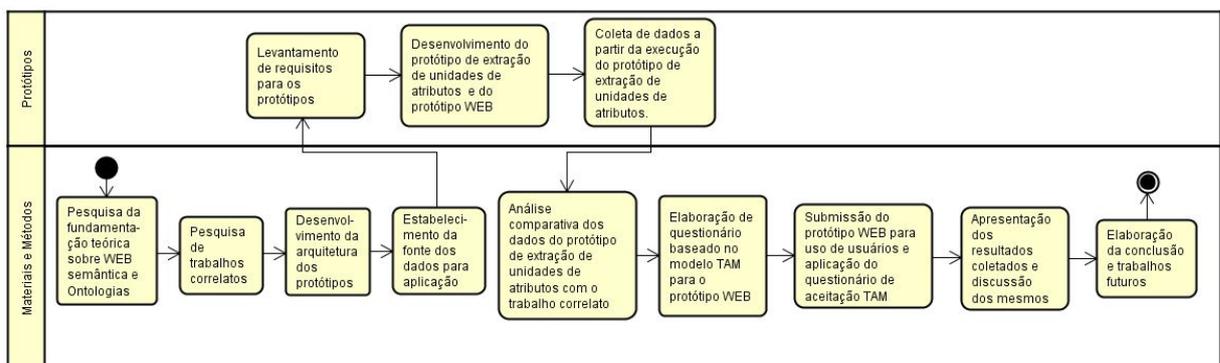
Para atingir os objetivos deste estudo foi definida a natureza como aplicada, pois tem como objetivo gerar conhecimento com a aplicação prática, dirigido à solução de um problema específico relacionado à WEB semântica. (GERHARDT; SILVEIRA, 2009).

A abordagem da pesquisa é quantitativa por meio da coleta dos dados a partir do uso dos protótipos e de questionários. Segundo Gerhardt e Silveira (2009), umas das características da abordagem quantitativa é o foco na objetividade, coleta e análise dos dados, o que se enquadra no objetivo desse estudo.

O método de pesquisa empregado neste estudo foi a explicativa. Segundo Gil (2007) a preocupação deste tipo de pesquisa é na identificação de fatores determinantes para a ocorrência de fenômenos, o que se enquadra neste estudo, pois por meio da experimentação e coleta de dados foi realizada uma comparação com o cenário apresentado por Horch, Kett e Weisbecker (2015). Além disso, foi aplicada a experimentação com base nos mesmos segmentos do estudo de Horch, Kett e Weisbecker (2015), fazendo-se um estudo comparativo.

A Figura 6 ilustra o desenho da metodologia que representa a Metodologia de Pesquisa utilizada neste estudo, além do desenvolvimento dos protótipos.

Figura 6: Desenho da Metodologia de Pesquisa



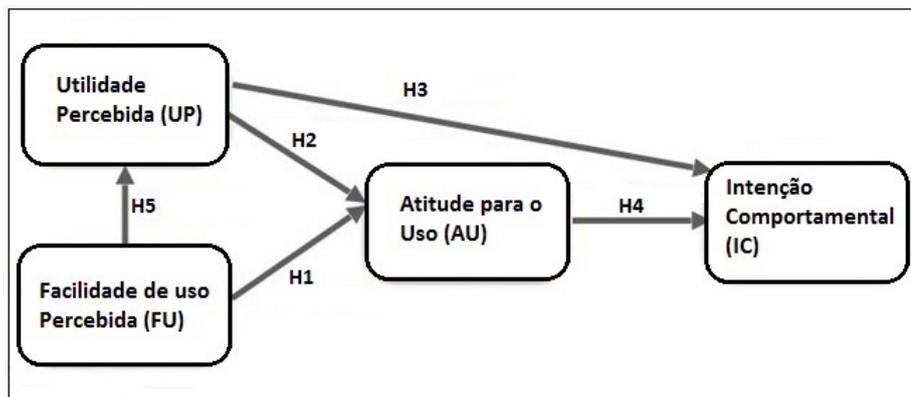
Fonte: Elaborada pelo autor.

De acordo com a Figura 6, a pesquisa para fundamentação teórica teve como foco a WEB Semântica e ontologias, temas fundamentais para o estudo. Já a pesquisa de trabalhos correla-

tos teve como base a WEB Semântica e extração de atributos de produtos de páginas da WEB, tendo como principal trabalho relacionado o estudo de Horch, Kett e Weisbecker (2015). Também foram levantados os requisitos para o desenvolvimento de ambos os protótipos, sendo os requisitos do protótipo de extração de unidades de atributos os mesmos do trabalho correlato de Horch, Kett e Weisbecker (2015), já os requisitos do protótipo WEB tiveram como origem sites conhecidos de busca, como o Buscapé e Bondfaro. Após o desenvolvimento, execução e coleta dos dados do protótipo de extração de unidades de atributos, foi possível realizar a análise comparativa dos dados extraídos com o estudo de Horch, Kett e Weisbecker (2015). Após, foram elaboradas as questões para o método TAM (Technology Acceptance Model, usado para a análise de aceitação) e realizada sua aplicação a um grupo de pessoas que teve acesso ao protótipo WEB.

Liu (2014) explica que o TAM tem como base duas variáveis primárias, compostas por variáveis independentes, que são: Utilidade percebida (UP) e Facilidade de uso percebida (FU); e duas variáveis dependentes, sendo elas, Atitude para o Uso (AU) e Intenção Comportamental (IC). O modelo de pesquisa TAM original de Davis (1989) é apresentado na Figura 7.

Figura 7: Modelo de pesquisa TAM Davis (1989)



Fonte: Adaptado de Liu (2014).

Liu (2014) explica os relacionamentos ilustrados na Figura 7, sendo eles:

- Facilidade de Uso Percebida (FU)

*H1: A facilidade de uso percebida (FU) influenciará positivamente a atitude dos usuários em relação ao sistema.*

- Utilidade Percebida (UP)

*H2: A utilidade percebida (UP) influenciará positivamente a atitude dos usuários em relação ao sistema.*

*H3: Utilidade Percebida (UP) influenciará positivamente a intenção comportamental dos usuários para o uso do sistema.*

- Atitude para o Uso (AU) e Intenção Comportamental (IC)

*H4: A atitude para o uso do sistema irá influenciar positivamente a intenção comportamental dos usuários do sistema.*

- Facilidade de Uso Percebida (FU) e Utilidade Percebida (UP)

*H5: Facilidade de Uso Percebida (FU) influenciará positivamente a Utilidade Percebida (UP) do sistema.*

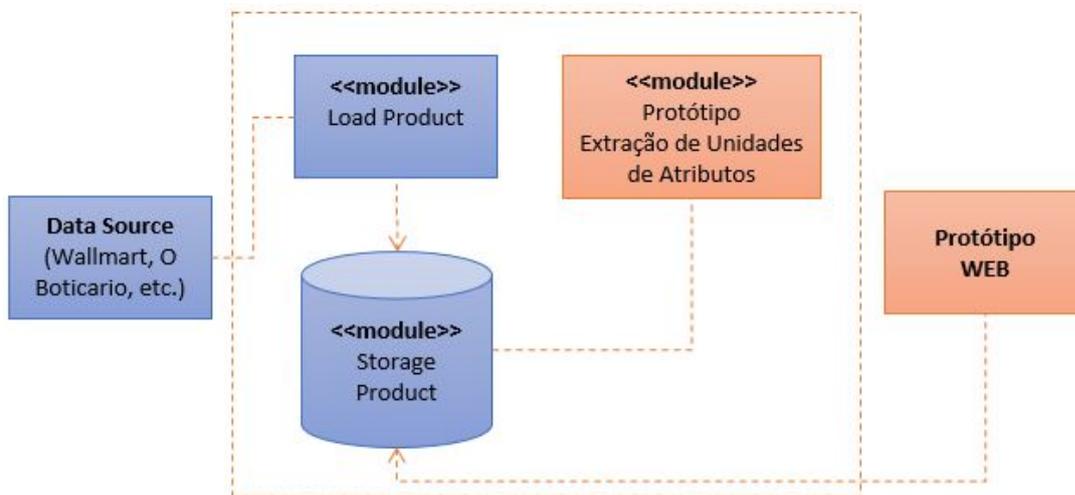
Davis (1989) definiu a utilidade percebida como “o grau em que uma pessoa acredita que usar um determinado sistema aumentaria seu desempenho no trabalho” e, a facilidade de uso percebida, como “o grau em que uma pessoa acredita que usar um determinado sistema estaria livre de esforço”. Davis (1993) afirma que a intenção comportamental, que é determinada em conjunto pela atitude dos usuários em relação ao uso do sistema e à utilidade percebida, determina o uso real do sistema de informação.

A técnica de coleta de dados para o modelo TAM ocorreu através de questionário quantitativo utilizando a escala Likert. Amaro, Macedo e Macedo (2005) explicam que a escala Likert é composta por uma série de cinco proporções e que, dessas alternativas, deve ser selecionada uma pelo entrevistado, podendo a alternativa ser: discorda totalmente, discorda, sem opinião (ou indeciso), concorda e concorda totalmente. Para análise do resultado é estabelecida uma cotação das alternativas com valores respectivos de: -2, -1, 0, +1, +2 ou utilizando pontuações de 1 a 5.

## **5 PROTÓTIPO E TECNOLOGIAS UTILIZADAS**

Esta seção apresenta as tecnologias empregadas nesse estudo além da arquitetura proposta. A Figura 8 ilustra a arquitetura geral proposta para a extração de unidades de produtos em lojas virtuais brasileiras.

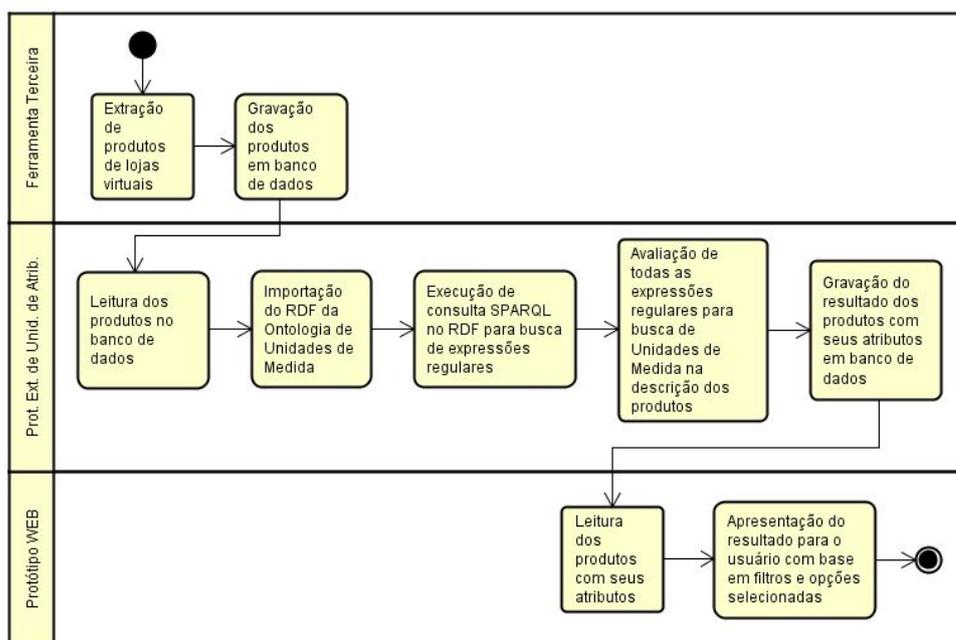
Figura 8: Arquitetura Proposta



Fonte: Elaborada pelo autor.

Os itens destacados na cor azul da Figura 8 não estão sendo abordados neste estudo, sendo que os demais contemplam os protótipos desenvolvidos. Na Figura 9 é apresentado o fluxo de funcionamento da arquitetura para extração de unidades de produtos.

Figura 9: Fluxograma da Arquitetura



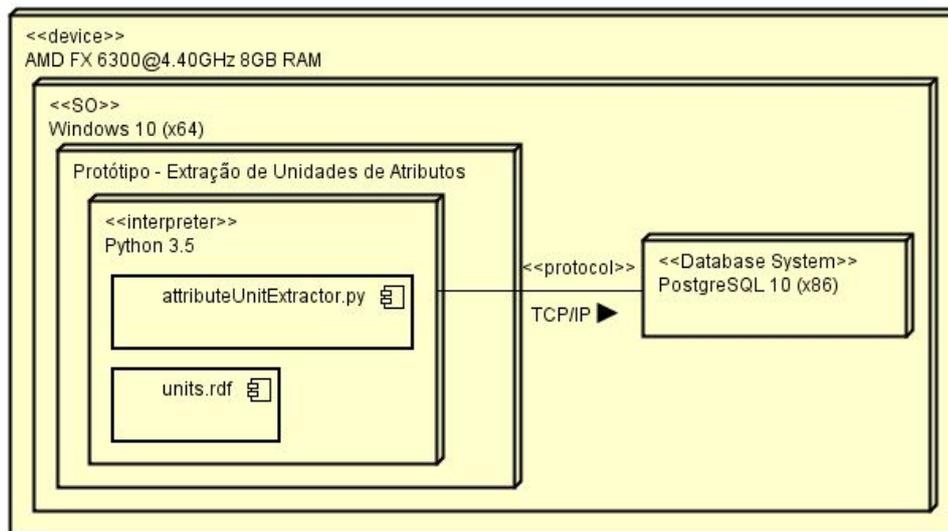
Fonte: Elaborada pelo autor.

De acordo com a Figura 9, o processo inicia com a extração (módulo Load Product) e gravação (módulo Storage Product) de produtos de lojas virtuais em um banco de dados com origem em uma fonte de dados. Em seguida, os dados previamente extraídos são processados pelo protótipo para a extração das unidades de atributos utilizando a consulta SPARQL (Simple

Protocol and RDF Query Language) ao arquivo RDF (Resource Description Framework) que contém a ontologia de unidades de medida para avaliação da descrição dos produtos; após, as unidades extraídas são gravadas em banco de dados para leitura e exibição posterior no protótipo WEB.

A Figura 10 ilustra a arquitetura do protótipo de Extração de Unidades de Atributos.

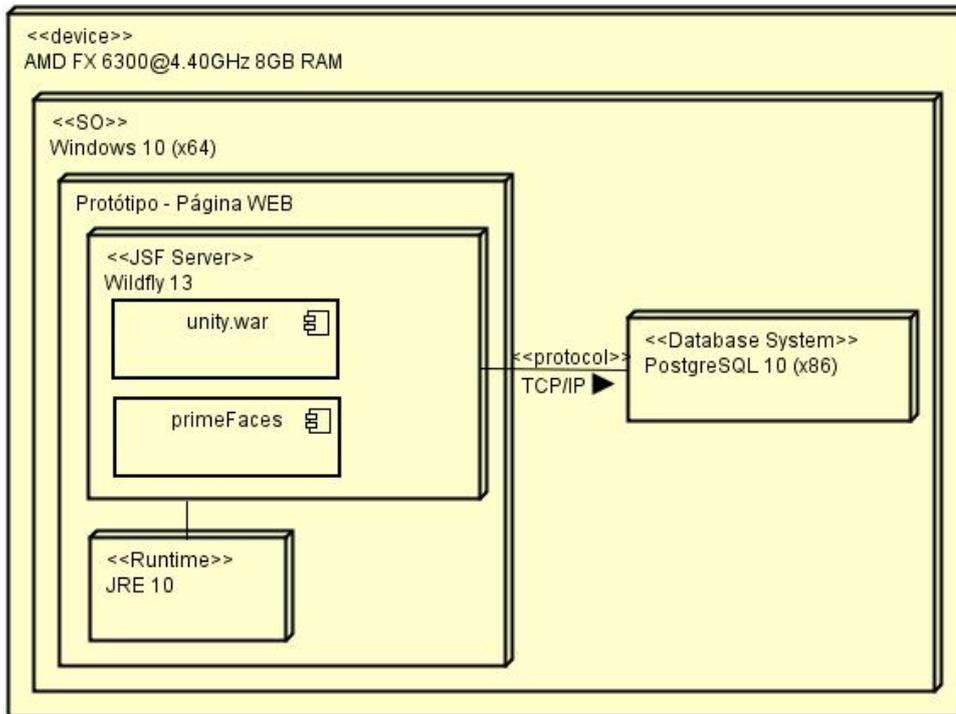
Figura 10: Arquitetura do Protótipo de Extração de Unidades de Atributos



Fonte: Elaborada pelo autor.

Conforme pode ser visto na arquitetura acima, o protótipo é composto pelo componente `attributeUnitExtractor.py`, que é responsável pela análise da descrição de produtos para a extração e armazenamento de seus atributos, que é um dos objetivos deste estudo. Ainda, para dar suporte ao desenvolvimento desse protótipo, foi utilizado o IDE PyCharm 2019.1.2. Como linguagem de programação foi escolhido o Python 3.5; como linguagem de consulta de RDF o SPARQL 1.1, além do RDF de unidades de medida. A seguir, a Figura 11 mostra a arquitetura do protótipo WEB.

Figura 11: Arquitetura do Protótipo WEB



Fonte: Elaborada pelo autor.

Para a construção do protótipo da arquitetura acima, foi utilizado o IDE Eclipse 2019.03, sendo que a linguagem de programação adotada foi o Java 10. Como servidor de aplicação optou-se pelo Wildfly 13. Ademais, na camada de apresentação foi usado o JSF (JavaServer Faces) 2.3 com a biblioteca de componentes Primefaces 6.2. Para a injeção de dependência foi utilizado o Java Enterprise Edition 8. Ressalta-se que todas as tecnologias foram escolhidas em virtude do conhecimento do autor desse estudo. Além disso, é possível verificar no Apêndice A informações sobre os protótipos.

## 6 APRESENTAÇÃO E ANÁLISE DE DADOS

Nesta seção é apresentada a análise e discussão dos resultados obtidos através da execução do protótipo de extração de unidades de atributos, assim como a comparação com o trabalho de Horch, Kett e Weisbecker (2015). Também são apresentados e discutidos os resultados obtidos a partir do uso do protótipo WEB.

## 6.1 Protótipo de Extração de Unidades de Atributos

### 6.1.1 Análise

O protótipo processou 300 produtos de três lojas virtuais brasileiras. As lojas virtuais foram escolhidas usando como critério sua avaliação e reputação no Ebit<sup>5</sup>, além de possuírem as mesmas categorias de produtos em relação ao trabalho correlato de Horch, Kett e Weisbecker (2015). A Tabela 1 sumariza os departamentos e categorias de produtos extraídos.

Tabela 1: Departamentos e Categorias de Produtos

Loja	Portal	Departamento	Categoria	Produtos
O Boticario	www.boticario.com.br	Cosméticos	Hidratantes	60
O Boticario	www.boticario.com.br	Cosméticos	Perfumaria Marculina	40
Walmart	www.walmart.com.br	Alimentos	Condimentos e Ervas	20
Walmart	www.walmart.com.br	Alimentos	Doces e Sobremesas	20
Walmart	www.walmart.com.br	Alimentos	Frios Laticíneos	20
Walmart	www.walmart.com.br	Alimentos	Grãos, Sementes e Oleaginosas	20
Walmart	www.walmart.com.br	Alimentos	Bebidas Não Alcoolicas	20
Magazine Luiza	www.magazineluiza.com.br	Eletrônicos	Computadores	50
Magazine Luiza	www.magazineluiza.com.br	Eletrônicos	Celulares	50
				<b>300</b>

Fonte: Elaborada pela autor.

Com base na amostra acima, a tabela seguir apresenta um resumo do processamento obtido pelo protótipo.

Tabela 2: Extração de Atributos de Produtos

	Produtos	Perc.
Extraído parcialmente correto	73	24,33%
Extraído incorretamente	0	0,0%
Extraído corretamente	196	65,33%
Não extraído	31	10,33%
	<b>300</b>	<b>100%</b>

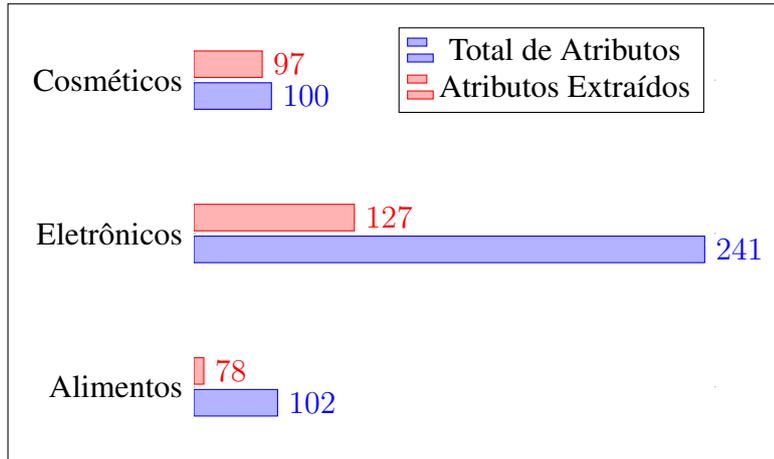
Fonte: Elaborada pela autor.

Observa-se que, do total de 300 produtos, 24,33% tiveram seus atributos extraídos parcialmente; nenhum produto teve atributos extraídos de forma incorreta; 65,33% dos produtos tiveram todos seus atributos extraídos de forma correta e 10,33% dos produtos não tiveram nenhum atributo extraído.

Ao considerar os 300 produtos, há um total de 443 atributos possíveis. Desses, 302 foram extraídos corretamente pelo protótipo, totalizando 68% de atributos extraídos. O Gráfico 1 apresenta a relação de extração por departamento.

<sup>5</sup>Ebit - Reputação de lojas virtuais e dados para o mercado on-line - <https://www.ebit.com.br/>

Gráfico 1 - Extração de Atributos por Departamento



Fonte: Elaborado pelo autor.

Conforme apresentado no Gráfico 1, o departamento de alimentos tem 102 unidades, o de eletrônicos tem 241 e o de cosméticos 100. Para o departamento de cosméticos foram extraídos 97% dos atributos, sendo que a alta taxa de extração ocorreu devido a boa qualidade da ortografia das unidades de medida, sua homogeneidade (a grande maioria é unidade "ml") e a quantidade de atributos por produto, pois nesse caso, 100% dos produtos continham somente um atributo. Os três atributos não extraídos estavam com a ortografia incorreta. Para o departamento de eletrônicos, 47,3% dos atributos não foram extraídos em função de existir em média 2,41 atributos por produto e, na maioria dos casos, há uma duplicidade da mesma unidade de medida para duas características como, por exemplo, o GB para memória RAM e para o armazenamento de disco rígido, causando a exportação somente da primeira unidade. No departamento de alimentos, 23,5% dos atributos não foram extraídos, sendo que o principal motivo foi a baixa qualidade no cadastro dos produtos desse departamento, em muitos casos a sigla de unidade de medida estava informada fora do padrão como a representação de Mililitro "Ml", quando o correto seria "ml".

### 6.1.2 Análise Comparativa

A análise comparativa com o estudo correlato de Horch, Kett e Weisbecker (2015) utilizou as seguintes métricas: precisão, recall, extração correta e extração incorreta, sendo que as fórmulas para cada métrica são apresentadas a seguir.

$$\text{Precisão} = \frac{\text{positivos verdadeiros}}{\text{positivos verdadeiros} + \text{falsos positivos}} \quad (1)$$

$$\text{Recall} = \frac{\text{positivos verdadeiros}}{\text{positivos verdadeiros} + \text{falsos negativos}} \quad (2)$$

$$\text{Extração Correta} = \frac{\text{positivos verdadeiros}}{\text{total de atributos}} \quad (3)$$

$$\text{Extração Incorreta} = \frac{\text{falsos negativos}}{\text{total de atributos}} \quad (4)$$

Horch, Kett e Weisbecker (2015) definem os *positivos verdadeiros* como atributos reais extraídos corretamente (Ex.: ml para mililitro); *falsos positivos* como atributos não reais, porém extraídos como sendo (Ex.: código de produto LJ-42B sendo identificado como 42 Bytes); e *falsos negativos* como atributos reais que não foram extraídos (Ex.: atributo 60 ML não identificado devido a unidade estar fora da ontologia, que seria *ml* para mililitro). A seguir, a Tabela 3 apresenta o resultado do comparativo.

Tabela 3: Comparação com trabalho correlato

	<b>Trabalho Correlato</b>	<b>Este Estudo</b>	<b>Diferença</b>
<b>Precisão</b>	99,07%	100%	0,93pp
<b>Recall</b>	77,74%	68,17%	-9,57pp
<b>Extração Correta</b>	77,17%	68,17%	-9,00pp
<b>Extração Incorreta</b>	22,83%	31,83%	9,00pp

Fonte: Elaborada pela autor.

Observa-se que houve convergência entre as métricas, destacando-se a precisão e a extração correta. O estudo identificou que as variações observadas relacionam-se com a diferença no padrão de dados informados na descrição dos produtos em lojas virtuais brasileiras em comparação ao cenário europeu.

Além disso, há precisão de 100%, ou seja, não ocorreu nenhuma extração de atributo de forma incorreta, já no trabalho correlato dois atributos foram extraídos incorretamente. Quanto ao recall, ocorreu uma diferença de 9,57 pontos percentuais, atingindo 68,17%, valor igual a métrica de extração correta devido a não ocorrência de falsos positivos (fator diferenciador da métrica recall conforme fórmulas apresentadas anteriormente).

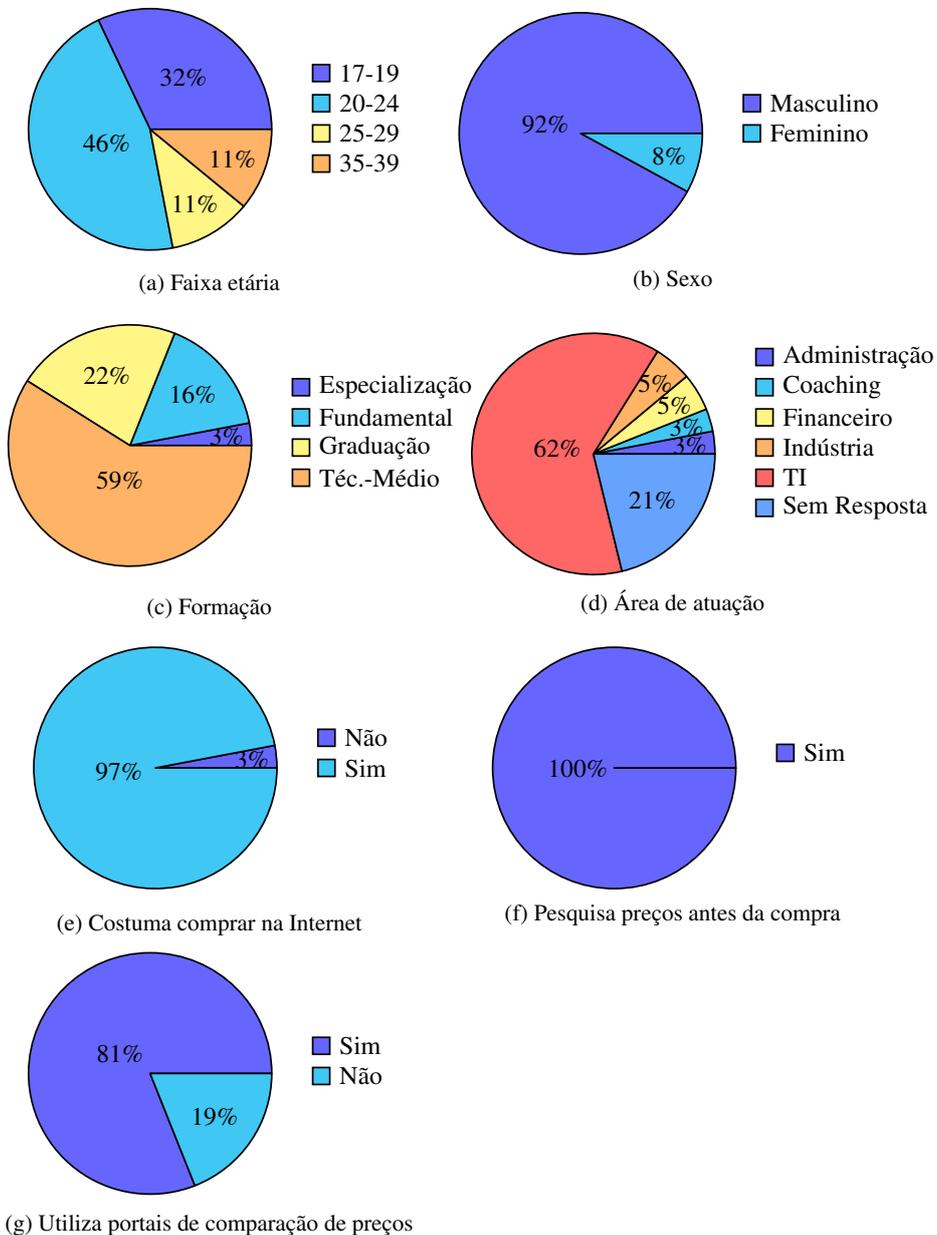
Foram observadas duas características que influenciaram na diferença dos resultados dos cenários: (i) a não ocorrência de falsos positivos, ou seja, no cenário brasileiro há uma melhor qualidade na descrição dos produtos que influenciou na não identificação de informações genéricas como sendo atributos de produtos; (ii) observou-se no cenário brasileiro uma grande quantidade de atributos na descrição de produtos de forma incorreta, ou seja, sem padronização da unidade de medida, como, por exemplo, quilograma sendo descrito como Kg (primeira letra maiúscula), quando o correto para sua abreviação seria kg. Esse comportamento do cenário brasileiro influenciou na taxa menor de extração de atributos da amostra geral.

## 6.2 Protótipo WEB

### 6.2.1 Perfil Demográfico

Nesta seção é apresentado, através do Gráfico 2, o perfil demográfico obtido a partir da aplicação do modelo TAM (ver Apêndice B) a 37 respondentes, sendo que, a amostra de respondentes foi obtida através do envio da pesquisa por e-mail e redes sociais a alunos da Universidade do Vale do Rio dos Sinos (UNISINOS).

Gráfico 2 - Perfil Demográfico



Fonte: Elaborado pelo autor.

Com base nos gráficos do perfil demográfico, os respondentes têm entre 17 e 39 anos de idade, sendo que destes, 92% são do sexo masculino e 8% do sexo feminino. Em relação a maior formação concluída, 59% tem formação técnica/ensino médio, 22% Graduação, 16% Fundamental e 3% Especialização. Sobre a área de atuação, 62% atuam com Tecnologia da Informação, 17% em outras áreas e 21% ainda não atuam ou preferiram não responder. O questionário também indagou sobre costumes de compra e, nesses quesitos, 97% costumam comprar na internet, 100% pesquisam preços antes da compra e 81% utilizam algum portal de comparação de preços da internet. Com base nesses costumes, pode-se observar a importância que uma ferramenta teria no dia-a-dia.

### 6.2.2 Análise

As análises apresentadas a seguir foram realizadas utilizando a ferramenta *SPSS Statistics Subscription* versão 24. A primeira análise foi a de confiabilidade, que tem como objetivo verificar a validade e consistência dos itens usados em cada constructo. O resultado da análise de confiabilidade é apresentado na Tabela 5.

Tabela 5: Análise de Confiabilidade

<b>Constructos</b>	<b>Itens</b>	<b>Média (<math>\bar{x}</math>)</b>	<b>Desvio Padrão (<math>\sigma</math>)</b>	<b>Alfa de Cronbach (<math>\alpha</math>)</b>
Facilidade de Uso (FU)	6	4,25	0,58	0,742
Utilidade Percebida (UP)	6	4,02	0,74	0,897
Atitude para o Uso (AU)	4	3,81	0,98	0,916
Intenção Comportamental (IC)	3	3,72	0,95	0,866
<b>Todos constructos</b>	<b>19</b>	<b>4</b>	<b>0,63</b>	<b>0,923</b>

Fonte: Elaborada pela autor.

Baseado nas classificações de confiabilidade de acordo com o Alfa de Cronbach, apresentadas por Taber (2017), a classificação foi alta para o constructo Facilidade de Uso, confiável para os constructos Intenção Comportamental e Utilidade Percebida, ótima para o constructo Atitude de Uso e também ótima considerando todos os constructos. As médias indicam que as respostas estão entre indeciso e concordo plenamente, sendo que o desvio padrão mostra que houve uma dispersão máxima de 0,98 em relação à média, mostrando variação significativa entre as respostas.

A próxima análise foi a de correlação bilateral utilizando o coeficiente de correlação de Pearson. Essa análise tem como objetivo identificar uma relação entre os constructores, sendo o seu resultado apresentado na Tabela 6.

Tabela 6: Correlação de Pearson

	FU	UP	AU	IC
<b>Facilidade de Uso (FU)</b>	1	0,576**	0,336*	0,171
<b>Utilidade Percebida (UP)</b>	0,576**	1	0,700**	0,660**
<b>Atitude para o Uso (AU)</b>	0,336*	0,700**	1	0,787**
<b>Intenção Comportamental (IC)</b>	0,171	0,660**	0,787**	1

\*\* . A correlação é significativa no nível 0,01 (bilateral).  
\* . A correlação é significativa no nível 0,05 (bilateral).

Fonte: Elaborada pela autor.

Os resultados apresentados na Tabela 6 mostram correlações positivas e significativas com  $p < 0,05$  ou  $p < 0,001$  entre todos os constructos exceto entre Facilidade de Uso (FU) e Intenção Comportamental (IC), que apresentaram uma correlação desprezível de somente 0,171.

A seguir são apresentados os resultados do teste das cinco hipóteses definidas neste estudo, sendo que as análises foram executadas utilizando a regressão linear considerando os níveis de significância 0,05, 0,01 e 0,001. As duas primeiras hipóteses testadas foram H1 e H2, utilizando FU e UP para o cálculo da variável dependente AU. O resultado é apresentado na Tabela 7.

Tabela 7: Preditores FU e UP e Variável Dependente AU

Resumo do Modelo					
Modelo	R	R Quadrado	R Quadrado Ajustado	Erro Padrão da Estimativa	
1	0,704 <sup>a</sup>	0,496	0,466	0,72208	
a. Preditores: (Constantes), UP, FU					
Coeficientes					
Modelo	Coeficientes não Padronizados		Coeficientes Padronizados		Sig
	B	Erro Padrão	Beta( $\beta$ )	t	
(Constante)	0,506	0,896		0,565	0,576
1 FU	-0,168	0,251	-0,100	-0,671	0,507
UP	1,000	0,197	0,757	5,084	0,000
Variável Dependente: AU					

Fonte: Elaborada pela autor.

De acordo com o resultado apresentado na Tabela 7, o valor de R quadrado afirma que a Facilidade de Uso e a Utilidade Percebida explicam 49,6% da Atitude para o Uso da ferramenta. Também é possível verificar através do coeficiente padronizado  $\beta$  o impacto de cada constructo sobre a variável dependente. Os resultados indicam que a Utilidade Percebida tem grande impacto sobre a Atitude de Uso diferente da Facilidade de Uso. Com isso, a hipótese H1 foi rejeitada com  $\beta$  de -0,100 e nível de significância maior do que 0,05. Já a hipótese H2 foi confirmada com  $\beta$  de 0,757 e nível de significância menor do que 0,001. A seguir é apresentado na Tabela 8 o resultado da análise de regressão linear para as hipóteses H3 e H4, levando em conta UP e AU para o cálculo da variável dependente IC.

Tabela 8: Preditores UP e AU e Variável Dependente IC

Resumo do Modelo					
Modelo	R	R Quadrado		R Quadrado Ajustado	Erro Padrão da Estimativa
1	0,801 <sup>a</sup>	0,642		0,621	0,58722
a. Preditores: (Constantes), UP, AU					
Coeficientes					
Modelo	Coeficientes não Padronizados			Coeficientes Padronizados	
	B	Erro Padrão	Beta( $\beta$ )	t	Sig
(Constante)	0,279	0,535		0,522	0,605
1 UP	0,274	0,183	0,215	1,497	0,144
AU	0,614	0,139	0,636	4,434	0,000
Variável Dependente: IC					

Fonte: Elaborada pela autor.

Conforme apresentado na Tabela 8, o valor de R quadrado afirma que a Utilidade Percebida e a Atitude para o Uso explicam 64,2% da Intenção Comportamental. Os resultados de cada constructo mostram que a Utilidade Percebida tem pouco impacto sobre a Intenção comportamental com  $\beta$  de 0,215 e nível de significância maior do que 0,05, enquanto que a Atitude para o Uso tem grande impacto com  $\beta$  de 0,636 e nível de significância menor do que 0,001. Dessa forma, a hipótese H3 foi rejeitada e a hipótese H4 confirmada. Por ultimo, conforme apresentado na Tabela 9, foi realizada a análise de regressão linear para a hipótese H5, considerando apenas FU para o cálculo da variável dependente UP.

Tabela 9: Preditor FU e Variável Dependente UP

Resumo do Modelo					
Modelo	R	R Quadrado		R Quadrado Ajustado	Erro Padrão da Estimativa
1	0,576 <sup>a</sup>	0,332		0,313	0,62038
a. Preditor: (Constante), FU					
Coeficientes					
Modelo	Coeficientes não Padronizados			Coeficientes Padronizados	
	B	Erro Padrão	Beta( $\beta$ )	t	Sig
(Constante)	0,902	0,755		1,196	0,240
1 FU	0,734	0,176	0,576	4,167	0,000
Variável Dependente: UP					

Fonte: Elaborada pela autor.

A Tabela 9 mostra que a Facilidade de Uso explica apenas 33,2% da Utilidade Percebida. Porém, levando-se em conta o coeficiente padronizado  $\beta$ , é possível observar que a Facilidade de Uso tem impacto significativo sobre a Utilidade Percebida com  $\beta$  de 0,576 e nível de significância menor do que 0,001, o que confirma a hipótese H5. O resultado dos testes de hipóteses é resumido na Tabela 10 a seguir.

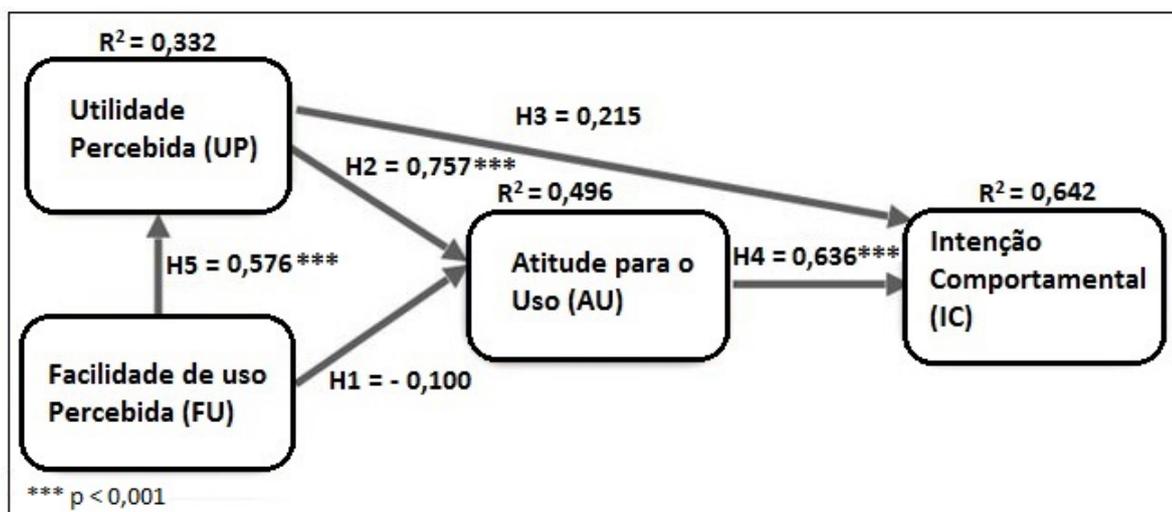
Tabela 10: Resultado do Teste de Hipóteses

Hipótese	Especificação	Resultado
H1	Facilidade de uso irá influenciar positivamente a Atitude para o Uso da ferramenta	Não confirmada ( $\beta = -0,100, p > 0,05$ )
H2	Utilidade percebida irá influenciar positivamente a Atitude para o Uso da ferramenta	Confirmada ( $\beta = 0,757, p < 0,01$ )
H3	Utilidade percebida irá influenciar positivamente a Intenção Comportamental	Não confirmada ( $\beta = 0,215, p > 0,05$ )
H4	Atitude para o uso irá influenciar positivamente a Intenção Comportamental	Confirmada ( $\beta = 0,636, p < 0,01$ )
H5	Facilidade de uso irá influenciar positivamente a Utilidade Percebida da ferramenta	Confirmada ( $\beta = 0,576, p < 0,01$ )

Fonte: Elaborada pela autor.

De acordo com os dados apresentados na Tabela 10, três hipóteses foram confirmadas e duas não. É possível observar as hipóteses H2 e H4 com melhores resultados, confirmando que a Utilidade Percebida exerce grande influência sobre a Atitude para o Uso, que por sua vez, impacta positivamente na Intenção Comportamental para o uso da ferramenta. A hipótese rejeitada H1 mostra que a Facilidade de Uso não influencia na Atitude para o uso da ferramenta, bem como, devido a rejeição da hipótese H3, a Utilidade Percebida não influencia positivamente a Intenção Comportamental. Na Figura 12 é apresentado o resultado da regressão linear com os valores inseridos no modelo TAM.

Figura 12: Resultados da Análise de Regressão Linear



Fonte: Adaptado de Liu (2014).

### 6.3 Discussão

Este estudo tem como foco criar uma arquitetura para extração de unidades de medidas e, para isso, foram desenvolvidos dois protótipos: o primeiro para a extração de atributos de produtos de lojas virtuais brasileiras e, o segundo, um protótipo WEB para apresentação e comparação de preços de produtos, mostrando a aplicação da ferramenta. Além do desenvolvimento dos dois protótipos, foi realizado um estudo comparativo com o estudo correlato, mostrando que o resultado desta comparação apresentou similaridade nos resultados, porém, com algumas diferenças apontadas, que se deram principalmente devido às diferenças na qualidade da informação do cadastro de produtos da amostra deste estudo em relação ao trabalho correlato. Também foi aplicado um questionário utilizando o método TAM para verificar a aceitação dos usuários em relação ao protótipo WEB. O resultado deste questionário evidenciou que a Utilidade Percebida exerce grande influência sobre a Atitude para o Uso que, por sua vez, impacta positivamente na Intenção Comportamental para o uso da ferramenta. Ainda, foi confirmado que a Facilidade de Uso não influencia na Atitude para o Uso da ferramenta, assim como a Utilidade Percebida não influencia a Intenção Comportamental.

Os resultados obtidos através do estudo comparativo e a análise de aceitação foram suficientes para responder a questão de pesquisa deste estudo: **qual a relevância de uma arquitetura de extração e comparação de produtos em lojas on-line?** Sua relevância é confirmada devido a automatização da extração de características (atributos) de produtos a partir de sua descrição, possibilitando a extração de produtos para comparação de preços de qualquer loja on-line, independente do seu formato e forma de apresentação dos produtos, assim como, a possibilidade de uma forma prática e simples de comparar produtos levando em conta seus atributos e possibilitando a visualização do custo por unidade de medida (Ex.: custo por mililitro). A relevância do método de comparação de produtos também pode ser observado através do feedback dos respondentes, onde pode-se destacar alguns como: "Excelente!!"; "Bem prático e rápido"; "Achei a interface muito simples e direta, realmente é uma ferramenta super útil na realidade Brasileira.", demais feedbacks estão listados no Apêndice C.

## 7 CONCLUSÃO

Em relação ao mercado de comércio eletrônico brasileiro, o crescimento tem ocorrido ano após ano e as projeções são ainda mais otimistas E-BIT (2019). Dessa forma, cada vez mais se faz necessária uma padronização na apresentação de ofertas de produtos, tornando relevante um processo automatizado de extração e comparação de atributos de produtos de lojas on-line. Para a fundamentação dos conceitos que guiam esse trabalho, foi realizada uma pesquisa bibliográfica, além da análise de trabalhos relacionados. Durante essa pesquisa foi possível observar a relevância do tema de extração automatizada através da WEB Semântica e ontologias.

O estudo comparativo teve como base o trabalho de Horch, Kett e Weisbecker (2015) e cor-

roborou a eficácia do protótipo de extração de atributos de produtos, mostrando que a ontologia de unidades de medida é uma alternativa viável no processo de automação. Além disso, ocorreram algumas diferenças entre o cenário brasileiro e o cenário europeu, em função do padrão de cadastro de produtos e na limitação do método quando os produtos têm mais de um atributo de mesmo tipo. Ainda, a análise de aceitação de tecnologia com o uso do protótipo WEB mostrou resultados satisfatórios, com boa receptividade pela maioria dos respondentes e, também, mostrou que a Utilidade Percebida exerce grande influência sobre a Atitude para o Uso, que por sua vez, impacta positivamente na Intenção Comportamental para o uso da ferramenta.

Conclui-se nesse estudo que foi possível responder a questão de pesquisa e cumprir com os objetivos apresentados, possibilitando uma discussão relevante sobre a Web Semântica e ontologias.

Como possibilidade de continuidade desse estudo, sugere-se o aperfeiçoamento do método de extração de atributos, possibilitando a extração de mais de uma unidade do mesmo tipo, além do desenvolvimento de uma ontologia de unidades de medida que possa prever a extração de unidades fora do padrão, se antecipando a possíveis erros de cadastro de produtos. Também sugere-se o aperfeiçoamento do protótipo WEB, pois o mesmo teve boa aceitação pela sua simplicidade e praticidade, porém, nos comentários dos respondentes (ver Apêndice C) foram sugeridas melhorias como o aperfeiçoamento do filtro de unidades de medida e a opção de escolha do tamanho da unidade de medida (exemplo ml ou L).

## AN ARCHITECTURE FOR EXTRACTION OF PRODUCT UNITS IN BRAZILIAN VIRTUAL SHOPS

**CONTEXT:** E-commerce grows every year since it offers to the e-consumers an easy way to find the best offers and make their purchases online. To enable this task, there are several processes on the search engines used by e-consumers, allowing the comparison of similar products. **PROBLEM:** This task of comparing similar products becomes a challenge due to the lack of standardization of the registered information in the WEB. In most cases, product attribute information is described freely, unformatted and directly in the product description, making it difficult to establish a standard for comparison. Another important factor is the growth of Brazilian e-commerce and, consequently, the need to simplify the process of e-consumer analysis. **GOAL:** Due to the problems presented, this study presents an architecture for automated extraction of product units and two prototypes, the first one is a product attribute extraction engine, and the second, one, a WEB prototype for presentation and comparison of product prices by virtual store. **PROPOSED METHOD:** In order to reach the main goal of this study, we opted for the experimentation-type research method, where the extraction prototype was compared with the developed study related by Horch, Kett and Weisbecker(2015). Also, prototyping was used for the development of the WEB prototype to show the application of the result of the extraction of products attributes, achieved by the extration engine. In order to validate the WEB prototype, the TAM (Technology Acceptance Model) acceptance model was applied. **CONCLUSION:** Finally, we present an analysis and discussion of the results, besides the acceptance of the prototype, showing that the architecture is feasible and useful for e-commerce.

Keywords: Semantic Web. Ontology. OWL (Web Ontology Language). E-commerce. TAM (Technology Acceptance Model). Extraction of Units.

## REFERÊNCIAS

- ALMEIDA, M. A.-c. B. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação**, [S.l.], v. 31, p. 5 – 13, 08 2002.
- AMARO, A.; MACEDO, A.; MACEDO, L. **A arte de fazer questionários**. 2005. Dissertação (Mestrado em Ciência da Computação) — Faculdade De Ciências Da Universidade Do Porto, 2005. Disponível em: <<https://sites.google.com/site/sociologiaemacao/2-metodologia-da-investigacao-sociologica/a-arte-de-fazer-questionarios.doc>>. Acesso em: nov. 2018.
- BALOG, K. On the investigation of similarity measures for product resolution. In: LDH, 2011. **Anais...** [S.l.: s.n.], 2011. p. 49–54. Disponível em: <<http://krisztianbalog.com/files/ldh2011-produCLpdf>>. Acesso em: 8 ago. 2018.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. In: SCIENTIFIC AMERICAN, 2001. **Anais...** [S.l.: s.n.], 2001. Disponível em: <[https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf)>. Acesso em: 08 out. 2018.
- DAVIS, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. **MIS Quarterly**, Michigan, v. 13, n. 3, p. 319–340, Sept. 1989.
- DAVIS, F. D. User acceptance of information technology: system characteristics, user perceptions and behavioural impacts. **International Journal of Man-Machine Studies**, Ann Arbor, M1 48109, USA, v. 38, n. 3, p. 475–487, Mar. 1993.
- RODRIGUES, J. **Html básico - códigos html**. [S.l.: s.n.], 2010. Disponível em: <<https://www.devmedia.com.br/html-basico-codigos-html/16596>>. Acesso em: 06 out. 2018.
- E-BIT. **39ª edição do estudo webshoppers**. 2019. Disponível em: <<https://www.ebit.com.br/webshoppers>>. Acesso em: 27 apr. 2019.
- FENSEL, D. **Ontologies: a silver bullet for knowledge management and electronic commerce**. [S.l.: s.n.], 2000. v. 31.
- FREITAS, F. L. G. de. Ontologias e websemântica. In: XXIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 2003. **Anais...** [S.l.: s.n.], 2003. p. 52. Disponível em: <[http://www.inf.ufsc.br/fernando.gauthier/EGC6006/material/Aula%203/Ontologia\\_Web\\_semantica%20Freitas.pdf](http://www.inf.ufsc.br/fernando.gauthier/EGC6006/material/Aula%203/Ontologia_Web_semantica%20Freitas.pdf)>. Acesso em: 22 set. 2018.
- GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. 2009. Dissertação (Mestrado em Ciência da Computação) — UFRGS, 2009. Disponível em: <<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>>. Acesso em: 28 out. 2018.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. [S.l.]: Atlas, 2007.
- GUARIDO, N.; OBERLE, D.; STAAB, S. What is an ontology? In: STAAB, S., STUDER, R. (EDS.) **HANDBOOK ON ONTOLOGIES**, 2009, Springer, Heidelberg. **Anais...** [S.l.: s.n.], 2009. p. 1–17. Disponível em: <<https://pdfs.semanticscholar.org/8d71/9f74e24e30aea03178152d630ec09aa01a58.pdf>>. Acesso em: 08 ago. 2018.

HORCH, A.; KETT, H.; WEISBECKER, A. Extracting product unit attributes from product offers by using an ontology. In: SECOND INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE, COMPUTER ENGINEERING, AND SOCIAL MEDIA (CSCESM), 2015., 2015. **Anais...** [S.l.: s.n.], 2015. p. 67–71.

ISOTANI, S. et al. Estado da arte em web semântica e web 2.0: potencialidades e tendências da nova geração de ambientes de ensino na internet. **Brazilian Journal of Computers in Education**, [S.l.], v. 17, n. 01, p. 30, 2009.

LAUFER, C. **Guia de web semântica**. [S.l.]: Governo do estado de São Paulo, 2015. Disponível em: <<http://www.governoaberto.sp.gov.br/wp-content/uploads/2016/05/Book-Web-Semantica.pdf>>. Acesso em: 21 set. 2018.

LIU, Z. Y. **An analysis of technology acceptance model**. 2014. Dissertação (Mestrado em Ciência da Computação) — University of Gothenburg, 2014. Disponível em: <<https://gupea.ub.gu.se/handle/2077/38592>>. Acesso em: 02 nov. 2018.

PICKLER, M. E. V. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, [S.l.], v. 12, p. 65 – 83, 04 2007.

RIJGERSBERG, H.; ASSEM, M. van; TOP, J. Ontology of units of measure and related concepts. **Semant. web**, Amsterdam, The Netherlands, The Netherlands, v. 4, n. 1, p. 3–13, Jan. 2013.

ROQUE, C. Ontology: a survey. In: SYMPOSIUM ON AUTOMATED SYSTEMS BASED ON HUMAN SKILL AND KNOWLEDGE, 8., 2003. **Anais...** [S.l.: s.n.], 2003. p. 28–41. Disponível em: <<http://christophe-roche.fr/Bibliographie/2003/Ontology%20Survey%20-%20Roche.pdf>>. Acesso em: 14 out. 2018.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, [S.l.], v. 33, p. 132 – 141, 04 2004.

SU, W.; WANG, J.; LOCHOVSKY, F. H. Ode: ontology-assisted data extraction. **ACM Trans. Database Syst.**, New York, NY, USA, v. 34, n. 2, p. 12:1–12:35, July 2009.

TABER, K. The use of cronbach's alpha when developing and reporting research instruments in science education. **Research in Science Education**, [S.l.], p. 1–24, 06 2017.

## APÊNDICE A – PROTÓTIPOS DE EXTRAÇÃO DE UNIDADES DE ATRIBUTOS E WEB

Figura 13: Código fonte attributeUnitExtractor.py

```

94
95     return size
96
97     def getAttributesFromOntology(self, product_description, product_name):
98
99         product_attributes = []
100         #product_attributes2 = {}
101         graph = Graph()
102         graph.parse(self.ontology_path)
103         rdf_class = ""
104         size = None
105         unit = None
106         sizes = []
107
108         # get all regex and labels of units of measurement
109         result = graph.query(
110             """SELECT ?regex ?label ?individual
111             {SELECT DISTINCT ?regex ?label ?individual
112             WHERE {
113                 ?individual units:regex ?regex .
114                 ?individual rdfs:label ?label .
115                 FILTER(LANG(?label) = "") .
116             } ORDER BY ?label GROUP BY ?regex""")
117
118         #print(product_description)
119         #print(product_name)
120         for row in result:
121             regex_string = str("%s" %row[0])
122             regex = re.compile(regex_string)
123             res_description = regex.search(product_description.strip())
124             res_product_name = regex.search(product_name.strip())
125             if res_description:
126                 size = re.sub("[0-9\\.\s]", "", str(res_description.group()))
127             elif res_product_name:
128                 size = re.sub("[0-9\\.\s]", "", str(res_product_name.group()))
129             #print(size)
130             if res_description or res_product_name:
131                 unit = str(row[1]).encode("utf-8")
132                 rdf_class = str(row[2]).split("/")[-1].split(".")[0]
133                 sizes.append({"size":size, "unit":unit, "class":rdf_class})
134
135         #print(sizes)
136         for size_dict in sizes:
137
138             converted_size = None
139             converted_unit = None
140             conversion_factor = None
141
142             # get smallest class of measurement for conversion
143             individual_of_smallest_unit = None
144             label_of_smallest_unit = None

```

Fonte: Elaborada pelo autor.

Figura 14: SPARQL para consulta de expressões regulares com base no RDF

```

"""SELECT ?regex ?label ?individual
{SELECT DISTINCT ?regex ?label ?individual
WHERE {
    ?individual units:regex ?regex .
    ?individual rdfs:label ?label .
    FILTER(LANG(?label) = "") .
} ORDER BY ?label} GROUP BY ?regex"""

```

Fonte: Elaborada pelo autor.

Figura 15: RDF contendo a expressão regular para a unidade Litro

```

<NamedIndividual rdf:about="&units:Liter">
  <rdf:type rdf:resource="&units:VolumeUnit"/>
  <rdfs:label>L</rdfs:label>
  <rdfs:label xml:lang="de">Liter</rdfs:label>
  <rdfs:label xml:lang="pt">Litro</rdfs:label>
  <rdfs:label xml:lang="en">Liter</rdfs:label>
  <units:conversionFactorToGallon rdf:datatype="&xsd;float">0.22</units:conversionFactorToGallon>
  <units:conversionFactorToPint rdf:datatype="&xsd;float">0.47317648</units:conversionFactorToPint>
  <units:conversionFactorToQuart rdf:datatype="&xsd;float">1.0566882</units:conversionFactorToQuart>
  <units:conversionFactorToMilliliter rdf:datatype="&xsd;float">1000.0</units:conversionFactorToMilliliter>
  <units:conversionFactorToFluidOunce rdf:datatype="&xsd;float">35.2</units:conversionFactorToFluidOunce>
  <units:conversionFactorToCup rdf:datatype="&xsd;float">4.2267528</units:conversionFactorToCup>
  <units:regex>\d+(,|\.)?\d*\s?(l|L)(\s|$|\s|</units:regex>
  <units:isLargerThan rdf:resource="&units:Cup"/>
  <units:isLargerThan rdf:resource="&units:FluidOunce"/>
  <units:isSmallerThan rdf:resource="&units:Gallon"/>
  <units:isLargerThan rdf:resource="&units:Milliliter"/>
  <units:isLargerThan rdf:resource="&units:Pin">
  <units:isSmallerThan rdf:resource="&units:Quart"/>
</NamedIndividual>

```

Fonte: Elaborada pelo autor.

Figura 16: Página Inicial do Protótipo WEB

**Projeto Unity**

Departamento:  Ajuda

Categoria:

Subcategoria:

Descrição Produto:

Valor de Atributos:

Atributo Preço Por Unidade:

Lista de Produtos						
Departamento	Categoria	Subcategoria	Descrição	Unidades	Preço	Preço por Unidade
ALIMENTOS	FRIOS E LATICINEOS	Manteiga	Manteiga Ghee Me More Sal Rosa - 200g	1	R\$27.90	R\$0,1395/g
ALIMENTOS	GRAOS E SEMENTES	Pipoca	Pipoca para Microondas Pop Corn Cobertura Chocolate 160g - Yoki	1	R\$4,34	R\$0,0271/g
ALIMENTOS	CONDIMENTOS	Zimbro	Zimbro - villa Cerroni - 70 g	1	R\$22.90	R\$0,3271/g
ALIMENTOS	FRIOS E LATICINEOS	Manteiga	Manteiga Ghee Me More Ervas Finas - 200g	1	R\$27.90	R\$0,1395/g
ALIMENTOS	DOCES	Geléia	Geléia premium banana caramelizada 310g	1	R\$11,64	R\$0,0375/g
ALIMENTOS	BEBIDAS NÃO ALCOOL	Água	Água Voss Sparkling Gaseificada 375ml	1	R\$29.70	R\$0,0792/mL
ALIMENTOS	GRAOS E SEMENTES	Sementes	Sementes Amendoim Forrageiro Caixa com 1,0 kg	1	R\$239,00	R\$239,00/kg
ALIMENTOS	DOCES	Base	Base Líquida Tutti Frutti 970ml -	1	R\$7,38	R\$0,0076/mL

Patterns

Fonte: Elaborada pelo autor.

Figura 17: Exemplo de pesquisa no Protótipo WEB

**Projeto Unity**

Departamento:  Ajuda

Categoria:

Subcategoria:

Descrição Produto:

Valor de Atributos:

Atributo Preço Por Unidade:

Lista de Produtos						
Departamento	Categoria	Subcategoria	Descrição	Unidades	Preço	Preço por mL
ALIMENTOS	BEBIDAS NÃO ALCOOL.	Água	Água Mineral Perrier Gaseificada 330ml	1	R\$14,96	R\$0,0453/mL
ALIMENTOS	BEBIDAS NÃO ALCOOL.	Água	Água Mineral Evian Sem Gás 330ml	1	R\$8,25	R\$0,025/mL

1

Patterns

Fonte: Elaborada pelo autor.

**APÊNDICE B – QUESTIONÁRIO DE ACEITAÇÃO DA FERRAMENTA PROTÓTIPO WEB**

Neste Apêndice é apresentado o questionário elaborado com base no modelo TAM, que foi submetido a alunos da Unisinos para análise da aceitação do protótipo WEB.

Figura 18: Questionário - Página 1

## Uma Arquitetura para Extração de Unidades de Produtos em Lojas Virtuais Brasileiras.

### Objetivo:

Esse questionário tem como objetivo avaliar a aceitação de um protótipo de uma ferramenta para comparação de preços de produtos (Protótipo Unity), desenvolvido como Trabalho de Conclusão do Curso de Sistemas de Informação da Universidade do Vale do Rio dos Sinos (UNISINOS).

Link com o protótipo: <http://vps6546.publiccloud.com.br/unity/>

Contato: Márcio Kolling - [kolling.marciop@gmail.com](mailto:kolling.marciop@gmail.com)

\* Required

*Skip to question 1.*

### Sobre você

#### 1. Idade \*

---

#### 2. Sexo \*

*Mark only one oval.*

- Masculino  
 Feminino  
 Prefiro não responder

#### 3. Maior Formação Concluída \*

*Mark only one oval.*

- Pós-Doutorado  
 Doutorado  
 Mestrado  
 Especialização  
 Graduação  
 Técnico/Ensino Médio  
 Fundamental

#### 4. Área de Atuação

---

#### 5. Costuma comprar na Internet? \*

*Mark only one oval.*

- Sim  
 Não

Fonte: Elaborada pelo autor.

Figura 19: Questionário - Página 2

Uma Arquitetura para Extração de Unidades de Produtos em Lojas Virtuais Brasileiras.

6. Caso a resposta anterior seja negativa, explique por quê?

---



---



---



---



---

7. Faz pesquisa de preço na Internet antes de realizar uma compra? \*

Mark only one oval.

- Sim
- Não

8. Caso a resposta anterior seja negativa, explique por quê?

---



---



---



---



---

9. Utiliza portais de comparativo de preços como o Buscapé e BondFaro? \*

Mark only one oval.

- Sim
- Não

10. Caso a resposta anterior seja negativa, explique por quê?

---



---



---



---



---

## Sobre o Protótipo

### Facilidade de uso percebida

---

11. Aprender a usar a ferramenta de comparação de preços Unity seria fácil para mim. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

Fonte: Elaborada pelo autor.

Figura 20: Questionário - Página 3

12. **Eu sei facilmente como proceder com a ferramenta Unity para fazer o que eu quero. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

13. **A maneira de interação com a ferramenta Unity é clara e facilmente compreendida. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

14. **A ferramenta Unity é flexível para que eu possa usá-la da maneira que melhor me convier. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

15. **Seria fácil ficar mais habilidoso no uso da ferramenta Unity. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

16. **Eu considero ferramentas de comparação de preços fáceis de usar. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

## Sobre o Protótipo

### Utilidade percebida

---

17. **Usando a ferramenta Unity meu trabalho de pesquisa e comparação de preços fica mais rápido. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

Fonte: Elaborada pelo autor.

Figura 21: Questionário - Página 4

18. Usando a ferramenta Unity meu desempenho de pesquisa e comparação de preços melhora. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

19. Usando a ferramenta Unity minha produtividade aumenta. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

20. Meu trabalho de comparação de preços fica mais efetivo (eficiente e eficaz) usando a ferramenta Unity. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

21. Meu trabalho de comparação de preços fica mais fácil usando a ferramenta Unity. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

22. As ferramentas de comparação de preço são úteis para mim. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

## Sobre o Protótipo

### Atitude para uso

---

23. Utilizar a ferramenta Unity é uma ótima ideia. \*

Mark only one oval.

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

Fonte: Elaborada pelo autor.

Figura 22: Questionário - Página 5

24. **Eu desejo utilizar a ferramenta Unity. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

25. **Seria muito melhor utilizar a ferramenta Unity. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

26. **Eu gosto da ideia de utilizar a ferramenta Unity no trabalho de comparação de preços. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

## Sobre o Protótipo

### Intenção Comportamental

---

27. **Eu pretendo utilizar a ferramenta Unity, sempre que possível. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

28. **Eu tenho a intenção de aumentar o uso de ferramentas de comparação de preço. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

29. **Eu adotaria a ferramenta Unity, no futuro. \***

*Mark only one oval.*

	1	2	3	4	5	
Discordo totalmente	<input type="radio"/>	Concordo totalmente				

Fonte: Elaborada pelo autor.

Figura 23: Questionário - Página 6

**30. Observações/comentários**

---

---

---

---

---

---

Powered by  
 Google Forms

Fonte: Elaborada pelo autor.

## APÊNDICE C – COMENTÁRIOS DOS RESPONDENTES AO QUESTIONÁRIO

Os respondentes deixaram as observações e comentários a seguir:

- Gosto de ferramentas de comparação e pesquisa de preço e utilizo sempre que possível. Sobre o Unity, ele poderia ter um filtro melhor sobre as medidas. Por exemplo a categoria computador, não fica claro se os GB são de memória ou armazenamento. E nem especifica qual tipo de armazenamento.
- No preço por unidade poderia ter a opção de selecionar a unidade. Ex: 1/ml ou 1000/L
- Excelente!!
- Ótima ideia da ferramenta. Seria mais interessante o projeto ter um design mais intuitivo e interessante, visto que está muito simples, mas creio que os criadores já estão cientes disso. E sobre o questionário, seria mais interessante ser mais objetivo, algumas perguntas acabaram sendo muito parecidas, deixando o questionário um pouco mais confuso do que poderia ser. De resto, o projeto tem tudo para dar certo. Boa sorte.
- O preço por unidade ficou meio confuso a utilização para mim.
- Não sei se é uma ideia para o futuro, mas seria interessante adicionar uma ferramenta que diz o nome da loja na qual o produto está (ou eu não notei pelo menos).
- Entendo que este é apenas um projeto, mas os sites atuais de comparação de preço possuem muito mais informações sobre os produtos que estão sendo comparados (comparação de especificações técnicas, etc..)
- A ferramenta, como melhoria, pode ser adaptada para dispositivos móveis, e ter um help mais próximo das caixas de seleção
- A funcionalidade que eu mais uso nestes filtros online não possui na ferramenta unity, que é o range de preços (de R\$ XX,XX até R\$ XXX,XX) e também uma sugestão de melhoria é a de limpar todos os campos do filtro que forem dependentes de um campo que for alterado, por exemplo, quando o campo subcategoria for alterado, limpar todos os campos abaixo (pois os valores de atributos são diferentes para diferentes tipos de subcategorias) e quando o departamento for alterado limpar todos. No geral eu gostei, é uma ferramenta bem interessante, já utilizei no passado o Buscapé para pesquisar preços mas hoje não utilizo mais e não pretendo utilizar este tipo de ferramentas por enquanto.
- Bem prático e rápido.
- Achei o formulário confuso, uma mudança interessante seria limpar os campos mais específicos quando selecionando os menos específicos, exemplo, estou pesquisando por

refrigerantes e vou para computadores, limpar o Valor de Atributos e o Atributo Preço Por Unidade, pois não faz sentido medir computador por ml

- Achei a interface muito simples e direta, realmente é uma ferramenta super útil na realidade Brasileira. Se eu puder fazer uma crítica construtiva, as perguntas em geral são muito repetitivas, pois a resposta de uma torna óbvia a próxima resposta na maioria dos casos, porém entendo que muitas vezes não temos como fugir de alguma didática específica. Parabéns pela criação da ferramenta!
- Trabalharia um pouco mais no layout da ferramenta.