

# **A Produção do Conhecimento na Engenharia da Computação**

**Ernane Rosa Martins  
(Organizador)**

 **Atena**  
Editora  
Ano 2019

Ernane Rosa Martins  
(Organizador)

# A Produção do Conhecimento na Engenharia da Computação

Atena Editora  
2019

2019 by Atena Editora  
Copyright © Atena Editora  
Copyright do Texto © 2019 Os Autores  
Copyright da Edição © 2019 Atena Editora  
Editora Executiva: Prof<sup>a</sup> Dr<sup>a</sup> Antonella Carvalho de Oliveira  
Diagramação: Rafael Sandrini Filho  
Edição de Arte: Lorena Prestes  
Revisão: Os Autores

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores. Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

### **Conselho Editorial**

#### **Ciências Humanas e Sociais Aplicadas**

Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas  
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná  
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília  
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa  
Prof<sup>a</sup> Dr<sup>a</sup> Cristina Gaio – Universidade de Lisboa  
Prof. Dr. Deyvison de Lima Oliveira – Universidade Federal de Rondônia  
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná  
Prof<sup>a</sup> Dr<sup>a</sup> Ivone Goulart Lopes – Istituto Internazionale delle Figlie de Maria Ausiliatrice  
Prof<sup>a</sup> Dr<sup>a</sup> Juliane Sant’Ana Bento – Universidade Federal do Rio Grande do Sul  
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense  
Prof<sup>a</sup> Dr<sup>a</sup> Lina Maria Gonçalves – Universidade Federal do Tocantins  
Prof<sup>a</sup> Dr<sup>a</sup> Natiéli Piovesan – Instituto Federal do Rio Grande do Norte  
Prof<sup>a</sup> Dr<sup>a</sup> Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa  
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará  
Prof<sup>a</sup> Dr<sup>a</sup> Vanessa Bordin Viera – Universidade Federal de Campina Grande  
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

#### **Ciências Agrárias e Multidisciplinar**

Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul  
Prof. Dr. Alexandre Igor Azevedo Pereira – Instituto Federal Goiano  
Prof<sup>a</sup> Dr<sup>a</sup> Daiane Garabeli Trojan – Universidade Norte do Paraná  
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista  
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul  
Prof<sup>a</sup> Dr<sup>a</sup> Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia  
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul  
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará  
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas

### **Ciências Biológicas e da Saúde**

Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria  
Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás  
Prof.<sup>a</sup> Dr.<sup>a</sup> Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina  
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará  
Prof.<sup>a</sup> Dr.<sup>a</sup> Natiéli Piovesan – Instituto Federal do Rio Grande do Norte  
Prof.<sup>a</sup> Dr.<sup>a</sup> Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão  
Prof.<sup>a</sup> Dr.<sup>a</sup> Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa  
Prof.<sup>a</sup> Dr.<sup>a</sup> Vanessa Bordin Viera – Universidade Federal de Campina Grande

### **Ciências Exatas e da Terra e Engenharias**

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto  
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná  
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará  
Prof.<sup>a</sup> Dr.<sup>a</sup> Natiéli Piovesan – Instituto Federal do Rio Grande do Norte  
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista

### **Conselho Técnico Científico**

Prof. Msc. Abrãao Carvalho Nogueira – Universidade Federal do Espírito Santo  
Prof.<sup>a</sup> Dr.<sup>a</sup> Andreza Lopes – Instituto de Pesquisa e Desenvolvimento Acadêmico  
Prof. Msc. Carlos Antônio dos Santos – Universidade Federal Rural do Rio de Janeiro  
Prof.<sup>a</sup> Msc. Jaqueline Oliveira Rezende – Universidade Federal de Uberlândia  
Prof. Msc. Leonardo Tullio – Universidade Estadual de Ponta Grossa  
Prof. Dr. Welleson Feitosa Gazel – Universidade Paulista  
Prof. Msc. André Flávio Gonçalves Silva – Universidade Federal do Maranhão  
Prof.<sup>a</sup> Msc. Renata Luciane Polsaque Young Blood – UniSecal  
Prof. Msc. Daniel da Silva Miranda – Universidade Federal do Pará

<b>Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)</b>	
P964	A produção do conhecimento na engenharia da computação [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa, PR: Atena Editora, 2019.  Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web Inclui bibliografia ISBN 978-85-7247-339-2 DOI 10.22533/at.ed392192405  1. Computação – Pesquisa – Brasil. 2. Sistemas de informação gerencial. 3. Tecnologia da informação. I. Martins, Ernane Rosa. CDD 004
<b>Elaborado por Maurício Amormino Júnior – CRB6/2422</b>	

Atena Editora  
Ponta Grossa – Paraná - Brasil  
[www.atenaeditora.com.br](http://www.atenaeditora.com.br)  
contato@atenaeditora.com.br

## APRESENTAÇÃO

Segundo o dicionário Aurélio a Engenharia é a “Arte de aplicar conhecimentos científicos e empíricos e certas habilitações específicas à criação de estruturas, dispositivos e processos que se utilizam para converter recursos naturais em formas adequadas ao atendimento das necessidades humanas. A Engenharia de Computação é definida como o ramo da engenharia que se caracteriza pelo projeto, desenvolvimento e implementação de sistemas, equipamentos e dispositivos computacionais segundo uma visão integrada de hardware e software, apoiando-se em uma sólida base matemática e conhecimentos de fenômenos físicos.

Este livro, possibilita conhecer algumas das produções do conhecimento no ramo da Engenharia da Computação, que abordam assuntos extremamente importantes, tais como: as transformações sofridas nos processos de projeto desde a implementação das ferramentas digitais; o armazenamento, indexação e recuperação de formulários digitais; a reabilitação motora assistida por computadores; a reflexão acerca do realismo e da representação visual em jogos digitais; os padrões de players em ambientes virtuais; as soluções tecnológicas relevantes usadas em países africanos; a complexa relação existente entre jogos digitais e o humano; a dinâmica da comunicação de um grupo de Facebook criado em um processo de urbanismo bottom-up; o estado da arte das pesquisas e estudos acadêmicos acerca dos elementos visuais contidos na interface de jogos digitais; as estratégias de design que integrem tecnologia computacional digital a artefatos e instalações para a interação de visitantes em museus; os jogos que abordam o tema de mitologia e religião.

Deste modo, espero que este livro seja um guia para os Engenheiros de Computação auxiliando-os em assuntos relevantes da área, fornecendo conhecimentos que podem permitir especificar, conceber, desenvolver, implementar, adaptar, produzir, industrializar, instalar e manter sistemas computacionais, bem como perfazer a integração de recursos físicos e lógicos necessários para o atendimento das necessidades informacionais, computacionais e da automação de organizações em geral. Por fim, agradeço a todos que contribuíram de alguma forma para a construção desta obra e desejo a todos os leitores, novas e significativas reflexões sobre os temas abordados.

Ernane Rosa Martins

## SUMÁRIO

<b>CAPÍTULO 1</b> .....	<b>1</b>
VIRTUAL REALITY AS A TOOL TO REGAIN TACTUAL PROCEDURES IN DIGITAL DESIGN	
Tales Lobosco	
DOI 10.22533/at.ed3921924051	
<b>CAPÍTULO 2</b> .....	<b>15</b>
UMA PROPOSTA DE SISTEMA DE BUSCA PARA RECUPERAÇÃO DE FORMULÁRIOS DIGITAIS	
Afonso Henrique Anastácio Calábria	
Talles Brito Viana	
DOI 10.22533/at.ed3921924052	
<b>CAPÍTULO 3</b> .....	<b>26</b>
REVISÃO SISTEMÁTICA: APLICABILIDADE DO MS KINECT EM REABILITAÇÃO MOTORA	
Tiago Pereira Remédio	
Alexandro José Baldassin	
DOI 10.22533/at.ed3921924053	
<b>CAPÍTULO 4</b> .....	<b>43</b>
REFLEXÕES ACERCA DO REALISMO E DA REPRESENTAÇÃO VISUAL EM GAMES	
TENDÊNCIAS DE MERCADO E JOGOS AAA	
Ana Carolina Generoso de Aquino	
Rosane de Fatima Antunes Obregon	
Heitor Dias Couto	
DOI 10.22533/at.ed3921924054	
<b>CAPÍTULO 5</b> .....	<b>52</b>
PLAYER GAME DATA MINING FOR PLAYER CLASSIFICATION	
Bruno Almeida Odierna	
Ismar Frango Silveira	
DOI 10.22533/at.ed3921924055	
<b>CAPÍTULO 6</b> .....	<b>62</b>
INTERAÇÃO DA TECNOLGIA NA ÁFRICA	
Welington dos Santos Ayres	
DOI 10.22533/at.ed3921924056	
<b>CAPÍTULO 7</b> .....	<b>69</b>
INTEGRAÇÃO DA TECNOLOGIA CUDA AO MODELO DE PREVISÃO DO TEMPO ETA	
Henrique Gavioli Flores	
Alex Lima de Mello	
Marcelo Trindade Rebonatto	
Carlos Amaral Hölbig	
DOI 10.22533/at.ed3921924057	
<b>CAPÍTULO 8</b> .....	<b>84</b>
HORIZONTES DA PESQUISA EM CULTURA DE GAMES SOB A ESTÉTICA DA PRODUÇÃO	
Nilson Valdevino Soares	
Luís Carlos Petry	
Guilherme Sousa Vieira	

Ana Carolina Simões de Freitas Cabral  
Felipe Blanco  
Saulo de Oliveira Machado  
José Guilherme dos Santos Gomes

DOI 10.22533/at.ed3921924058

**CAPÍTULO 9 ..... 100**

HORIZONTAL DIALOGUES AND OPEN DATA: THE COMMUNICATION SPACES OF BOTTOM-UP URBANISM.

José Eduardo Calijuri Hamra

DOI 10.22533/at.ed3921924059

**CAPÍTULO 10 ..... 115**

ELEMENTOS VISUAIS EM JOGOS DIGITAIS: UMA REVISÃO SISTEMÁTICA DA LITERATURA.

Ana Carolina Generoso de Aquino

Rosane de Fatima Antunes Obregon

DOI 10.22533/at.ed39219240510

**CAPÍTULO 11 ..... 131**

MEDIAÇÃO DE CONTEÚDO E TECNOLOGIA DIGITAL EM MUSEUS: ESTRATÉGIAS PROJETAIS PARA ENRIQUECIMENTO DA EXPERIÊNCIA DO VISITANTE.

Diego Enéas Peres Ricca

Clíce de Toledo Sanjar Mazzilli

DOI 10.22533/at.ed39219240511

**CAPÍTULO 12 ..... 151**

BRINCANDO COM OS DEUSES: A VIABILIDADE DA DISSEMINAÇÃO DA CULTURA FOLCLÓRICA E POPULAR AFRO-BRASILEIRA EM JOGOS DIGITAIS.

Igor Rocha dos Santos

Marcos Wendell S. de O. Santos

Larissa Cardillo Acconcia Dias

Maurício Acconcia Dias

DOI 10.22533/at.ed39219240512

**CAPÍTULO 13 ..... 166**

A OBRA DANTESCA E SEMIOSES DA CULTURA DE JOGOS DE VIDEOGAME: REFLEXOS EM QUESTÕES DE LETRAMENTO

Caio Túlio Olímpio Pereira da Costa

Leandro Paz da Silva

DOI 10.22533/at.ed39219240513

**CAPÍTULO 14 ..... 176**

A BATTLING BEHAVIOR ANALYSIS OF SHOOTER GAMES BOTS BASED ON THE BARTLE'S PLAYER TYPES AND FINITE STATE MACHINES

Felipe Oviedo Frosi

Isabel Cristina Siqueira da Silva

DOI 10.22533/at.ed39219240514

**SOBRE O ORGANIZADOR..... 194**

## UMA PROPOSTA DE SISTEMA DE BUSCA PARA RECUPERAÇÃO DE FORMULÁRIOS DIGITAIS

**Afonso Henrique Anastácio Calábria**

Instituto Federal do Ceará – Campus Crato  
Crato-CE, Brasil

**Talles Brito Viana**

Instituto Federal do Ceará – Campus Crato  
Crato-CE, Brasil

**RESUMO:** A grande quantidade de documentação gerada dentro das organizações torna difícil a tarefa de armazenar, recuperar ou acompanhar o processamento dos documentos dentro das organizações. Neste cenário, destaca-se um tipo de documento: os formulários, estes que na maioria das vezes são criados digitalmente, em seguida são impressos em papel e transmitidos entre pessoas e departamentos de uma organização de acordo com um fluxo de processo organizacional. Com o objetivo de tratar os problemas de precisão, eficiência e desempenho na manipulação de formulários digitais dentro das organizações, este trabalho propõe um Sistema de Busca para recuperação de formulários digitais, com ênfase na descrição semiestruturadas e indexação dos mesmos. Especificamente, neste trabalho é proposta uma nova abordagem para o armazenamento, indexação e recuperação de formulários digitais baseada na definição de: interfaces de usuário para criação de formulários digitais, repositórios de formulários

digitais, motores de indexação e busca de formulários digitais, e por fim, interfaces de usuário para a consulta e recuperação de formulários digitais. Espera-se que o emprego do sistema proposto vise cessar o uso do papel ao longo do tempo, e assim, facilitar a gestão dos documentos digitais das organizações.

**PALAVRAS-CHAVE:** Formulários Digitais, XFDL, Indexação, Recuperação de Informação, Sistemas de Busca.

**ABSTRACT:** The larger quantity of generated documentation within organizations makes difficult to store, retrieve and track the processing of documents. In such a scenario, it is important to note a kind of document: the forms, in which these most often are created digitally, then are printed on paper and transmitted between people and departments in an organization according to an organizational process flow. In order to address the problems of accuracy, efficiency and performance when handling forms within organizations, this paper proposes a search system for recovering digital forms, with an emphasis on forms semi-structured description and indexing. Specifically, this paper proposes a new approach for storage, indexing and retrieval of digital forms based on the definition of: user interfaces for creating digital forms, repositories of digital forms, search engines for indexing digital forms, and finally,



user interfaces for query and retrieval of digital forms. It is expected that the adoption of the proposed system aims to cease the use of paper over time, and thus to facilitate the management of digital documents of organizations.

**KEYWORDS:** Digital Forms, XFDL, Indexing, Information Retrieval, Search Systems.

## 1 | INTRODUÇÃO

A produção de documentos e informações na sociedade atual é cada vez maior (OLIVEIRA, 2014), então, neste cenário surge o *Gerenciamento Eletrônico de Documentos* (GED). O gerenciamento eletrônico de documentos visa gerir o ciclo de vida das informações documentais de uma organização desde a criação de um documento até o armazenamento (arquivamento do mesmo). Considerando que em geral as organizações fazem uso de documentos nos processos internos, então gerir tais documentos pode ser visto como um meio de tratar e entender toda a vida informacional da organização (PENA; SILVA, 2004).

Considerando que quantidade substancial dos documentos das organizações é ainda tratada de forma de papel, dessa forma, surge a seguinte problemática: Quão difícil é armazenar documentos na forma de papel? Em (PALETTA; DIAS, 2015), um documento é definido como qualquer base de conhecimento, fixada materialmente e disposta de maneira que se possa utilizar para consulta, estudo, prova, dentre outros. A grande quantidade de documentação gerada dentro das organizações, como por exemplo, formulários, memorandos, requisições e documentação fiscal, torna difícil a tarefa de armazenar, recuperar ou acompanhar o processamento dos documentos dentro da organização (ANDRADE, 2002). Especialmente destaca-se um tipo de documento: os formulários, estes que na maioria das vezes são criados digitalmente, em seguida são impressos em papel e transmitidos entre pessoas e departamentos de uma organização para que um fluxo de processo organizacional seja seguido de acordo com as informações preenchidas nos mesmos. Assim, formulários seguem de forma física para a próxima pessoa responsável por avaliar, encaminhar ou rejeitar determinada solicitação. O grande problema do uso formulários impressos em papel está na dificuldade para entregar e repassar estes documentos, bem como, na dificuldade de recuperar e processar informações de formulários previamente criados e armazenados. A utilização de formulários em papel gera problemas relacionados à comunicação, manutenção e recuperação de tais documentos, tornando os processos das organizações mais ineficientes.

No entanto, sistemas de Gerenciamento Eletrônico de Documentos podem reduzir o trabalho humano no armazenamento e na recuperação de documentos físicos, de forma a aumentar a eficiência e a precisão na recuperação dos documentos (JERVIS; MASOODIAN, 2009). Nesse contexto, este trabalho propõe um Sistema de Busca para recuperação de documentos, especificamente, formulários digitais, com ênfase na descrição semiestruturada de formulários digitais e na indexação dos mesmos,

com o objetivo de facilitar a gestão de documentos digitais.

Este trabalho está organizado da seguinte forma: Na Seção 2 são apresentados trabalhos relacionados anteriormente propostos na literatura que já tentaram resolver o problema de recuperação e indexação de formulários digitais. Dadas as deficiências destas abordagens existentes na literatura, na Seção 3 é proposto e discutido um novo sistema de busca para recuperação de formulários digitais. Por fim, algumas considerações finais são apresentadas na Seção 4.

## 2 | TRABALHOS RELACIONADOS

A tecnologia de *Reconhecimento Óptico de Caracteres* (OCR) (MITHE et al., 2013) é empregada em diversos trabalhos relacionados ao gerenciamento eletrônico de documentos. As tecnologias de reconhecimento óptico de caracteres podem ser basicamente vistas como softwares que implementam um conjunto de algoritmos e modelos matemáticos que visam converter documentos com caracteres impressos em documentos digitais. Como discutido em (KAE; LEARNED-MILLER, 2009), apesar de diversos autores afirmarem que o reconhecimento óptico de caracteres é um problema solucionado, a variação das formas dos documentos continuam a fazer com que os modernos softwares de reconhecimento óptico de caracteres falhem devido à degradação moderada dos documentos com o passar do tempo e à dificuldade de reconhecimento de fontes obsoletas.

Em contrapartida, ao invés de empregar algum tipo de reconhecimento automático a partir do conteúdo dos documentos, outras abordagens propostas para o gerenciamento eletrônico de documentos são baseadas em ontologias para a classificação, indexação e recuperação de documentos. Uma ontologia é criada por especialistas e define as regras que regulam a combinação entre termos e relações em um domínio do conhecimento (ALMEIDA et al., 2003). O emprego de ontologias pode ser visto como um método em que termos e relações de uma ontologia devem ser associados a um documento ou partes do mesmo, com o objetivo de relacionar informações dos documentos com conceitos (em geral, conceitos do mundo real) pré-catalogados. Por exemplo, no protótipo YOUFILE (BARCETTI et al., 2008), para que os documentos sejam indexados de maneira adequada é necessário que um operador humano associe manualmente os campos (ou partes) dos documentos com informações correspondentes aos termos e relações pré-definidas através de ontologias. Tal método baseado em ontologias permite que grupos de documentos sejam indexados e, após isso, documentos sejam recuperados por meio da busca de documentos associados aos conceitos e termos definidos nas ontologias. Apesar disso, o desempenho e a eficiência são dois fatores que em geral dificultam o emprego dessa técnica na prática. Os usuários têm que ter conhecimento sobre os conceitos, termos, relações e atributos das ontologias e um processo manual para classificar os documentos é bastante custoso e pouco eficiente (BARCETTI et al., 2008). Já

em relação ao desempenho computacional, como apontado em (DING et al., 2004), os algoritmos para executar inferência sob uma base de dados de ontologias são bastante custosos, o que torna lento o processamento de consulta e recuperação de documentos.

Por fim, uma terceira tecnologia relacionada ao gerenciamento eletrônico de documento é denominada como *Reconhecimento de Manuscritos* (PLAMODON et al., 2000), este pode ser definido como métodos e algoritmos para converter informações escritas em caligrafia humana em caracteres digitais. A razão pela qual os manuscritos persistem na era digital é a conveniência que o papel e caneta possuem em comparação com teclados em inúmeras situações do dia-a-dia (PLAMODON et al., 2000). Entretanto, devido à quantidade e diversidade da capacidade de escrita humana, seria demasiadamente custoso o processo de conversão de uma base de dados de documentos legados manuscritos ainda não digitalizados. Apesar disso, no sistema proposto em (BARRUS et al., 2014) é demonstrado que o preenchimento de formulários através da captura de dados produzidos por informação manuscrita, via uma *stylus pen* em dispositivos móveis como um *tablet*, é uma evidência de que de fato o papel pode ser substituível de forma funcional e ecológica. Assim, ao passo que novas tecnologias surgem, talvez no futuro o papel perca espaço no cotidiano das organizações, bem como a produção de papel seja reduzida.

### 3 | PROPOSTA

Sendo assim, com o objetivo de tratar os problemas de precisão, eficiência e desempenho na manipulação de formulários digitais dos trabalhos existentes na literatura, este trabalho apresenta uma nova proposta que consiste na utilização de um modelo para descrição, armazenamento e recuperação de formulários digitais.

Para a descrição dos formulários digitais propõe-se o emprego de uma linguagem de descrição padronizada de formulários digitais, denominada XFDL (*Extensible Forms Description Language*) (BOYER et al., 1998). Como esta linguagem pode ser empregada na descrição de formulários digitais é discutido na Seção 3.1. Além disso, propõe-se que repositórios de formulários digitais sejam indexados através de um novo sistema de busca a fim de tratar os problemas anteriormente citados. O sistema de busca proposto para a recuperação de formulários digitais é apresentado e discutido na Seção 3.2.

#### 3.1 DESCRIÇÃO DE FORMULÁRIOS DIGITAIS

O XFDL (*Extensible Forms Description Language*) [13] é uma linguagem que possibilita descrever e tratar a representação de documentos digitais do tipo formulário. XFDL é uma linguagem baseada na linguagem XML (*eXtensible Markup Language*) (BRAY et al., 1998) que descreve sintaticamente a estrutura de formulários digitais,

visando lidar com problemas como: precisão de layout, cálculos integrados, validação de *inputs*, múltiplas assinaturas digitais e suporte a transações (BOYER et al., 1998). O XFDL assim como XML e outras linguagens de marcação, possui, por exemplo, *tags* para a descrição dos elementos do formulário. Porém, é importante ressaltar que para cada tipo de formulário podem existir variadas instâncias. Ou seja, uma instância de um formulário é o mesmo formulário em adição dos dados preenchidos.

O sistema de busca proposto neste trabalho lida com as descrições padronizadas dos formulários bem como as com instâncias de formulários. Em específico, é proposto que tanto os formulários digitais não preenchidos (formulários “*em branco*”) bem como instâncias de formulários devem ser armazenados em uma fonte de dados para posterior indexação e recuperação dos mesmos de acordo com necessidades de informação do usuário.

Discutiremos um exemplo prático: supondo que um aluno de uma instituição de ensino superior deseje solicitar o requerimento de segunda chamada para uma prova de determinada disciplina, então é preciso que o mesmo preencha um formulário de requerimento. Este formulário é composto de campos que poderiam incluir: o nome do aluno, o curso que o aluno faz e a disciplina tal que o mesmo requer prova de segunda chamada.

A linguagem XFDL permite representar o formulário digital de segunda chamada em um arquivo XML de acordo com um formato definido padronizado. De forma geral, a linguagem permite descrever cada um dos campos e conteúdo do formulário. Na Figura 1 é mostrado um trecho da linguagem XFDL que descreve especificamente o campo do nome do aluno (de um hipotético formulário de segunda chamada).

```
<XFDL>
  <page>
    ...
    <field sid="aluno">
      <label>Nome do aluno: </label>
      <value></value>
      <format>
        <constraints>
          <message>Deve-se informar o nome completo do aluno
            que requer fazer segunda chamada. </message>
        </constraints>
      </format>
    </field>
    ...
  </page>
</XFDL>
```

Figura 1. Trecho XFDL que descreve o campo nome do aluno

É possível observar na Figura 1 que temos um *atributo* denominado ‘*sid*’ (do elemento *field*) que tem a função de associar cada elemento do formulário a um

identificador único (no caso mostrado temos que *sid="aluno"*). Além disso, um elemento *label* permite descrever os rótulos do campo de texto do formulário (no caso mostrado o rótulo é “*Nome do Aluno:*”). Por fim, é importante destacar que é possível descrever informações textuais em linguagem natural que detalham qual é a objetivo do campo dentro do contexto formulário e como ele poderia ser preenchido pelo usuário. Para isto, tal descrição é incluída dentro do elemento *message*, ilustrado na Figura 1.

No caso do exemplo mostrado na Figura 1 temos que o arquivo XFDL descreve um formulário sem dados (formulário “*em branco*”). Sendo assim, definimos que: Um *formulário vazio* tem valor vazio para todos os campos da descrição XFDL associada. Por exemplo, o formulário ilustrado na Figura 1 tem o elemento *value* do campo “*aluno*” vazio.

Neste contexto, definimos que uma *instância de um formulário vazio* representa um formulário que já teve os dados preenchidos. As instâncias são cópias de formulários vazios que são posteriormente modificados com a inserção dos valores dos campos. Na Figura 2 é mostrada uma instância do formulário vazio anteriormente apresentado na Figura 1, neste caso, temos que o valor do nome preenchido do aluno é “*João*”, conforme indicado através do elemento *value* do campo (*field*) “*aluno*”.

```
<XFDL>
...
<field sid="aluno">
  <label> Nome do aluno: </label>
  <value>João</value>
  ...
</field>
...
</XFDL>
```

Figura 2. Trecho XFDL que descreve o campo nome do aluno

### 3.2 SISTEMA DE BUSCA PROPOSTO

Neste trabalho é proposta uma nova abordagem para o armazenamento, indexação e recuperação de formulários digitais que é baseada na definição de: (i) interfaces de usuário para criação de formulários digitais; (ii) repositórios de descrições de formulários digitais; (iii) motores de indexação e busca de formulários digitais; (iv) interfaces de usuário para a consulta e recuperação de formulários digitais de acordo com as necessidades dos mesmos. Neste sentido, é proposto um novo sistema de busca de recuperação de formulários digitais com a arquitetura ilustrada na Figura 3. A arquitetura proposta é composta dos seguintes módulos: Interface de Criação e Preenchimento de Formulários, Repositórios de Formulários Digitais (repositórios de formulários vazios e preenchidos), Motor de Indexação, Motor de Busca e Interface

de Consulta e Recuperação. Cada um destes módulos é discutido em cada uma das subseções seguintes.

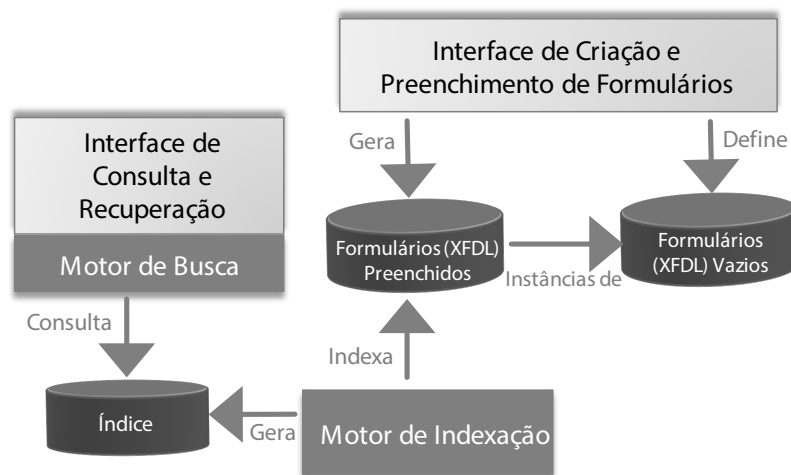


Figura 3. Arquitetura do sistema de busca proposto

**A) Criação e preenchimento de formulários:** A *Interface de Criação e Preenchimento de Formulários* constitui uma interface gráfica que permite criar novas descrições XFDL de formulário vazios e armazená-las em repositórios. Desta forma é possível definir variados tipos de formulários digitais e armazená-los. Além disso, esta mesma interface possibilita criar formulários XFDL preenchidos (instâncias de formulários vazios) a partir de repositórios de formulários vazios existentes. É importante ressaltar que esta interface deve ser portátil e estar disponível para ser utilizada em diferentes dispositivos, desde *desktops* até dispositivos móveis.

**B) Indexação de formulários digitais:** O *Motor de Indexação* pode ser visto como o módulo responsável por indexar os formulários preenchidos descritos em XFDL com a finalidade de gerar índices específicos a partir de um algoritmo especial, de forma a facilitar a posterior recuperação de formulários digitais. Como o XFDL se trata de um modelo baseado em XML, então é necessário empregar algoritmos para indexação da estrutura e conteúdo de arquivos XML, isto é, algoritmos para indexação de dados semiestruturados. Diversas técnicas para indexação de fontes de dados semiestruturadas têm sido propostas na literatura (por exemplo, tais como as técnicas (KOTSAKIS, 2002) e (BRITO; ELIAS, 2015)). Estas técnicas permitem manter a ocorrência de diversos caminhos de elementos extraídos a partir da fonte de dados XFDL, bem como, dos valores textuais armazenados nos documentos XFDL. Por exemplo, na Tabela 1 são destacados alguns caminhos extraídos a partir do exemplo XFDL da Figura 1.

Caminho	Significado	Tipo
<i>/XFDL/page/field/label</i>	Rótulo de um campo.	Descritivo
<i>/XFDL/page/field/format/constraints/message</i>	Objetivo do campo no contexto do formulário.	Descritivo
<i>/XFDL/page/field/value</i>	Valor do campo.	Valorado

Tabela 1. Exemplos de caminhos de descrições XFDL

Assim, os termos textuais contidos nas descrições XFDL são indexados em relação aos caminhos em que ocorrem. Cada caminho tem um significado associado (como mostrado na Tabela 1) e isto pode ser utilizado para aprimorar os resultados de busca e recuperação. Por exemplo, os termos textuais: “segunda”, “chamada” e “aluno” ocorrem dentro do caminho */XFDL/page/field/format/constraint/message* que representa qual é o objetivo do campo do nome do aluno no contexto formulário, ou seja, que o campo está relacionado com a prova de segunda chamada de um determinado aluno. Desta forma, este formulário deveria ser retornado caso procurado por formulários de segunda chamada através de uma interface de consulta.

**C) Consulta e recuperação de formulários:** Na arquitetura proposta é definido um *Motor de Busca*, que mediante uma função de ranqueamento (score) deve retornar os resultados mais relevantes a uma consulta feita por um usuário que procura por formulários digitais. O modelo de recuperação e ranqueamento proposto neste trabalho é fundamentado e baseado no algoritmo clássico proposto em (KOTSAKIS, 2002): palavras chaves devem ser informadas pelo usuário e procuradas em um índice invertido de termos textuais. Todos os caminhos tal que ocorrem as palavras chaves da consulta são descobertos e enumerados, e cada caminho tem um valor de peso associado que afeta o resultado do ranqueamento.

Neste trabalho é definido que caminhos das descrições XFDL podem ser classificados em relação aos seguintes dois tipos: (i) tipo *Descritivo*: caminhos relacionados à descrição dos campos, explanação sobre o objetivo dos campos e formatação visual do próprio formulário; (ii) tipo *Valorado*: caminhos relacionados aos valores preenchidos no formulário. Por exemplo, como listado na Tabela 1 temos que o caminho */XFDL/page/field/label* é do tipo *Descritivo* já que descreve os rótulos dos campos dos formulários. Já o caminho */XFDL/page/field/value* é do tipo *Valorado* pois descreve valores preenchidos no próprio formulário.

Sendo assim, definimos a seguinte heurística para o ranqueamento dos formulários: *no ranqueamento, os caminhos de tipo Descritivo devem contribuir com maior peso no score em relação aos caminhos de tipo Valorado*. Os efeitos esperados por esta heurística de ranqueamento são que os termos textuais de rótulos e explicações sobre o próprio formulário (isto é, termos em caminhos de tipo *Descritivo*) impactam mais no ranqueamento do que os termos textuais dos valores preenchidos (isto é, termos associados a caminhos de tipo *Valorado*). Assim, por exemplo, se consultado pelas palavras chaves: “João”, “segunda” e “chamada”, tem-

se que o formulário de segunda chamada da Figura 1 seria retornado no topo dos resultados de busca providos. Apesar disso, outros formulários referentes a “João” também poderiam ser retornados, mas ranqueados com valor de escore mais baixo, posicionados mais abaixo na lista de resultados de busca providos.

Por fim, é importante ressaltar que o módulo de *Interface de Consulta e Recuperação*, responsável por prover uma interface gráfica de interação do usuário com o *Motor de Busca*, também deve ser portátil e estar disponível para utilização em diferentes dispositivos, desde *desktops* até dispositivos móveis.

#### 4 | CONSIDERAÇÕES FINAIS

O sistema proposto neste trabalho emprega o modelo de descrição XFDL para representar o formato, bem como, conteúdo de formulários preenchidos de forma digital, com o objetivo de tornar mais eficiente a manipulação, armazenamento e recuperação dos mesmos. No contexto da aplicabilidade do sistema proposto, e, em relação ao significativo crescimento da popularização dos dispositivos móveis espera-se que o emprego do sistema proposto neste trabalho vise cessar o uso do papel ao longo do tempo.

Em relação às tecnologias baseadas em *OCR* (MITHE et al., 2013), podemos afirmar que o emprego do modelo de descrição baseado em XFDL remove a necessidade de reconhecer caracteres opticamente, uma vez que o sistema utilizará apenas o meio digital, sem a influência direta ou indireta do papel. Ou seja, já que os formulários serão criados e preenchidos através do sistema proposto, não há necessidade de digitalização de documentos na forma de papel.

Além disso, em comparação com tecnologias baseadas em *Reconhecimento de Manuscritos* (BARRUS et al., 2014), temos que ao considerarmos que o sistema proposto não conta com a utilização de papel, podemos concluir que o mesmo não enfrenta problemas de eficiência em relação ao processo de conversão de uma base de dados de documentos manuscritos, bem como, problemas de precisão no reconhecimento de caligrafia.

O sistema proposto opta por utilizar um algoritmo de recuperação de informação baseado na indexação de informações textuais em linguagem natural, bem como na indexação de caminhos XFDL, de forma a aprimorar o valor semântico dos termos textuais na indexação. Sabe-se que tal processo é computacionalmente mais eficiente do que fazer inferência em base de dados de ontologias (DING et al., 2004). Além disso, não ocorre o excessivo trabalho de classificação manual humana para relacionar partes dos documentos com conceitos de ontologias antes da indexação.

Por fim, é importante ressaltar que o escopo deste artigo está em definir uma proposta de um novo sistema de busca de formulários digitais em um alto nível de abstração. Como passo direto de continuação deste trabalho têm-se os seguintes



direcionamentos: primeiramente devemos citar a definição detalhada e formalizada da heurística de ranqueamento que será adotada no sistema proposto. Isto envolve o desenvolvimento de um mapeamento dos tipos de caminhos que podem ser extraídos ao utilizar descrições XFLD, bem como, a fixação expressa dos pesos de ranqueamento em relação a cada tipo de caminho. Em segundo, uma implementação da proposta apresentada neste trabalho será desenvolvida. Em estudos preliminares foi concluído que, em termos de viabilidade tecnológica, a concretização das ideias abstratas apresentadas neste trabalho é plenamente possível. Para isto, será empregado o uso da API Lucene (MCCANDLESS et al., 2010) na construção dos módulos dos Motores de Busca e Indexação. Lucene é uma API escalável e de alto desempenho para indexação e recuperação de informação.

Sendo assim, como trabalhos futuros, devemos citar o projeto e implementação do sistema proposto neste trabalho. Além disso, serão executados experimentos a partir de um repositório de formulários digitais gerados dentro de organizações reais, bem como, validação dos requisitos de usabilidade, eficiência e precisão de busca provida pelo sistema proposto.

## REFERÊNCIAS

- Oliveira, C.T. 2014. **O Gerenciamento Eletrônico de Documentos Sob a Ótica da Representação da Informação Arquivística**. Archeion Online 2.1.
- Pena, M., & Silva, A. 2004. **A Digitalização De Documentos Históricos E A Gestão Eletrônica De Documentos Para Disponibilização On Line**. Revista Vozes, p10.
- Paletta, F. C., & Dias, D. 2015. **Gestão Eletrônica De Documentos E Conteúdo**. Revista PRISMA. COM, p. 25.
- Andrade, M. 2002. **Gerenciamento eletrônico da informação: ferramenta para a gerência eficiente dos processos de trabalho**. Seminário Nacional de Bibliotecas Universitárias, Recife.
- Jervis, M., & Masoodian, M. 2009. **Digital management and retrieval of physical documents**. In Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (pp. 47-54). ACM.
- Mithe, R., Indalkar, S., & Divekar, N. 2013. **Optical Character Recognition**. International Journal of Recent Technology and Engineering (IJRTE) Volume, 2; p. 72-75.
- Kae, A., & Learned-Miller, E. 2009. **Learning On The Fly: Font-Free Approaches To Difficult OCR Problems**. In Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on (pp. 571-575). IEEE.
- Almeida, Mauricio B. and Bax, Marcello P. **An overview about ontologies: survey about definitions, types, applications, evaluation and building methods**. Ci. Inf. [online]. 2003, vol. 32, no. 3
- Barchetti, Ugo et al. 2008. **How Can Ontologies Support Enterprise Digital And Paper Archives?: A Case Study**. In: Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology. p. 627-636. ACM.

L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs. “**Swoogle: A semantic web search and metadata engine**”. In Proceedings of the thirteenth ACM international conference on Information and knowledge management (CIKM '04).

Plamondon, Réjean, and Sargur N. Srihari. 2000. **Online And Off-Line Handwriting Recognition: A Comprehensive Survey**. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.1 p. 63-84.

Barrus, John W. Schwartz, Edward L. 2014. **Image-Based Document Management: Aggregating Collections Of Handwritten Forms**. In: Proceedings of the 2014 ACM symposium on Document engineering. ACM, p. 117-120.

Boyer, J., Bray, T., & Gordon, M. 1998. **Extensible Forms Description Language (XFDL) 4.0**. Draft Specification NOTE-XFDL-19980902, W3C. Acessado em Agosto de 2015: <http://www.w3.org/TR/NOTE-XFDL>

Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., & Yergeau, F. 1998. **Extensible Markup Language (XML) 1.0**. W3C. Available at <http://www.w3.org/TR/1998>.

Kotsakis, E. 2002. **Structured information retrieval in XML documents**. Proceedings of the 2002 ACM Symposium on Applied Computing (SAC'02). pp. 663-667.

Brito, T., Elias G. 2015. **BranchGuide: an indexing technique for efficient, lossless processing of branching path queries**. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (SAC '15).

McCandless, M., Hatcher, E., and Gospodnetic, O. 2010. **Lucene in Action, Second Edition: Covers Apache Lucene 3.0**. Manning Publications Co., Greenwich, CT, USA.

## **SOBRE O ORGANIZADOR**

**ERNANE ROSA MARTINS** - Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas, possui Pós-Graduação em Tecnologia em Gestão da Informação, Graduação em Ciência da Computação e Graduação em Sistemas de Informação. Professor de Informática no Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia), ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE), certificado pelo IFG no CNPq.

Agência Brasileira do ISBN  
ISBN 978-85-7247-339-2

