

## Desenvolvimento de um Sistema para Gestão de Propriedade Intelectual para Instituições de Ciência e Tecnologia

Arthur L. Teles, Eduardo C. de Souza, Eloy R. P. Maciel, Isaac P. da Silva, Izabela A. M. de Andrade, Lara G. Santos, Mateus L. Sobreira, Mayara C. Mendes, Marcela P. Silvério, Tássyla L. Lima, André R. da Cruz, Daniel H. Dalip

<sup>1</sup>Departamento de Computação  
Centro Federal de Educação Tecnológica de Minas Gerais

{arthurleaoteles, eduardbh2014, eloymaciel06}@gmail.com

{isaacportela2003, contatobelaaa, laragamasantos}@gmail.com

{may.dcmendes, marcelapsilverio, tassyla.tll}@gmail.com

{mateuslealsobreira}@hotmail.com

{dacruz, hasan}@cefetmg.br

**Abstract.** *This paper describes a system focused on intellectual property management, capable of extracting and organizing data published in the journal of the Instituto Nacional de Propriedade Industrial regarding an Intellectual Property (IP). In order that an institution is aware of the information concerning their IPs, such as patents and software registration, it is necessary that the journal's publications are monitored on a weekly basis, which allows for human error and demands considerable time. Therefore, the system presents a solution while extracting and interpreting semi-structured data, performing a deduplication process in different occurrences of an institution name and displaying the information to the user in an organized manner.*

**Resumo.** *Este artigo apresenta um sistema para gestão de propriedade intelectual, capaz de realizar a extração e organização dos dados publicados na revista do Instituto Nacional de Propriedade Industrial acerca de uma propriedade intelectual (PI). Para uma instituição estar a par das informações sobre suas PIs, tais como patentes e registros de software, é necessário que sejam acompanhadas semanalmente as publicações realizadas na revista, abrindo margem para erro humano e podendo demandar tempo considerável. Assim, o sistema propõe uma solução ao extrair e interpretar dados semiestruturados, realizar um processo de deduplicação em diferentes ocorrências do nome de uma instituição e exibir as informações ao usuário de maneira organizada.*

### 1. Introdução

O Instituto Nacional de Propriedade Industrial (INPI) é o órgão responsável pela proteção intelectual de patentes, registros de *software*, marcas, desenho industrial e informações geográficas. Além disso, o INPI apresenta semanalmente uma revista<sup>1</sup> com demandas que

---

<sup>1</sup><http://revistas.inpi.gov.br>

os titulares das propriedades intelectuais (PI) devem atender, denominadas exigências. Com isso, as Instituições de Ciência e Tecnologia (ICTs) devem acompanhar semanalmente essa revista para atender tais exigências, bem como realizar pagamentos ao INPI para mantê-las protegidas e gerenciar os dados das suas PIs. Mesmo com a possibilidade de acesso aos arquivos dessa revista por meios computacionais, algumas ICTs públicas, por dificuldade de contratação de sistemas de informação, acessam-na manualmente.

A forma na qual os dados são apresentados pelo INPI possui limitações e não é considerada a mais adequada em alguns aspectos. Por exemplo, não é simples verificar quais são os próximos pagamentos a serem realizados por PI, visualizar quais temas a instituição mais produz e verificar quais são as próximas pendências a serem atendidas. Outrossim, é comum ocorrer erros humanos tanto no cadastro da informação, quanto na checagem de demandas frente ao INPI.

Este artigo apresenta um sistema para gestão denominado SIGPPI<sup>2</sup> (Sistema Integrado de Gestão de Propriedade Intelectual), desenvolvido com o framework Python Django e o SGBD MariaDB. O objetivo deste é gerenciar as informações acerca da PI desenvolvida por uma ICT e facilitar a análise e busca de tecnologias. O sistema foi implementado a partir da automática coleta, limpeza e organização dos dados provenientes do INPI [Nelli 2018]. Tendo em vista que os eventos da revista são exibidos como dados semiestruturados, em formato XML, é preciso realizar um pré-processamento para torná-los estruturados em um banco de dados relacional. Além disso, cada instituição é referenciada apenas pelo seu nome, podendo haver assim necessidade de deduplicação.

Outros trabalhos relacionados à gestão de propriedade intelectual também podem ser encontrados, como *Symphony*<sup>3</sup> ou SGPI PRO [dos Santos Passos Brito 2020]. Entretanto, o SIGPI se destaca na integração com o INPI e na filtragem e organização dos dados de exigências e pagamentos coletados a partir das notificações da revista.

A Seção 2 apresenta a descrição do sistema e delinea detalhes sobre os módulos. A Seção 3 apresenta a forma como é executada a extração e integração dos dados, o tratamento de deduplicação e a configuração. A Seção 4 conclui o artigo.

## 2. Módulos do Sistema

Uma ICT deve estar a par de cada despacho publicado na revista do INPI, assim como seus pagamentos e a situação de cada PI. Para isso, o presente sistema permite a visualização de despachos e o cadastro de eventos. A Figura 1 apresenta um exemplo de despacho, que consiste em demandas que o INPI solicita que o inventor/titular da PI cumpra, além de ocorrências diversas. Nesse caso, foi informado que a PI foi anulada por falta de cumprimento de uma exigência formal. Dessa forma, o sistema atualiza automaticamente a situação da PI e/ou emite lembrete considerando tais despachos. Em adição, o sistema é responsável pelo controle de pagamento de cada PI, considerando sua data de depósito.

O sistema pode ser representado em 4 módulos, explicados a seguir: (1) gerenciamento de dados de Propriedade Intelectual; (2) pagamentos de uma PI; (3) coleta da revista; (4) cumprimento de exigências. O modelo de armazenamento utilizado foi um banco de dados relacional [Silberschatz et al. 2011].

---

<sup>2</sup><https://youtu.be/QPjkEbbLEhc>

<sup>3</sup><https://www.getapp.com.br/software/2056974/symphony-1>

(21) BR 10 2023 004469-7	Código 15.21 - Numeração Anulada
	(22) 10/03/2023
	(71) INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ (BR/CE)
	Pedido com Numeração Anulada tendo em vista falta de cumprimento de exigência formal

**Figura 1. Exemplo de despacho**

O módulo de gerenciamento de dados de Propriedade Intelectual armazena dados da PI, tais como título, código, titulares e inventores. O módulo de pagamento é responsável por gerenciar os prazos, situação, data do pagamento. O módulo de coleta das revistas do INPI é responsável por extrair os despachos acerca das PI, assim como atualizar a sua situação. Visando identificar se a PI pertence à instituição, é realizado um processo de deduplicação. Os despachos coletados podem possuir exigências gerenciadas pelo módulo de "cumprimento de exigências". Há um módulo para gerenciar as exigências, presentes nas revistas coletadas, no qual toda PI está sujeita a responder.

### 3. Desenvolvimento do Sistema

Esta seção apresenta como foram realizados os aspectos considerados mais importantes para o desenvolvimento do sistema. Na Seção 3.1, apresentamos como foi realizada a extração dos dados da revista do INPI para, assim, salvar os dados sobre a PI e seus despachos em um banco de dados relacional. A Seção 3.2 apresenta como foi realizada a deduplicação do nome da ICT. Finalmente, a Seção 3.3 apresenta como foram configurados os dados de exigências, pagamentos e a situação da PI de tal forma a deixar o sistema flexível a modificações futuras.

#### 3.1. Extração e Integração dos Dados

Conforme apresentado na Figura 2, os despachos do INPI, juntamente com os dados de cada PI, são obtidos no formato XML pela revista do INPI. Cada XML, dependendo do tipo de PI, pode ter uma representação diferente. Dessa forma, seus dados são extraídos e transformados em um JSON, Figura 3, que, por sua vez, foi armazenado em um banco de dados relacional, conforme estrutura apresentada na Figura 4.

Nesse diagrama, pode-se observar que toda Propriedade Intelectual possui seu responsável, associado pela tabela Pessoa, e seus participantes. Uma Pessoa pode ser física ou jurídica, pessoas jurídicas podem estar associadas à ICT responsável - armazenado em InstituicoesNIT. Os participantes podem ser titulares ou inventores, identificados pelo TipoParticipacao. Ademais, a situação da PI é definida por meio da tabela StatusPropriedadeIntelectual e sua tabela de histórico. Uma PI pode possuir diversos eventos cadastrados pelo usuário ou os próprios despachos da revista do INPI. Cada evento tem um tipo que indica se é cadastrado pelo usuário ou um despacho. Caso seja um despacho, são armazenados o código e o título. As informações relacionadas a pagamentos são obtidas por meio da data de depósito de uma PI e gerenciados manualmente pela ICT responsável.

#### 3.2. Deduplicação de dados

Tendo em vista que as publicações da revista do INPI apresentam pouca padronização, é possível que uma mesma ICT seja representada de maneiras diferentes, ou seja, em cada despacho, o nome da ICT tenha sido cadastrado de diferentes formas. Desse modo, o

```

<revista numero="1234" dataPublicacao="29/05/2023" diretorio="Patente">
  <despacho>
    <codigo>123</codigo>
    <titulo>Titulo Despacho</titulo>
    <processo-tipo>
      <titulo>Titulo PI</titulo>
      <numero>BR 32023-2 3</numero>
      <data-deposito>20/05/2023</data-deposito>
      <titular-lista>
        <titular>
          <nome-completo>titular_nit</nome-completo>
          <endereco>
            <pais><sigla>BR</sigla></pais>
            <uf>MG</uf>
          </endereco>
        </titular>
      </titular-lista>
    </processo-tipo>
    <comentario>Um exemplo de comentário</comentario>
    <comentario>Podemos ter inumeros</comentario>
  </despacho>
</revista>
  
```

```

<revista numero="1234" data="29/05/2023">
  <processo numero="BR 32023-2 3" data-deposito="20/05/2023">
    <despachos>
      <despacho codigo="123" nome="Titulo Despacho">
        <texto-complementar>Um exemplo de comentário; Podemos ter inumeros
        </texto-complementar>
      </despacho>
    </despachos>
    <titulares>
      <titular nome-razao-social="titular_nit" pais="BR" uf="MG"/>
    </titulares>
    <marca>
      <nome>Titulo PI</nome>
    </marca>
  </processo>
</revista>
  
```

(a) Patente XML

(b) Marca XML

Figura 2. XMLs Possíveis

```

{'despacho': [{'titulo': 'Titulo Despacho', 'codigo': '123',
               'comentario': 'Um exemplo de comentário; Podemos ter inumeros'}],
 'num_revista': '1234',
 'numero': 'BR 32023-2 3',
 'nome': 'Titulo PI',
 'pertence-filtro': True,
 'data_publicacao': '29/05/2023',
 'data-deposito': '20/05/2023',
 'titulares': [{'nome': 'titular_nit', 'pais_sigla': 'BR', 'uf': 'MG'}],
 # Campos não presentes em uma patente ou em uma marca
 }
  
```

Figura 3. Dicionário python derivado dos XML

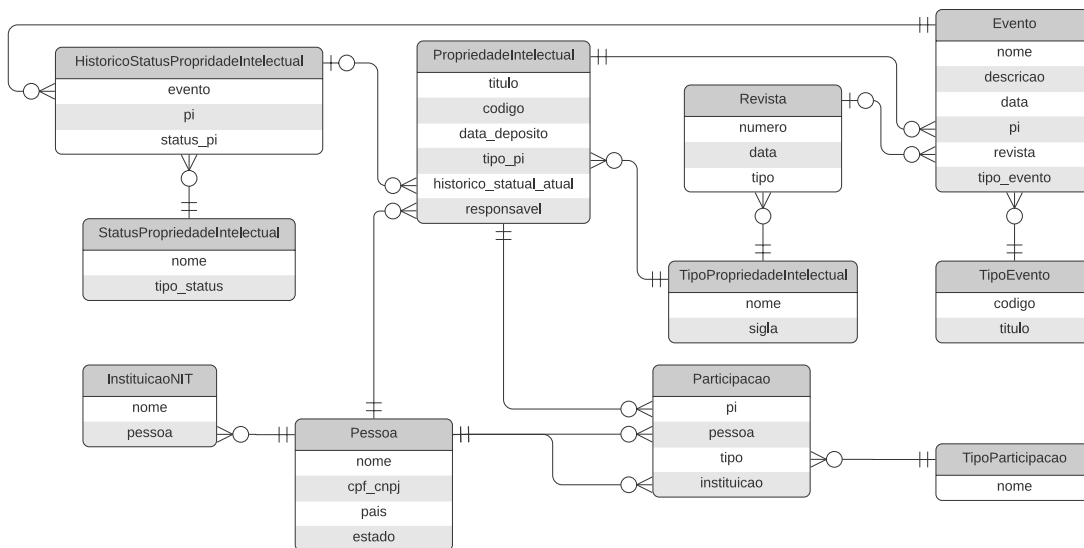


Figura 4. Diagrama lógico do modelo de banco de dados simplificado do sistema

**Tabela 1. Exemplos de regras para deduplicação de uma instituição**

Instituição	Inicia com	Termina com	Contém
CEFET-MG	CEFET	MG	
CEFET-MG	Centro Federal de Educação	Minas Gerais	
CEFET-MG	Centro Fed	Minas Gerais	Tec

o sistema irá realizar deduplicação com o objetivo associar a mesma ICT às suas diferentes representações. Conforme discutido em [Christen 2012], a remoção de duplicatas garante a limpeza dos dados coletados e a extração precisa das informações desejadas. Ademais, existem ferramentas implementadas disponíveis para utilização, como a biblioteca *Dedupe*<sup>4</sup>, que primeiramente divide os dados em grupos (*blocking*) e, em seguida, realiza comparações e identifica as duplicatas dentro de cada grupo por meio de aprendizado ativo. No que tange aos dados inseridos no sistema desenvolvido, a título de exemplo, o Centro Federal de Educação Tecnológica de Minas Gerais pode ser identificado por esse mesmo nome, pela abreviação “CEFET-MG” ou por uma forma alternativa “Centro Fed de Educ Tec de Minas Gerais”.

Assim, observa-se a necessidade de tratar e interpretar esses nomes como um único, considerando que, quando o usuário acessar a plataforma, é fundamental serem exibidos todos os eventos referentes à ICT em que está inserido. Por conseguinte, para este sistema, foi implementado um processo de deduplicação baseado em regras, em que todas as possíveis ocorrências de um nome devem ser cadastradas no sistema, conforme ilustrado na Tabela 1. Essas regras consideram como o termo se inicia, como ele termina e o que está contido em seu meio, o que possibilita que todas as referências a uma instituição sejam englobadas a partir das principais palavras-chave. Essa configuração é armazenada no banco de dados do sistema. A princípio, devido ao custo computacional, foi realizado a deduplicação apenas da ICT na qual estamos extraíndo os despachos. Em trabalhos futuros, será utilizada a ferramenta *Dedupe* e a deduplicação de instituições cotitulares e a desambiguação dos nomes de inventores de uma PI.

### 3.3. Configuração do sistema

Visando um sistema com configurações flexíveis, foram desenvolvidas tabelas responsáveis por configurar os pagamentos, cobranças, exigências e *status* de uma PI.

As configurações dos pagamentos possui informações relativas à periodicidade, repetições da cobrança e datas de referência para prazos de pagamento de uma PI. Após configurada, uma cobrança pode ser associada a uma PI, processo semelhante é feito com as exigências obtidas por meio de um despacho. Um despacho é associado a uma exigência por meio de uma tabela, com essa tabela é possível, para o usuário, configurar de maneira prática a relação entre determinado despacho e uma exigência associada ao código desse despacho. No entanto, é importante destacar que as informações referentes às exigências têm a finalidade de auxiliar no gerenciamento feito pela ICT. A Tabela 2 representa exemplos de exigências e seus respectivos prazos, considerando que pode ser concedido um prazo adicional para o seu cumprimento.

<sup>4</sup><https://docs.dedupe.io/en/latest/how-it-works/How-it-works.html>

**Tabela 2. Exemplo de configuração de exigências**

Exigência	Prazo (dias)	Prazo adicional (dias)
9.2 - Indeferimento	60	
404 - Arquivamento de ofício de registro	60	
029 - Deferimento de pedido de registro	60	30

A situação (*status*) de uma PI, assim como a exigência, é obtida por meio de um despacho do INPI. Para atualizar *status* de uma PI, criou-se uma tabela que relaciona o tipo de despacho e quais alterações esse despacho realiza, podendo ele adicionar, restaurar ou anular um *status*. Dessa forma, permitimos que o usuário configure a atualização gerada na ocorrência de um determinado despacho. A Tabela 3 apresenta alguns tipos de despachos e as atualizações de *status* associadas a eles.

**Tabela 3. Exemplos da associação de despachos e atualizações de *status***

Tipo do despacho	Tipo da atualização	Novo <i>status</i>
8.6 - Arquivamento	Adição	Arquivada
4.3 - Desarquivamento	Restaurar	<i>Status</i> anterior
11.14 - Publicação de Arquivamento Anulada	Anulação	Arquivada

#### 4. Conclusões

Este artigo descreveu a criação de um sistema desenvolvido para facilitar a gestão de propriedade intelectual, focando na organização e análise dos dados publicados na revista do INPI. O objetivo do sistema é atender às necessidades das Instituições de Ciência e Tecnologia (ICTs) ao acompanhar informações relacionadas às suas PIs. Tal sistema realiza a extração e interpretação dos dados semiestruturados presentes na revista, com ênfase na descrição do processo de extração elucidado neste trabalho. Além disso, o sistema propõe uma abordagem para lidar com a deduplicação de dados, sendo esse tópico, juntamente com a extração, o foco central do artigo. A importância desse sistema é evidenciada pelo contexto atual, em que a gestão eficaz da PI se torna cada vez mais essencial para impulsionar a inovação e o desenvolvimento tecnológico. Como trabalhos futuros, planejamos aperfeiçoar a deduplicação e ambiguidades de dados para possibilitar a identificação das instituições parceiras e inventores sem eventuais duplicidades. Além disso, fornecer os relatórios de pagamentos e de produção da ICT.

#### Referências

- Christen, P. (2012). A survey of indexing techniques for scalable record linkage and deduplication. *IEEE Transactions on Knowledge and Data Engineering*, 24(9).
- dos Santos Passos Brito, C. V. (2020). Desenvolvimento, implantação e análise do sistema “sgpi pro” no gerenciamento da propriedade intelectual no NIT da UNIVASF. *Universidade Federal do Vale do São Francisco*.
- Nelli, F. (2018). *Python data analytics with Pandas, NumPy, and Matplotlib*. Apress, 2nd edition.
- Silberschatz, A., Korth, H. F., and Sudarshan, S. (2011). *Database system concepts*. McGraw-Hill Education, 6th edition.