

Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Ciência da Computação

Bruno Boaventura Scholl

**Desenvolvimento de um sistema de informação para
coleta e análise de relatórios de produtividade científica**

São Paulo, 5 de dezembro de 2019

Durante a execução deste trabalho o aluno recebeu uma bolsa da Fundação Butantan

Desenvolvimento de um sistema de informação para coleta e análise de relatórios de produtividade científica

Monografia final da disciplina
MAC0499 – Trabalho de Formatura Supervisionado

Supervisor: Dr. Marcelo da Silva Reis
(Laboratório de Ciclo Celular, Instituto Butantan)

São Paulo, 5 de dezembro de 2019

Resumo

Avaliações de produtividade são parte essencial do monitoramento e controle da produção do conhecimento científico. O Centro de Desenvolvimento Científico (CDC) do Instituto Butantan elabora anualmente um detalhado relatório de produtividade, baseado em informações fornecidas pelos pesquisadores dos diversos laboratórios do Butantan. Todavia, os processos atuais de produção desse relatório, a despeito de avanços recentes, demandam muito tempo e trabalho tanto por parte dos pesquisadores, quanto do comitê nomeado pelo CDC para compilar e analisar esses dados. Dessa forma, neste projeto resolvemos essa questão através do desenho e implementação de um banco de dados relacional, populando-o com dados do relatório atual e analisando esses dados através de consultas relacionais. Além disso, também elaboramos um relatório de desempenho individual dos pesquisadores. Por fim, desenvolvemos uma interface web para que as próximas edições do relatório do CDC sejam coletadas de forma rápida e automatizada. O código-fonte está disponível de forma livre e gratuita, sob a licença GPLv3, o que permite que ele também seja empregado em outros Institutos de Pesquisa do Estado de São Paulo.

Palavras-chave: Cientometria, Bibliometria, Sistema de informação, Banco de dados.

Abstract

Productivity evaluations are an essential part of the assessment of scientific knowledge production. The Center of Scientific Development (CDC) of the Butantan Institute elaborates a detailed annual production report based on data provided by all researchers from the Institute. However, the current method of production of this report, despite recent advancements, is time-consuming and laborious for both the researchers and the committee nominated to compile and analyze those data. In this project, we solve this matter by designing and developing a relational database to store and query data, populating it with the current report data and assessing the yearly production report through database queries. We also compiled an individual performance report for all researchers, based on productivity data from the database. Finally, we developed a web interface for data submission, which will make future submissions fast and automated. The source code of the project is available in the form of free, open-source software under the GPLv3 license, which will enable other research institutions from São Paulo State to use that system in their own internal scientific production evaluations.

Keywords: Scientometrics, Bibliometrics, Information System, Database.

Sumário

Lista de Abreviaturas	vii
1 Introdução	1
1.1 Objetivos	3
1.1.1 Objetivos Específicos	3
1.2 Organização	3
2 Conceitos Fundamentais	5
2.1 Termos do Âmbito Científico	5
2.1.1 <i>Paper</i>	5
2.1.2 Revisão por Pares	6
2.1.3 DOI	6
2.1.4 Número de Citações	6
2.1.5 Índice H	7
2.1.6 Fator de Impacto (IF)	7
2.1.7 Qualis	9
2.1.8 Conceito CAPES	9
2.1.9 Credenciamento em Programa de Pós-Graduação	9
2.1.10 Agência de Fomento	10
2.1.11 Bolsa de Produtividade do CNPq	10
2.1.12 Nível PQ	10
2.1.13 ORCID	11
2.1.14 <i>ResearcherID</i>	11
2.1.15 Google Scholar	11
2.2 Conceitos técnicos e tecnologias	12
2.2.1 MVC	12
2.2.2 <i>Ruby on Rails</i>	12
2.2.3 Vue	13
3 Metodologia e Desenvolvimento	15
3.1 Regras de Negócio	15
3.2 Primeiro Passo - Banco de Dados	17

3.3	Módulo de Inserção de Dados	18
4	Caso de Uso	21
4.1	Cadastro	21
4.2	Login	23
4.3	Preenchimento do Formulário	23
4.3.1	Elementos de Inserção	24
4.3.2	Submissão do Formulário	24
4.3.3	Avalização dos Formulários	25
5	Discussão e Conclusões	29
A	Modelo Entidade Relacionamento	31
	Referências Bibliográficas	37

Lista de Abreviaturas

MVC	<i>Model View Controller</i>
DOI	<i>Digital Object Identifier</i>
CDC	Centro de Desenvolvimento Científico
MER	Modelo Entidade-Relacionamento
ORCID	<i>Open Researcher and Contributor ID</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
GPL	<i>General Public License</i>
REST	<i>Representational State Transfer</i>
IF	<i>Impact Factor</i>
SQL	<i>Structured Querying Language</i>
BD	Banco de Dados

Capítulo 1

Introdução

Avaliações de produtividade são mecanismos através dos quais a comunidade acadêmica, de forma periódica, monitora e controla a produção do conhecimento científico. Essas avaliações são utilizadas pelas lideranças institucionais para diversos fins, tais como a concessão de bolsas e auxílios e a contratação e/ou promoção de pessoal (Mattedi e Spiess, 2017). No caso particular do Instituto Butantan, uma das mais importantes avaliações de produtividade é a organizada anualmente pelo Centro de Desenvolvimento Científico (CDC). Os resultados obtidos nessa avaliação são enviados à Secretaria da Saúde do Estado de São Paulo, como forma de prestação de contas à sociedade paulista, e também utilizados internamente para definição de políticas científicas e distribuição de recursos entre os laboratórios do Butantan.

Tradicionalmente o relatório anual do CDC é elaborado em duas etapas: na primeira delas, pesquisadores de cada laboratório preenchem dois formulários. O primeiro deles é um documento de texto do Word no qual são inseridas informações tais como lista de publicações de artigos e de livros ao longo do ano, orientações concluídas e em andamento, participação em congressos, aulas ministradas, etc. Já o segundo, implementado em uma planilha Excel, serve especificamente para colocar dados financeiros de auxílios e de bolsas de alunos. A primeira etapa é concluída quando a secretaria de cada laboratório compila os dados de todos os pesquisadores em um único documento Word e uma única planilha Excel; ou seja, são produzidos dois arquivos que são referentes ao laboratório como um todo. Já na segunda etapa, também conhecida como *consolidação do relatório*, um comitê nomeado pelo CDC recebe todos os relatórios de laboratório (atualmente 29 deles, de quatro divisões diferentes, incluindo o próprio CDC) e gera estatísticas holísticas sobre eles: por exemplo, número de pesquisadores do Butantan (distinguindo entre estatutários e celetistas contratados pela Fundação Butantan), número de alunos orientados pelos pesquisadores (separando entre bolsistas e não bolsistas), número de trabalhos publicados (distinguindo entre artigos, livros e capítulos de livros), total de valores arrecadados em auxílios e bolsas (separando por valores em real e dólar), etc. Como todo esse processo é feito de forma manual, a consolidação do relatório é um processo muito trabalhoso, demorado e sujeito a erros na compilação das informações.

Recentemente, o comitê atualmente designado pelo CDC realizou algumas melhorias no processo de elaboração da edição de 2018 do relatório. Uma delas foi a simplificação da documentação, que passou a ser composta por apenas uma planilha Excel para todas as informações, facilitando assim seu preenchimento por parte do pesquisador; tal simplificação também eliminou redundâncias que existiam entre os dois arquivos da documentação original. Outro avanço importante foi na consolidação do relatório: como o novo formulário foi criado com a preocupação em manter um posicionamento sistemático das informações (e.g., nome de aluno bolsista só pode ser inserido na coluna B, entre as linhas 30 e 70), para realizar a consolidação bastou ao comitê inserir num mesmo arquivo todos formulários de laboratórios, um por aba; após isso, foram incluídas abas que referenciam as informações de todas as abas dos laboratórios para gerar as estatísticas necessárias; com o posicionamento sistemático das informações, as fórmulas nas abas de estatísticas buscam as informações necessárias nos locais predeterminados.

Todavia, embora o relatório de 2018 tenha tido os avanços descritos acima, ainda existem problemas. Ainda não é possível fazer de forma imediata e abrangente uma comparação desses resultados com os obtidos nos anos anteriores: seria preciso fazer antes uma curadoria manual da documentação das edições passadas, que está sujeita a todos os problemas mencionados anteriormente, e que portanto precisaria ser feita com tempo e cuidado. Além disso, planilhas tais como as do Excel não são o meio mais adequado para armazenar e analisar o grande volume de informações gerado pela série histórica de relatórios, nem facilita modificações, uma vez que a alteração da disposição das informações requereria o ajuste de todas as fórmulas. Uma alternativa escalável e, portanto, mais apropriada para esses fins, seria um banco de dados relacional, que é projetado para lidar com grandes volumes de dados e permite grande flexibilidade em relação a alterações futuras. A utilização de um banco de dados traz também a vantagem de que as informações podem ser buscadas de forma personalizada, com diversos filtros e arranjos dos dados, o que abre portas para novas análises e relatórios. Por fim, embora o formulário único tenha sido um avanço importante em relação ao relatório de anos anteriores, uma interface web cobrindo as mesmas informações seria uma alternativa que facilitaria ainda mais o preenchimento por parte do pesquisador, principalmente pela característica dinâmica de uma página web, que permitiria funcionalidades como sugestão de inserções (e.g. o nome de um laboratório) e validação de informações (e.g. busca de informações de um artigo a partir de seu DOI); ademais, se ao mesmo tempo essa interface alimentar um banco de dados relacional, o processo de consolidação do relatório anual do CDC tornar-se-ia totalmente automatizado, aliviando assim o comitê desse relatório e abrindo portas a novas possibilidades trazidas pela velocidade nesse processo.

1.1 Objetivos

Este trabalho teve como fim a melhoria no fluxo de submissão, armazenamento e análise dos dados para o relatório anual de produtividade científica do Instituto Butantan. Para tanto, **implementamos um sistema para coleta e submissão de dados de produção científica** para o relatório de produtividade científica do Instituto Butantan.

Dividimos este objetivo principal em passos incrementais de desenho e produção de software, formando, assim, três objetivos específicos:

1.1.1 Objetivos Específicos

1. Desenho e implementação de um banco de dados para armazenamento e busca dos dados do formulário de produtividade:

Este objetivo teve como resultado a especificação e implementação de um banco de dados capaz de armazenar e lidar com dados presentes nos formulários de 2018 já preenchidos.

2. Elaboração do relatório de produtividade referente a 2018 utilizando o banco de dados:

Este objetivo serviu para validarmos tanto o desenho e a implementação do banco de dados, quanto para avaliarmos as vantagens trazidas pela utilização de um banco de dados relacional nesse cenário.

3. Implementação de um sistema com interface gráfica, com base no projeto feito do banco de dados, para inserção de dados para armazenamento:

Por fim, este objetivo visou melhorar o banco de dados inicialmente projetado, adicionando um meio de interação que não requer experiência com programação e que pode ser utilizado diretamente pelos pesquisadores para inserção de dados e submissão dos formulários ao CDC.

1.2 Organização

Este trabalho está disposto em cinco capítulos. Neste primeiro capítulo, introduzimos o problema e as soluções com as quais o atacamos. No capítulo 2, apresentamos conceitos do âmbito da pesquisa científica em São Paulo e no Brasil, presentes no formulário, que geralmente não são conhecidos por aqueles sem contato com esse meio, como estudantes da graduação. No capítulo 3, relatamos o processo e metodologia utilizados na implementação de cada um dos objetivos específicos. No capítulo 4, mostramos um tutorial do sistema, no qual é feita a explicação de uso de alguns elementos presentes durante o preenchimento do formulário e o uso do sistema. Por fim, no capítulo 5, discutimos os resultados e também planos futuros para esta linha de pesquisa.

Capítulo 2

Conceitos Fundamentais

Neste capítulo apresentamos conceitos fundamentais para a leitura do restante da monografia. O capítulo está organizado em duas seções: na primeira delas (seção 2.1), descrevemos termos e jargões comuns no meio da pesquisa científica que, geralmente, são desconhecidos por aqueles que não estão inseridos nesse meio; tais termos e jargões aparecem ao longo do formulário do relatório de produção científica mencionado no capítulo 1. Já na seção 2.2, apresentamos conceitos técnicos e tecnologias utilizadas durante o desenho e desenvolvimento do sistema.

2.1 Termos do Âmbito Científico

2.1.1 *Paper*

Um *paper* é um artigo científico publicado em um periódico científico (*journal*, em inglês), geralmente com Revisão por Pares (subseção 2.1.2), com o objetivo de divulgar os avanços e resultados de pesquisas científicas. Artigos publicados em periódicos que contam com revisão por pares trazem consigo certo grau de confiabilidade, tendo em vista que os resultados e métodos demonstrados foram inspecionados por outros pesquisadores da mesma área. Além da revisão por pares, há outras métricas e métodos para avaliar e legitimar, não apenas os *papers* publicados em periódicos, mas também os periódicos propriamente ditos. A avaliação dos periódicos incentiva que estes mantenham altos os seus padrões de qualidade e os seus critérios de aceitação e publicação.

Papers são identificados pelo conjunto formado pelo seu título e seus autores. Para evitar possíveis ambiguidades entre nomes de artigos com títulos semelhantes e também permitir a indexação, busca e referência de artigos individuais, os periódicos atribuem um número DOI (subseção 2.1.3), que identifica unicamente cada artigo e pode ser acessado por um sistema centralizado na *web*.

2.1.2 Revisão por Pares

A revisão por pares é o processo de aferição de qualidade, veracidade e correção de artigos enviados a periódicos para publicação que traz confiabilidade à produção científica disseminada. De acordo com [Bornmann \(2011\)](#), essa verificação tem como objetivo julgar apenas a qualidade do trabalho produzido, sem levar em consideração características do autor, nem viés do avaliador. Para tanto, ela é, geralmente realizada por pesquisadores escolhidos de forma anônima, ou seja, aquele que submete o artigo não sabe quem o está avaliando, pelo menos até o artigo ser aceito. Durante o período de avaliação, os autores podem submeter o mesmo artigo com alterações apontadas pelo avaliador, com limite de submissões definido pelo periódico, até que seja aceito. Após a aceitação há alguns periódicos que divulgam quem foram os avaliadores e o histórico do artigo submetido. No entanto, isso pode variar de acordo com o periódico.

2.1.3 DOI

O Identificador Digital de Objeto, do inglês *Digital Object Identifier*, é um sistema de identificação persistente de objetos físicos, virtuais ou abstratos em redes (i.e. internet). O DOI, como especificado por [Paskin \(2010\)](#), garante a alocação de um número único para cada objeto identificado e também a capacidade de resolver seus identificadores para localizações na internet (URLs). A resolução a partir de um dado DOI, para localizações digitais, pode ser feita para endereços diferentes de acordo com requisições diferentes (veja figura 2.1), por exemplo, quando um navegador pede para que seja resolvido para uma URL, pode ser retornado o endereço onde o artigo está publicado; por outro lado, caso o navegador faça uma requisição para um outro formato, a nova informação requerida é retornada, caso esteja disponível.

O DOI é comumente utilizado no meio científico para identificar artigos e livros publicados, de forma que basta o número DOI para que alguém encontre o corpo do artigo e ainda metadados, como a citação já formatada do artigo referido.

2.1.4 Número de Citações

Assim como o número de citações em artigos é utilizado em métricas de popularidade e impacto de artigos e periódicos, ele também pode ser estendido a autores. Nesse caso, ele é utilizado na forma de número de citações totais, ou seja, o número total de vezes em que *papers* desse autor foram citados em outros artigos.

A informação que essa medida traz é da dimensão da contribuição e relevância dos trabalhos desse pesquisador para outras pesquisas e projetos; porém, essa métrica não deve ser a única utilizada para comparar a produtividade ou o impacto de autores, pois, por ser uma medida absoluta, diferenças no volume de publicações e na quantidade de citações por publicação trazem significados diferentes a um mesmo valor da métrica. Uma outra métrica

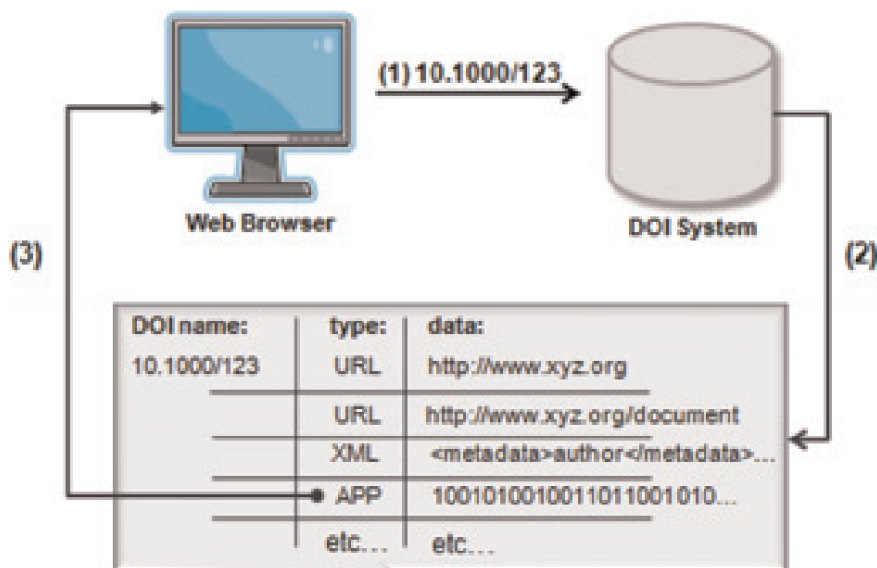


Figura 2.1: Ilustração de uma requisição feita pelo navegador (1) para a resolução do DOI *doi:10.1000/123* no formato de aplicativo (APP). O sistema DOI, por sua vez, verifica se o formato pedido está disponível (2) e então, em caso positivo, retorna a URL ou os dados associados ao formato (3). Figura extraída de Paskin (2010).

usada no lugar ou em conjunto com o número de citações totais é o índice h (subseção 2.1.5).

2.1.5 Índice H

A medida de desempenho, impacto ou alcance do trabalho de pesquisadores é geralmente utilizada para decisões como distribuição de recursos, contratações e premiações. Listar e classificar autores para tais fins deve ser feito de modo imparcial, objetivo e evitando vies, algo que pode ser atingido utilizando métricas adequadas. Uma medida usualmente utilizada para a comparação de autores baseada em impacto e relevância é o índice h , proposto por Hirsch (2005), uma métrica baseada em citações e número de publicações simultaneamente (figura 2.2).

Um pesquisador é dito com índice h se possui h artigos publicados, com pelo menos h citações cada. Ou seja, mesmo que o pesquisador publique em grande volume, se seus artigos não forem citados em grande quantidade também, seu índice h se manterá proporcional ao número de citações. O oposto também ocorre: caso um pesquisador tenha publicado poucos artigos, mas estes foram citados diversas vezes, seu índice h permanece proporcional ao número de publicações. Dessa forma, um pesquisador que tem um índice h alto não apenas publica bastante, mas também é citado em grande número.

2.1.6 Fator de Impacto (IF)

O Fator de Impacto, muitas vezes referido como IF, do inglês *Impact Factor*, é uma métrica idealizada por Garfield (1955) com o intuito de selecionar periódicos científicos de importância, mas que não necessariamente tivessem muitas citações, pois esses seriam apenas

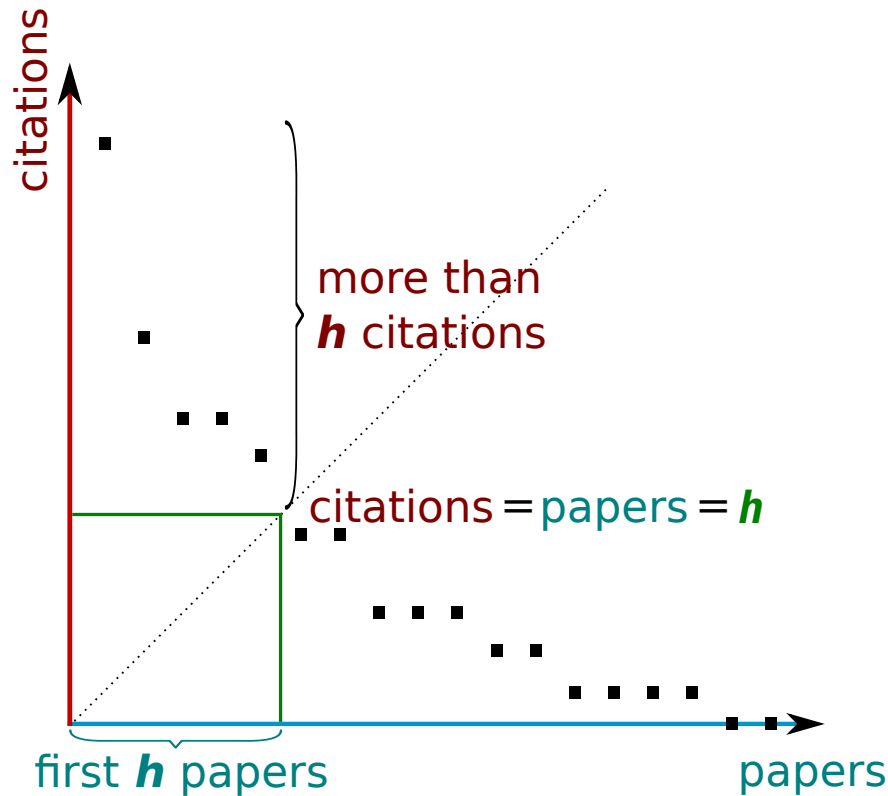


Figura 2.2: Cálculo do índice h . Eixos horizontal e vertical representam, respectivamente, o número de papers e o número de citações. Imagem em domínio público.

periódicos grandes.

O Fator de Impacto, sucintamente descrito em Garfield (1999), é calculado por uma razão que tem como numerador o número de citações do ano atual que citam artigos dos dois últimos anos e como denominador, o número de artigos publicados nesses dois anos. Matematicamente, sejam A_1, A_2, A_3 três anos consecutivos. O numerador N_{cit} é o número de vezes que artigos dos anos A_1 e A_2 são citados em artigos do ano A_3 . O denominador N_{papers} é o número de papers publicados nos anos A_1 e A_2 . Assim, temos o fator de impacto IF como:

$$IF = \frac{N_{cit}}{N_{papers}}$$

A fórmula descrita acima é a mais comumente utilizada no cálculo do IF, porém o número de anos anteriores levados em consideração pode ser alterado para se obter métricas que representam um período maior ou menor de tempo. A diminuição do período refletido pelo IF pode trazer uma pontuação maior para o periódico de uma área que muda rapidamente, enquanto que o aumento desse período, traria maior pontuação para um periódico de uma área cujos conceitos não se alteraram muito com o tempo.

Hoje em dia, o IF ainda é utilizado como métrica relacionada a periódicos, artigos e até autores, porém, devido a controvérsias quanto ao seu uso (Moustafa, 2015), com parte da comunidade acadêmica defendendo o uso de outras métricas.

2.1.7 Qualis

O Qualis Periódicos, ou apenas Qualis, é uma ferramenta de avaliação de periódicos criada pela CAPES (2014), utilizada no cálculo do conceito CAPES. Essa avaliação é feita anualmente por comitês responsáveis que avaliam os periódicos nos quais os artigos foram publicados e, de forma indireta, os artigos recebem a mesma nota do periódico. Os comitês e critérios são divididos em áreas de avaliação e atribuem a cada periódico avaliado uma das classificações: $A1, A2, A3, A4; B1, B2, B3, B4; C$; onde $A1$ é a classificação mais alta e C tem peso zero. Cada periódico pode ser avaliado por mais de um comitê, o que geralmente resulta em mais de uma nota; isso é até esperado, pois um periódico pode ser muito relevante em uma determinada área, porém nem tanto em outras. Quando isso ocorre, é considerada a maior nota.

Em 2018 houve uma mudança na metodologia da avaliação (CAPES, 2018) que especifica, dentre outras mudanças, a implantação de um Qualis referência que combina indicadores bibliométricos num modelo matemático para atribuir uma classificação de referência, que pode ser levemente alterada pelo comitê avaliador. Outra mudança importante, ocorreu no cálculo do qualis periódicos em si e agora utiliza a colocação do periódico ordenada por:

- **número de citações** (subseção 2.1.4) da plataforma Scopus (CiteScore); ou
- **fator de impacto** (subseção 2.1.6) da Web of Science; ou
- **índice h** (subseção 2.1.5) do Google Scholar, considerando um período de 5 anos.

A posição do periódico no *ranking* de uma dessas métricas categoriza o periódico em uma das categorias Qualis, a depender do percentil em que se encontra.

2.1.8 Conceito CAPES

Os programas de pós-graduação do Brasil passam anualmente por uma avaliação feita pela Diretoria de Avaliação da CAPES (PedroII). Essa avaliação utiliza dados de transparência disponibilizados através da plataforma sucupira e é feita por pares de cada área, ou seja, um programa de pós-graduação é avaliado por acadêmicos da mesma área do programa. A nota da avaliação, chamada de conceito CAPES, varia de 1 a 7, de forma que 1 e 2 reprovam o programa, 3 é regular, 4 é bom, 5 é muito bom e é a nota máxima para programas que oferecem apenas mestrado. Notas 6 e 7 são atribuídas para programas que, além de oferecerem o doutorado, também se destacam com excelência científica.

2.1.9 Credenciamento em Programa de Pós-Graduação

Pesquisadores, sejam ou não professores em universidades, interessados em orientar alunos de mestrado e/ou doutorado, podem se cadastrar como orientadores em programas de

pós-graduação dessas universidades, desde que estejam capacitados para tanto. Cada universidade, ou ainda, cada faculdade tem seus critérios que exigem, dentre outras coisas, título mínimo do orientador, produção científica e aulas ministradas previamente (EP-USP, 2004). O credenciamento em programa de pós-graduação permite que o pesquisador cadastrado oriente alunos daquela faculdade ou universidade, até um número máximo de orientandos, definido também pelo programa.

A formação de alunos da pós-graduação traz benefícios não apenas para o aluno e para a instituição que abriga o programa, mas também para o pesquisador, pois a formação de recursos humanos é contabilizada para o desempenho geral do pesquisador.

2.1.10 Agência de Fomento

A pesquisa científica, em grande parte dos projetos, recebe incentivos e verbas para viabilizar suas atividades. A seleção e curadoria de quais projetos são selecionados para receber valores é feita por instituições dedicadas a esse gerenciamento e incentivo à pesquisa. Essas instituições, chamadas Agências de Fomento (à pesquisa), atuam em nível federal ou estadual, redirecionando verba do país ou do estado para projetos de pesquisa. Essa verba é destinada a bolsas, equipamentos e custos necessários para a execução de cada projeto. O método de seleção dos projetos depende da agência; algumas, como a FAPESP, utilizam de um mecanismo semelhante à revisão por pares, mas para a avaliação, individual, de propostas submetidas.

2.1.11 Bolsa de Produtividade do CNPq

Bolsas de produtividade são prêmios concedidos pelo CNPq a pesquisadores por seu desempenho e dedicação e são usualmente vistas como um atestado de excelência dentre os acadêmicos. A bolsa é outorgada para aqueles que se destacam em produtividade científica, formação de pós-graduados, participação em grupos de pesquisa e liderança em sua área. Essas bolsas são divididas em duas categorias: 1 e 2, e em quatro níveis para categoria 1: *A, B, C, D*, dessa forma, as bolsas de produtividade têm as seguintes classificações, de forma ascendente em exigência: *2, 1D, 1C, 1B, 1A*. Existe ainda a categoria Sênior (SR), que é destinada a pesquisadores já aposentados mas que ainda mantêm vínculos com instituições de pesquisa, na forma de colaboração.

2.1.12 Nível PQ

A carreira de Pesquisador Científico, criada em 1975, pela lei complementar N°125 Estado de São Paulo (1975) se refere ao cargo de pesquisador científico em instituições públicas de pesquisa do Estado de São Paulo (e.g. Instituto Butantan). Pesquisadores que seguem essa carreira são divididos em seis classes, comumente chamadas de Nível PQ, que definem atividades desempenhadas, grau de complexidade, capacitação exigida e remuneração. Ao

se referir a essas classes é utilizado o nome *Pesquisador Científico*, seguido dos numerais romanos *I* a *VI*, de acordo com a classe ou a sigla PqC-, novamente acompanhada de um numeral segundo da classe (e.g. Pesquisador Científico I, PqC-II).

2.1.13 ORCID

ORCID é a sigla para *Open Researcher and Contributor ID*, um sistema de identificação única de pesquisadores e acadêmicos que visa combater o problema de nomes parecidos ou iguais no ambiente da pesquisa científica, proposto por HAAK *et al.* (2012). O objetivo deste identificador numérico é que ele seja usado nas publicações para referenciar os contribuidores e preservar a ligação, sem ambiguidades, entre pesquisador e pesquisa. O ORCID está disponível em: <https://orcid.org>.

2.1.14 ResearcherID

O ResearcherID é o identificador único de pesquisador da plataforma *Web of Science* (SIBi-USP). Nessa plataforma estão disponíveis dados sobre citações, publicações e índice h (subseção 2.1.5). Apesar desse serviço oferecer tais informações, quando comparado com outras plataformas, deixa a desejar devido à falta de automação no processo de atualização do perfil do pesquisador cadastrado, requerendo intervenção manual para atualização da lista de publicações a partir da rede *Web of Science* (Marín-Martín *et al.*, 2016). O ResearcherID está disponível em <http://www.researcherid.com>.

2.1.15 Google Scholar

Para que se tenha acesso aos artigos publicados, cada periódico tem seu próprio meio de disseminação, seja ele físico, de forma impressa, ou virtual, disponível na web. Usualmente o acesso a esses itens se dá de forma direta, ou seja, pelo *link* ou algum outro serviço de resolução como o DOI (subseção 2.1.3). Para que esses artigos possam ser encontrados mais facilmente, foram criadas ferramentas de busca de *papers* que varrem as bases de dados dos periódicos e permitem diversas formas de busca (e.g. textual, palavras-chave) a depender da ferramenta.

Em 2004, a Google Inc. lançou o *Google Scholar*, chamado em português de Google Acadêmico (<https://scholar.google.com.br>), uma ferramenta de busca para artigos científicos, teses, livros e até páginas *web* consideradas acadêmicas. Vine (2006) escreveu um artigo sobre o estado e utilidade do Google Scholar na época, pouco tempo após seu lançamento, quando a ferramenta contava apenas com busca textual, rápida e fácil, e ordenação baseada em relevância por um algoritmo não publicado (assim como o buscador Google). Vine aponta também, problemas do buscador que, por ser baseado em *crawling* (acesso automatizado a sites com o intuito de interpretar e buscar dados ou informações), muitas vezes apontava para páginas com apenas a amostra do artigo, principalmente em periódicos sem

acesso livre. Esse problema ainda persiste por causa de periódicos que requerem assinaturas ou pagamentos.

Hoje em dia, o Google Scholar conta também com a funcionalidade de contabilidade e análise de citações e é utilizado em paralelo com outras fontes como métrica de desempenho de pesquisadores. Esse uso em conjunto se dá pela diferença nos métodos de contagem utilizados por outras fontes, que resultam em números diferentes.

2.2 Conceitos técnicos e tecnologias

2.2.1 MVC

Desenvolver aplicações com interfaces visuais interativas apresenta desafios que os paradigmas de programação de aplicações não visuais não são capazes de transpor, ou simplesmente não são uma boa opção. Para atacar essa dificuldade, surgiu o MVC (Krasner *et al.*, 1988), sigla para *Model-View-Controller*, em português: Modelo, Visualização e Controlador, um paradigma de programação focado em interfaces visuais. Este paradigma propõe a separação da aplicação em três partes, ilustradas na figura 2.3, a saber:

- **Modelo:** parte do sistema responsável pela definição, estruturação e representação dos dados guardados em bancos de dados, ou na memória quando em tempo de execução.
- **Visualização:** parte do sistema responsável pela formatação e exibição das informações. Nela ficam definidos o leiaute da aplicação e em quais partes desse leiaute aparecerão quais informações.
- **Controlador:** parte do sistema responsável pela mudança de estado e/ou visualização da aplicação de acordo com a interação do usuário. Em algumas variações do MVC, essa parte é a ligação entre o modelo e a visualização e faz a resolução e filtragem de quais dados são necessários a cada vez que a visualização muda.

Separar o sistema nesses três componentes traz independência para as partes e torna o desenvolvimento mais flexível, de forma que alterar uma delas não necessariamente implica em alteração nas outras.

2.2.2 *Ruby on Rails*

Ruby on Rails (2005), ou apenas Rails, é um arcabouço de desenvolvimento de aplicações *web*, de código aberto, feito na linguagem de programação interpretada *Ruby*, que implementa o MVC (subseção 2.2.1) e oferece estruturas para lidar com bases de dados, sistemas *web* e páginas (interfaces) *web*. Sua maior vantagem é abstrair boa parte das ligações entre subsistemas, ficando a cargo do usuário escrever a lógica de negócio para cada uma das partes do MVC. Para o desenvolvimento dos modelos de banco de dados (*model*) e da

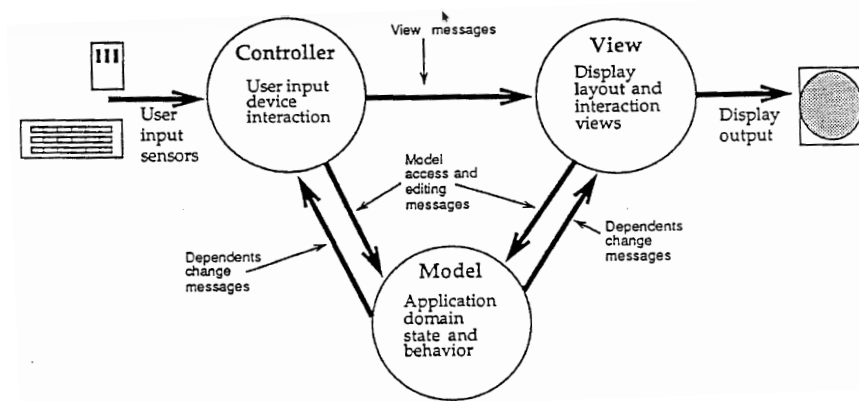


Figura 2.3: Separação das partes em Modelo, Visualização e Controlador. Cada uma das partes é responsável por uma parte do fluxo dos dados, de forma desacoplada. Imagem retirada de Krasner et al. (1988).

lógica de negócio (*controller*) é utilizada a linguagem Ruby, enquanto que para a escrita de páginas web são utilizadas as linguagens padrão: HTML, CSS e JavaScript (EcmaScript).

2.2.3 Vue

Vue.js, também chamado apenas de Vue, é um arcabouço para o desenvolvimento de páginas *web*, de código aberto e escrito em JavaScript. Sua principal vantagem é a divisão das páginas *web* em componentes, que permitem encapsulamento e reutilização de código. Um componente é uma parte do código (e.g. um formulário ou uma tabela), geralmente genérico para facilitar a reutilização, que pode ser inserido em partes da página e é capaz de receber dados ou realizar comportamentos específicos em cada instância. Uma das maiores características dos componentes do Vue é a reatividade, que permite que dados alterados em componentes sejam atualizados em todas as suas referências automaticamente, o que por sua vez facilita o desenvolvimento de interfaces dinâmicas e interativas.

Vue também permite a instalação de *plug-ins*, extensões que podem adicionar novos componentes ou funcionalidades já prontos, produzidos pela comunidade e disponíveis através de repositórios públicos, afim de agilizar e facilitar a criação de projetos que empregam componentes e partes comumente utilizadas pela comunidade que utiliza o arcabouço.

Capítulo 3

Metodologia e Desenvolvimento

3.1 Regras de Negócio

A submissão dos formulários de produção científica do Instituto Butantan é realizada anualmente por todos os 173 pesquisadores do instituto. Estes formulários são utilizados na avaliação de produção científica do instituto, realizada pelo Centro de Desenvolvimento Científico (CDC) com o intuito de gerar dados para tomadas de decisão baseadas no desempenho dos laboratórios do instituto.

O fluxo de coleta dos dados, ilustrado na figura 3.1, começa com o preenchimento dos formulários por parte dos pesquisadores de cada um dos laboratórios do instituto. Em seguida, o formulário preenchido segue para avaliação pelo diretor de cada laboratório, que verifica o preenchimento de cada formulário, requisitando correções caso necessário. Depois que os formulários de um laboratório são aprovados, eles são enviados ao diretor da divisão, por todos os laboratórios pertencentes à divisão. A verificação é mais uma vez repetida pelos diretores de divisão e, então disponibilizados ao CDC para análise.

O formulário de produção científica é constituído de cinco seções que abrangem desde dados pessoais, bolsas até publicações e atividades administrativas. As seções que compõem o formulário são:

- **Pessoal:** Seção na qual são informados dados pessoais do pesquisador, funcionários e alunos, assim como bolsas de produtividade do pesquisador;
- **Produção Científica:** Nesta seção são informadas publicações de artigos e capítulos de livros, juntamente com a informação de tipo de autoria (e.g. primeiro autor, autor correspondente), índices cientométricos (e.g índice h) e participação em congressos;
- **Formação de Recursos Humanos:** Esta seção inclui entradas como teses, dissertações e TCCs supervisionados, supervisões de pós-doutorados, disciplinas ministradas e credenciamentos em programas de pós-graduação;
- **Captação de Recursos:** Nesta seção são apresentados auxílios vigentes e os contratados no ano, e bolsas de alunos;

- **Participação em Atividades:** Esta seção indica as participações do pesquisador em atividades institucionais (e.g. comissões e comitês), culturais (e.g. organização de exposições), inovação ou translacionais, ou ainda, prestação de serviços;

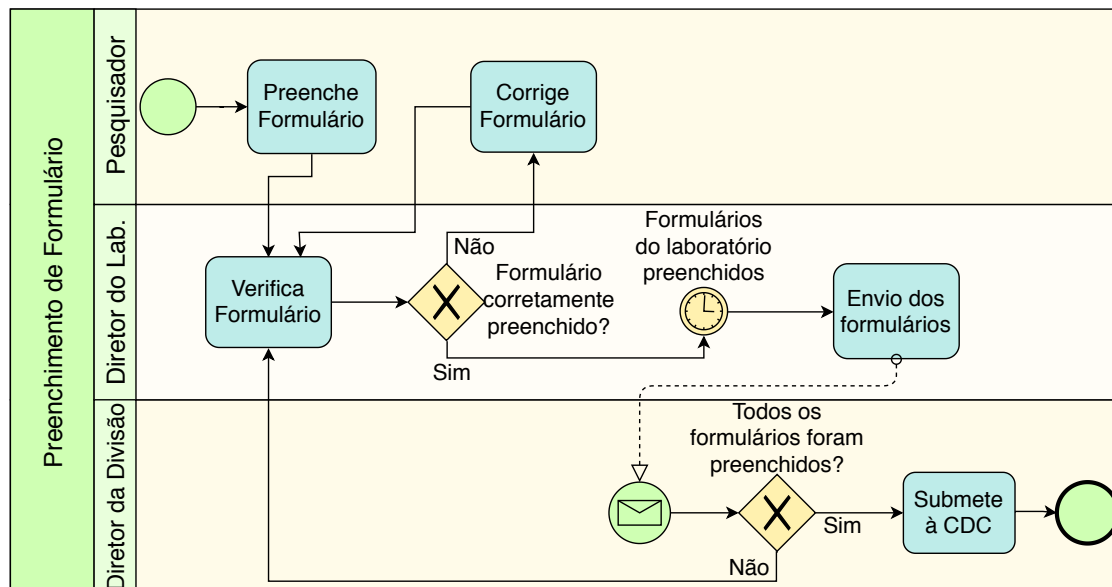


Figura 3.1: Modelo e Notação de Processo de Negócio (BPMN) que ilustra as regras de negócio do fluxo de preenchimento, validação e entrega dos formulários de produção científica do Instituto Butantan. Os formulários são preenchidos pelos pesquisadores, validados pelos diretores dos laboratórios, submetidos aos diretores de divisão de laboratórios e, então, entregues ao Centro de Desenvolvimento Científico (CDC) para sumarização e análise.

O preenchimento e submissão dos formulários era realizado de forma totalmente manual através de documentos de texto (Word) que tinham suas informações extraídas e sumarizadas à mão para formar o relatório. Recentemente esses documentos de texto foram trocados por planilhas (Excel), o que tornou a inserção dos dados mais organizada e estruturada através da formatação mais rígida. Além disso, o Excel permite o uso de fórmulas para facilitar a posterior sumarização dos dados.

A utilização de planilhas, apesar de trazer esses benefícios, não é a solução ideal, pois, para tanto, se vale da forma como o documento está estruturado, o que pode sofrer mudanças de acordo com o número de registros, e outras mudanças que afetam a estrutura da planilha. Outro problema com o uso de planilhas é a falta de integração e relacionamento entre os dados, isto é, a busca por dados não traz outros dados relacionados, nem dá essa opção. Também, a interface de edição de planilhas não traz artifícios que garantam o preenchimento consistente, sem a necessidade de extensiva codificação e personalização. Além disso, a planilha atual não foi bem estruturada, uma vez que guarda dados derivados e não a fonte deles (e.g. número de publicações, sem os dados das publicações propriamente ditos) e possui múltiplas tabelas similares que poderiam ser agrupadas em uma única tabela. Por fim, as planilhas não guardam informações temporais de cada uma das edições do formulário, tendo a necessidade de múltiplas planilhas para guardar essas informações e, novamente,

exigindo uma grande quantidade de código para comunicar e relacionar os dados de uma planilha com a outra.

3.2 Primeiro Passo - Banco de Dados

Desenho e modelagem do banco

Como primeiro passo em direção à melhoria do fluxo de submissão e análise, decidiu-se alterar a forma como os dados são armazenados. Assim, projetamos um banco de dados relacional, construindo um modelo de entidades e relacionamentos (veja apêndice A), no qual separamos os dados em entidades (e.g. um pesquisador, uma bolsa) e relacionamentos entre as entidades (e.g. um aluno receber uma bolsa, um pesquisador orientar um aluno). Dessa forma, os dados podem ser salvos no banco relacional e somos capazes de recuperar ligações entre os dados e dados relacionados.

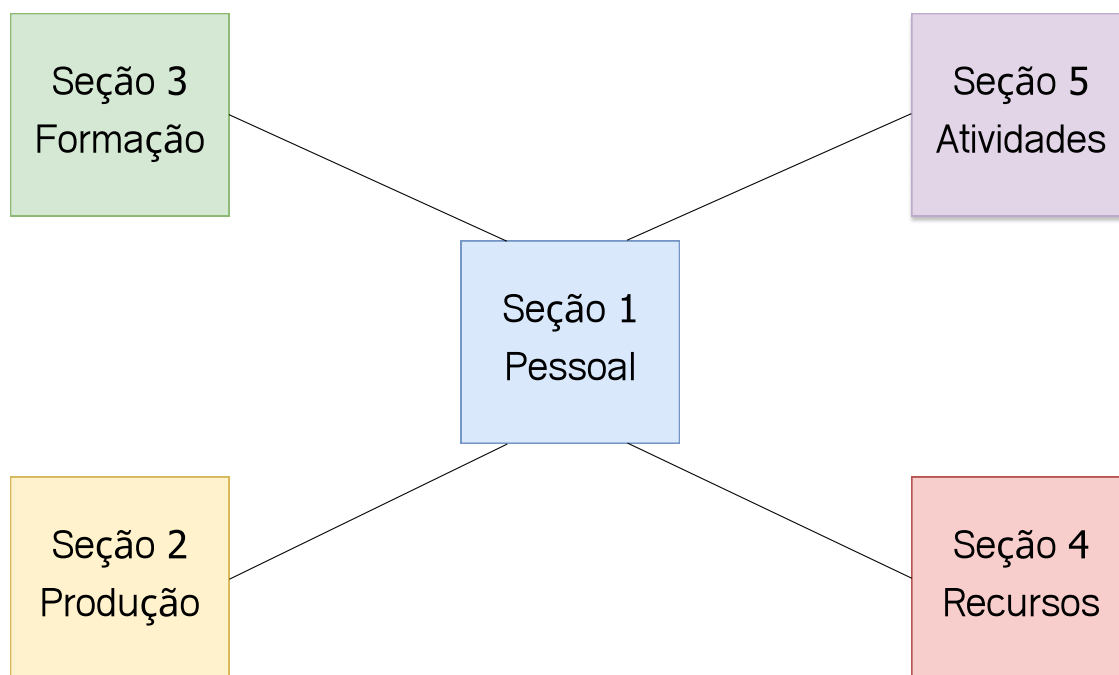


Figura 3.2: Organização em alto nível das entidades em cinco grupos, divididos por área; as linhas representam ligações de relacionamentos que ocorrem entre eles. Cada cor representa a mesma cor das entidades do grupo do MER (figura A.6). A seção 1 comporta dados pessoais de pesquisadores, funcionários e alunos (Ver figura A.1 para mais detalhes). A seção 2 compreende dados sobre produção científica como artigos (Ver figura A.2). A seção 3 representa a formação de recursos humanos, isto é, TCCs, teses e dissertações supervisionadas (Ver figura A.3). Na seção 4 estão as entidades referentes à captação de recursos, tais como bolsas e auxílios (Ver figura A.4). Por fim, na seção 5 estão dispostas informações a respeito de atividades exercidas pelos pesquisadores, como atividades administrativas (Ver figura A.5).

Carregamento do banco de dados

Com o banco de dados projetado, foi feita uma implementação preliminar, com a interação feita através de código e consultas SQL, com o intuito de fazer uma prova de conceito. Utilizamos essa primeira versão do banco de dados para fazer o relatório de produtividade do Instituto Butantan referente ao ano de 2018 e também para o relatório de análise individual.

A fonte dos dados para ambos os relatórios foram os formulários coletados no ano anterior. A inserção desses dados foi feita através de um *script* em Python, responsável por interpretar os dados das planilhas e inserí-los no banco de dados. Apesar do *script* funcionar bem de forma geral, ainda foi preciso monitorar o resultado durante a execução, devido a inconsistências no preenchimento (e.g. nomes escritos de formas diferentes em locais distintos da planilha; itens referenciados, mas nunca apresentados), com as quais o programa não conseguia lidar. Essas inconsistências foram tratadas manualmente, às vezes de forma rápida, como corrigindo um nome para usar a mesma abreviação em todo o documento, em outras ocasiões necessitando a comunicação com outro laboratório para correção.

Os *scripts* de carregamento, criação do banco de dados e outros *scripts* auxiliares dessa primeira versão do banco de dados estão disponíveis, como software livre de licença GPL-V3 (FSF, 2007), no repositório git do projeto (Scholl, 2019b).

Relatório de análise individual

Aproveitando a flexibilidade que o banco de dados traz, além do relatório anual de produção científica do Instituto Butantan, elaboramos também um relatório de análise individual: um relatório inspirado na análise feita no Instituto de Química da USP (IQ-USP), relatada por Marques (2015), que avalia o desempenho de cada pesquisador do instituto, levando em consideração, entre outros critérios, publicações de artigos e participação em atividades administrativas e atribuindo pontos a cada um desses itens. O resultado é uma pontuação atribuída a cada pesquisador, da qual é feito um *ranking* (veja figura 3.3). A finalidade dessa classificação, assim como foi no IQ-USP, é apenas a reflexão individual de cada pesquisador sobre sua colocação em relação aos seus pares e sobre a composição da sua pontuação (i.e. quais categorias predominam na pontuação).

Para garantir que essa avaliação seja apenas para reflexão individual, ela foi feita de forma anonimizada, isto é, a cada pesquisador foi atribuído um código que, por sua vez, foi enviado apenas ao seu respectivo pesquisador.

3.3 Módulo de Inserção de Dados

Para complementar o banco de dados já desenvolvido e resolver os problemas de interação com ele, principalmente na inserção dos dados da planilha, elaboramos uma interface gráfica interativa para coleta e inserção dos dados no banco. Essa interface é responsável por

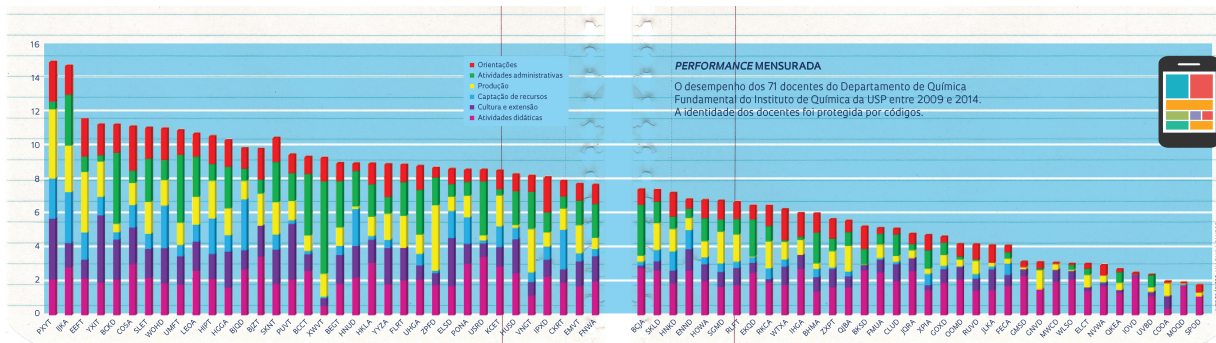


Figura 3.3: Classificação dos pesquisadores na análise realizada pelo Departamento de Química Fundamental do Instituto de Química da USP. Análise semelhante foi elaborada para os pesquisadores do Butantan, utilizando as informações de nosso banco de dados. Imagem retirada de Marques (2015), com autorização da editora. Autoria: Ana Paula Campos/Pesquisa FAPESP.

apresentar o formulário e fazer a ligação entre os dados preenchidos pelos pesquisadores e o banco de dados.

Apesar de já termos implementado um banco de dados, foi necessário adaptar e polir o desenho existente para torná-lo mais compatível com o esquema relacional e o modelo entidade-relacionamento. O banco de dados atual foi feito para aproveitar dados já existentes na planilha, mas que não deveriam ser guardados diretamente no banco de dados; por exemplo, dados calculados tal como o número de artigos publicados.

A interface gráfica do modulo foi feita em forma de sistema *web*, utilizando o arcabouço Vue (subseção 2.2.3) para o *front-end*, parte responsável pela parte visual, com a qual o usuário interage realmente e Ruby on Rails (subseção 2.2.2) para o *back-end*, a parte do sistema responsável pela lógica de negócio e a interação com o banco de dados. As partes se comunicam usando o padrão REST, comumente utilizado em aplicações desse tipo, que define operações sobre os dados, como leitura, escrita e atualização.

Ambos os sistemas de *front-end* e *back-end* estão disponíveis publicamente, em forma de software livre (licença GPLv3 FSF (2007)), nos repositórios git (Scholl, 2019c, Front-end), (Scholl, 2019a, Back-end).

Figura 3.4: Primeira seção do formulário na interface web.

Com esse módulo, a inserção dos dados preenchidos pelos pesquisadores passa a ser automática, porém a análise e criação de relatórios a partir do banco de dados permanece manual, requerendo um módulo dedicado a essa tarefa, que está nos planos futuros.

Capítulo 4

Caso de Uso

Neste capítulo apresentamos uma descrição da interface, de forma análoga a um tutorial. Para não tornar maçante a leitura do capítulo, foram ocultadas etapas intermediárias corriqueiras nesse tipo de sistema. No entanto, um manual mais detalhado e completo está em desenvolvimento. Essa descrição é voltada à explicação da utilização de alguns elementos mais complexos da interface, e apresentará os passos desde o cadastro do pesquisador à avaliação do preenchimento por parte dos diretores de divisão de laboratório.

A primeira tela em que o usuário se encontra quando entra no sistema é a página principal (figura 4.1), que conta com algumas informações e aponta para instruções. Em todas as páginas há um menu lateral para acesso às outras áreas do sistema (figura 4.2).

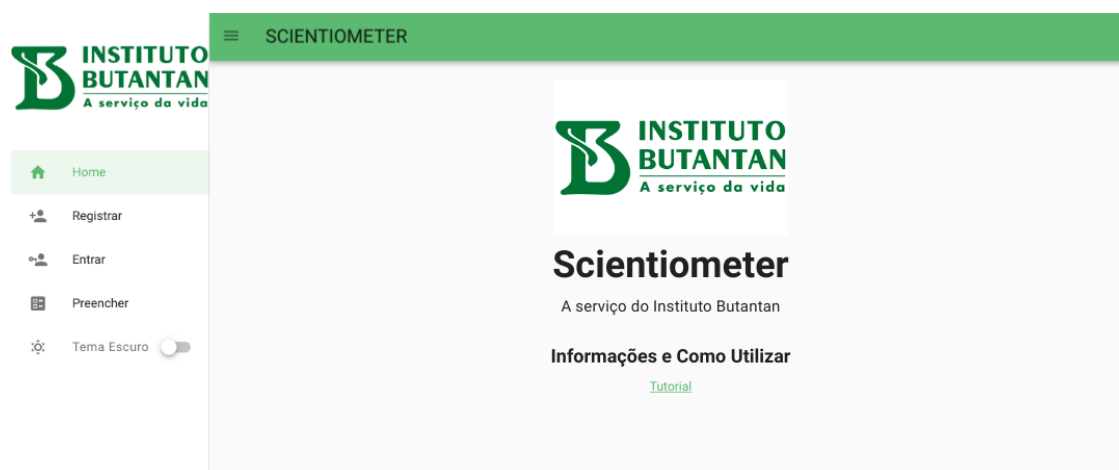


Figura 4.1: Página principal do sistema, a primeira a ser acessada.

4.1 Cadastro

O primeiro passo para a o preenchimento é a criação de uma conta, acessado através do menu lateral (figura 4.2), caso o pesquisador ainda não a tenha criado. Nesse cadastro é criada a credencial de acesso do pesquisador, que será utilizada para o acesso à plataforma. Usuário e senha podem ser de qualquer tamanho e conter quaisquer caracteres. A seleção

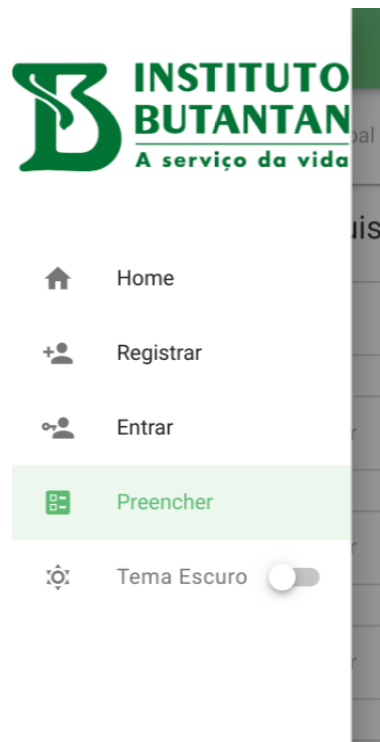


Figura 4.2: Menu lateral de navegação. Utilizado para alternar entre as telas do sistema.

de tipo de perfil é utilizada apenas no cadastro de administradores e normalmente oferece apenas uma opção (figura 4.3).

A imagem mostra a interface de usuário para o cadastro de uma nova conta. No topo, há uma barra verde com o texto 'SCIENTIOMETER' e um ícone de menu. Abaixo, há um formulário com o título 'Nova conta'. O formulário contém quatro campos de entrada: 'Usuário' (com ícone de pessoa), 'Senha' (com ícone de chave e botão de alternância de visibilidade), 'Repetir Senha' (com ícone de chave e botão de alternância de visibilidade) e 'Perfil' (com ícone de documento e seta para baixo). Abaixo dos campos, há um botão 'CRIAR CONTA'.

Figura 4.3: Página de cadastro do pesquisador, na qual é criada sua credencial de acesso ao sistema.

4.2 Login

Depois de realizado o cadastro, o pesquisador pode entrar no sistema com suas credenciais (figura 4.4).

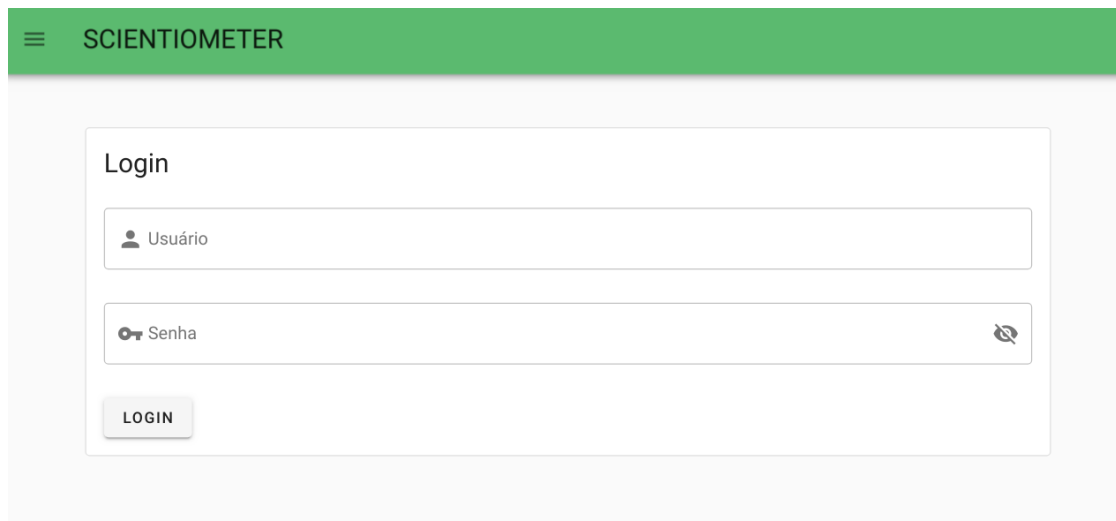


Figura 4.4: *Página de login. Utilizada para acessar o sistema a partir das credenciais cadastradas.*

4.3 Preenchimento do Formulário

O pesquisador, depois de autenticado, pode prosseguir para o preenchimento do formulário (figura 4.5). No decorrer do formulário, alguns elementos de entrada (e.g. caixas de seleção) aparecem com frequência. Alguns desses elementos são personalizados e a seguir terão instruções de como utilizá-los.



Figura 4.5: *Primeira seção do formulário, referente aos dados do pesquisador. Nela é possível ver também a navegação do formulário (topo), com as sete seções.*

4.3.1 Elementos de Inserção

Seleção Com Autocompletamento

O elemento de inserção com autocompletamento é uma entrada de elementos predefinidos, com a opção de digitar parte da entrada desejada para filtrar as opções exibidas (figura 4.6). É utilizada ao longo do formulário como entrada de agências de fomento, tipos de bolsas, seleção de alunos, entre outras.

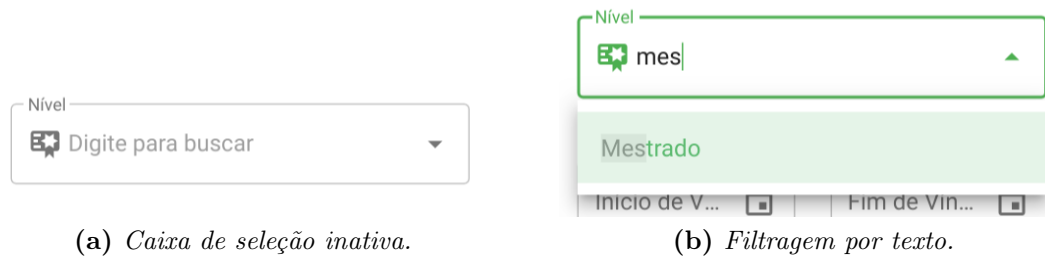


Figura 4.6: Elemento de seleção com autocompletamento. Na esquerda, é como ele aparece quando nada foi selecionado. Já na direita, a única opção que aparece é a que contém o texto pedido.

Seleção Múltipla

As caixas de seleção com autocompletamento são capazes de selecionar apenas uma opção por vez. Para suprir a necessidade de selecionar múltiplas opções, existem as seleções múltiplas (figura 4.7). Observe que, na figura 4.7(b), dentre as duas opções selecionadas, uma delas marcada para remoção. Para remover uma ou mais opções que foram selecionadas, basta tê-las marcadas e em seguida clicar em **remover**.

Múltiplas inserções

Esse elemento é composto por um formulário e uma tabela cujas entradas são feitas preenchendo esse formulário com dados diferentes (figura 4.8). Em particular, na figura 4.8(a) destacamos o elemento vazio, onde é possível ver o formulário e, logo acima, a tabela onde as entradas são colocadas. Para adicionar uma entrada basta preencher o formulário e clicar em **adicionar novo**. Por exemplo, na figura 4.8(b) a tabela contém duas entradas, uma delas selecionada para remoção. Para remover uma ou mais entradas, basta tê-las selecionadas e em seguida clicar em **remover**.

4.3.2 Submissão do Formulário

A entrega do formulário é feita de forma fracionada, onde cada seção é enviada separadamente. Isso possibilita ao pesquisador dedicar menos tempo cada vez que for trabalhar no preenchimento. O botão **enviar passo**, localizado no canto inferior esquerdo de cada passo

Participação de Mestrando

Nome

Nenhum selecionado

REMOVER

Participação de Mestrando

(a) Nenhuma opção selecionada.

Participação de Mestrando

Nome

Benício Souza

Mariela Lopes

REMOVER

Participação de Mestrando

Maurício Silva Santos

(b) Múltiplas opções selecionadas.

Figura 4.7: Elemento de **seleção múltipla**. Em Fig. 4.7(a), o elemento está sem opções selecionadas, enquanto que em Fig. 4.7(b) há duas opções selecionadas, uma delas marcada para remoção, restando ainda uma opção não selecionada.

do formulário, é responsável por submeter a seção e mudar a interface para o próximo passo (figura 4.9).

4.3.3 Avalização dos Formulários

A parte final da submissão de um formulário é a aprovação pelo diretor do laboratório para que os diretores de divisão possam fornecer os formulários ao CDC para análise. Esse passo passa a ser feito através da plataforma, por via das seguintes páginas.

Aprovação pela Diretoria de Laboratório

A lista de formulários submetidos (figura 4.10) pode ser acessada usando o menu lateral, estando autenticado como um diretor de laboratório. Os formulários estão listados exibindo informações como o nome do pesquisador que submeteu, última modificação e se foi aceito ou não, e podem ser ordenados por cada uma delas. Para acessar o formulário submetido,

Credenciamento em Programas de Pós-Graduação

<input type="checkbox"/>	Nome	Instituição
No data available		
REMOVER		

Programa

Instituição

Nível

ADICIONAR NOVO

(a) Elemento vazio.

Credenciamento em Programas de Pós-Graduação

<input type="checkbox"/>	Nome	Instituição
<input checked="" type="checkbox"/>	Ciência da Computação	IME-USP
<input type="checkbox"/>	Bioquímica	IQ-USP
REMOVER		

Programa

Instituição

Nível

ADICIONAR NOVO

(b) Múltiplas entradas adicionadas.

Figura 4.8: Elemento de inserção múltipla. Fig. 4.8(a): um elemento vazio; Fig. 4.8(b): um elemento com o formulário preenchido e duas entradas na tabela, uma delas selecionada para remoção.

ENVIAR PASSO

Figura 4.9: Botão enviar passo, responsável pelo envio dos dados da seção atual.

basta clicar na entrada e para aceitar a submissão usa-se o botão **aprovar**.

Pesquisador	Última Modificação	Aceito	Aprovar
Ronaldo da Silva	20/06/2019 08:43	Não	APROVAR
Roberto Miranda	12/04/2019 13:31	Não	APROVAR
Ana Pereira	22/08/2019 16:22	Não	APROVAR

Rows per page: 10 1-3 of 3

Figura 4.10: Lista de formulários com a aprovação pendente. A data e hora de última edição do formulário são exibidas para informação. Para que o formulário submetido seja aprovado, basta usar o botão **aprovar**.

Acompanhamento pela Diretoria de Divisão

Antes que os diretores de divisão possam enviar os formulários para análise, todos os formulários de todos os laboratórios da divisão devem ser preenchidos. Para que a diretoria da divisão de laboratórios possa acompanhar o estado das submissões nos laboratórios, há a relação de envios de formulários por laboratório, acessível pelo menu lateral (figura 4.11).

Laboratório	Aprovados
Laboratório Especial de Ciclo Celular	26/31
Laboratório Especial de Toxinologia Aplicada	12/28
Laboratório Ecologia e Evolução	5/12
Bacteriologia	4/20

Rows per page: 10 1-4 of 4

Figura 4.11: Lista de laboratórios da divisão, seguido do número de formulários já enviados, comparado com o total.

Capítulo 5

Discussão e Conclusões

Com o banco de dados preliminar fomos capazes de reduzir o tempo requerido para análise e sumarização dos resultados de produtividade científica, tornando o processo menos laborioso e mais rápido com a utilização de consultas reutilizáveis em SQL ao invés de um programa de edição de planilhas. Com o relatório de análise individual, demonstramos que a versatilidade do novo banco facilita a criação de novas consultas personalizadas e, dessa forma, a criação de novas avaliações que podem trazer novas perspectivas para a análise de produtividade e gerar resultados para o instituto. Ainda com o banco de dados preliminar, confirmamos que a coleta de dados a partir de planilhas, seja para sumarização manual ou inserção no banco, é ineficiente pois requer constante intervenção e correção manual de erros e inconsistências.

Para mitigar os problemas de inserção de dados elaboramos a interface *web* que apresenta o formulário para os pesquisadores de forma mais limpa e organizada do que o programa de planilhas, garante a ausência de inconsistências e previne erros de digitação por apresentar opções para seleção em diversos itens. Além disso, a interface insere os dados diretamente no banco, eliminando o passo que mais demandava tempo e intervenções, reduzindo substancialmente o tempo requerido para preparar os dados para análise.

Com a utilização interface *web*, toda a submissão dos dados pelos pesquisadores, avaliação por parte dos diretores de laboratório e avaliação e submissão por parte dos diretores de divisão passam a ser na mesma plataforma, sem a necessidades de envio de arquivos, nem diferentes versões de arquivos.

Para trabalhos futuros temos a implantação do sistema em um servidor do Instituto, para que possa ser utilizado já nas próximas coletas do formulário. Além disso, propomos a confecção de uma interface adicional para geração de análises e de relatórios diretamente no sistema *web*, o que permitirá que pessoas que não sabem programar gerem os relatórios e requisitem dados em poucos passos. Por fim, o trabalho realizado é apenas a base do que pode se tornar um sistema mais completo e, principalmente, mais conectado e integrado. Nesse primeiro passo foi feita a transformação do fluxo de submissão de dados em um sistema, o que gera oportunidades de elaboração de módulos complementares que facilitam ainda mais

o preenchimento ou, ainda, buscam dados de fontes externas, de forma automática, a partir de serviços como o Google Scholar e a plataforma Lattes.

Apêndice A

Modelo Entidade Relacionamento

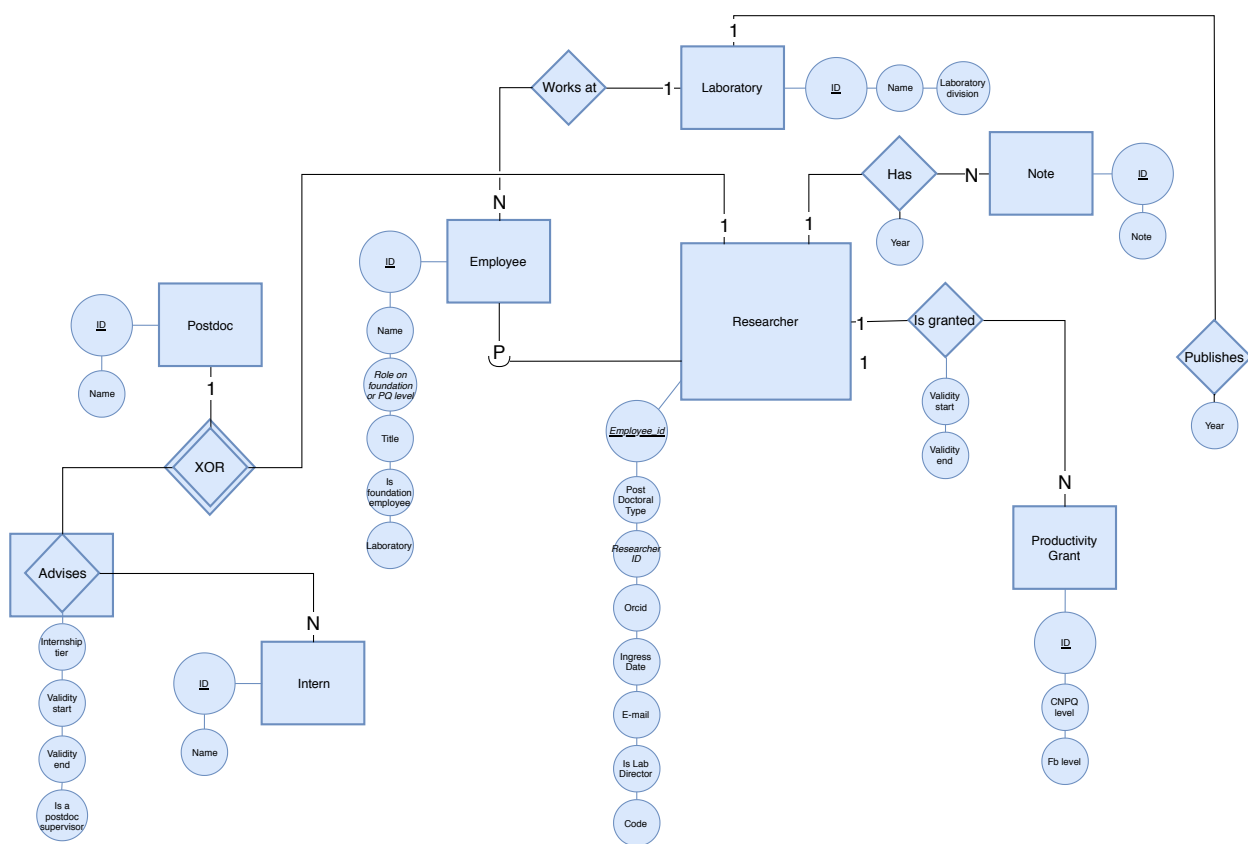


Figura A.1: Seção 1 do MER. Compreende os dados pessoais de pesquisadores, alunos, pós doutorandos e funcionários do laboratório. Note que bolsa de produtividade, na imagem Productivity Grant, está na seção de dados pessoais pois é um prêmio (veja subseção 2.1.11). Há um "abuso de notação" no losango com "XOR" que representa que apenas uma das duas entidades participa do relacionamento em uma dada instância.

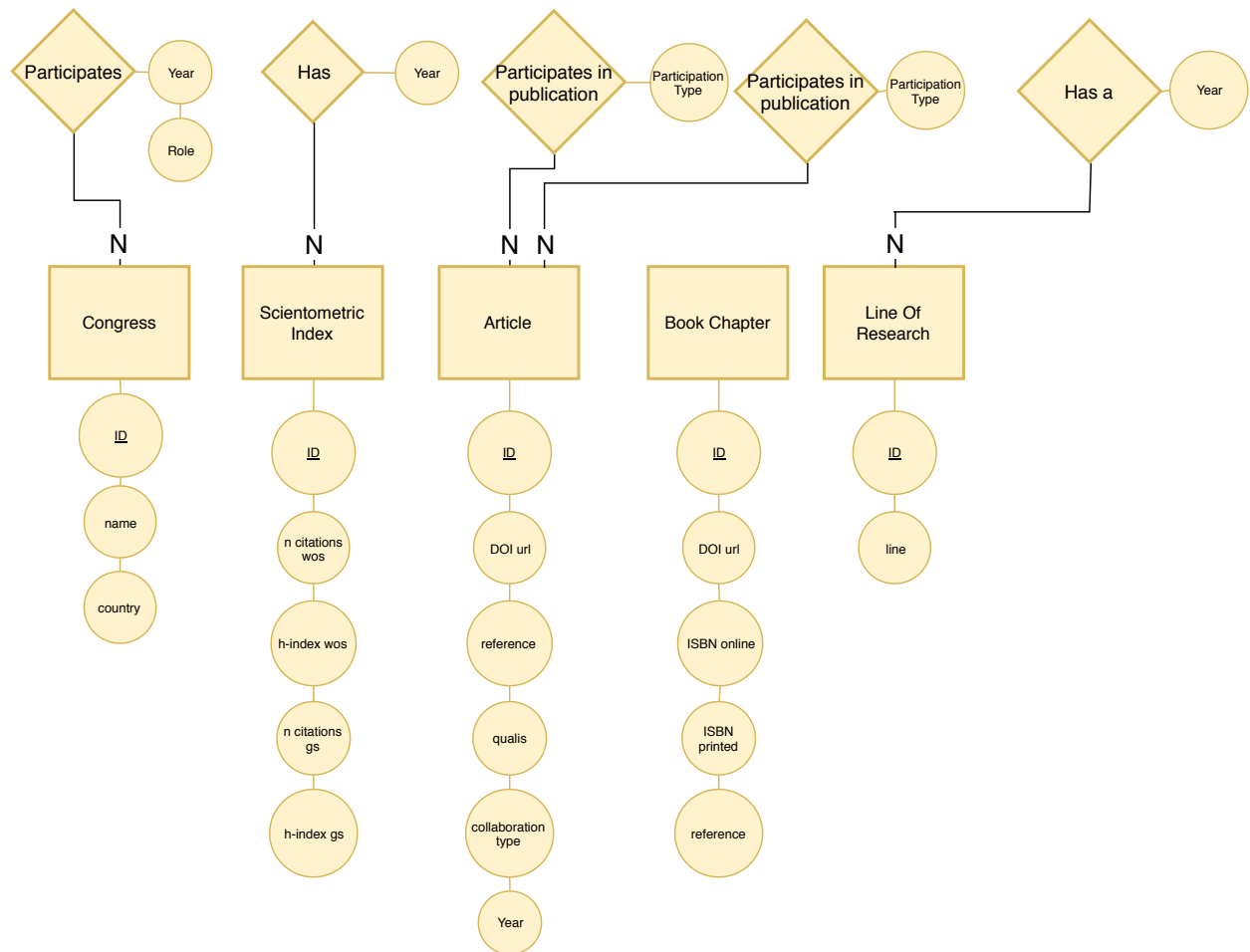


Figura A.2: Seção 2 do MER. Composta de dados de produção científica como artigos publicados, participações em congressos e também de dados de índices cientométricos do pesquisador.

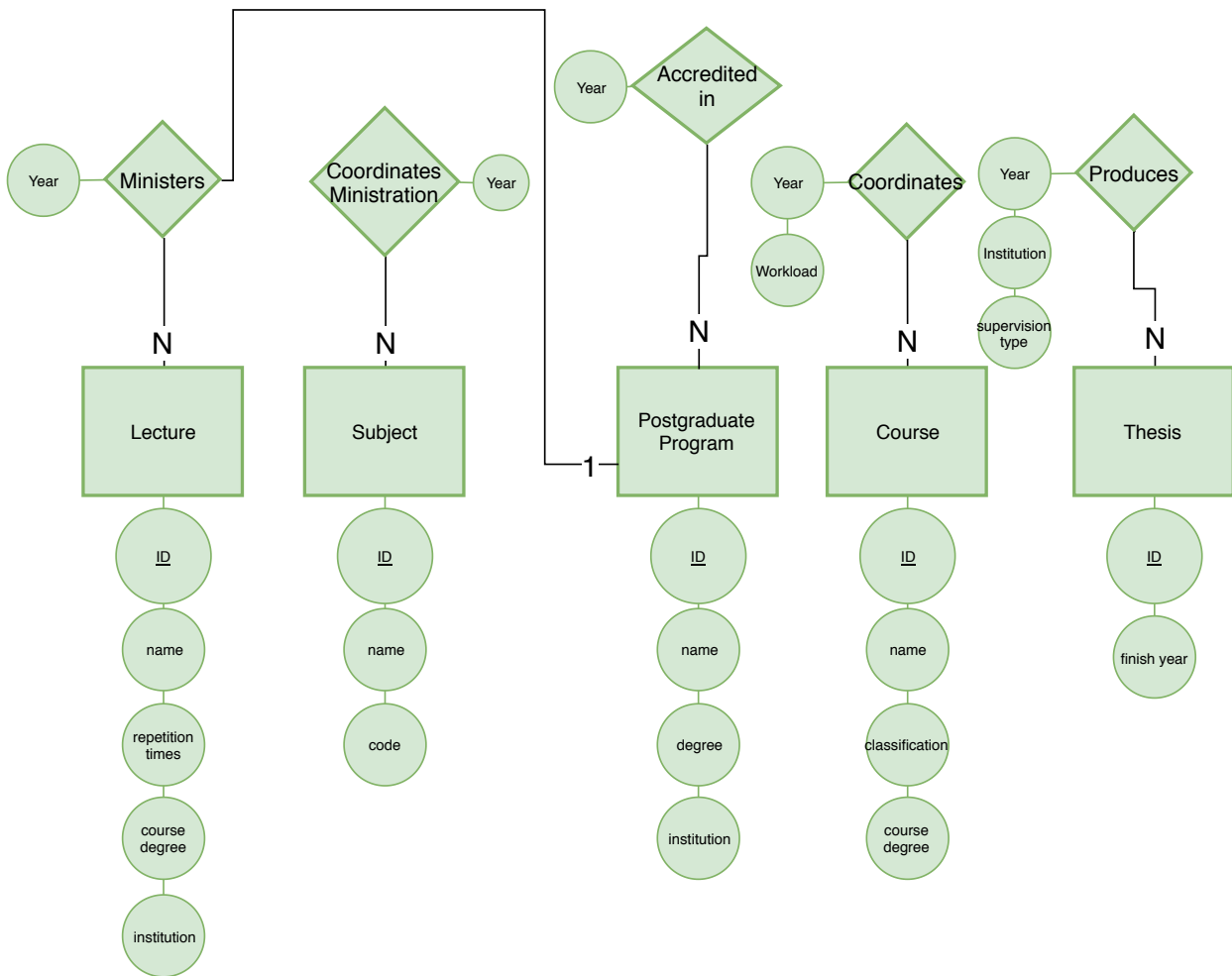


Figura A.3: Seção 3 do MER. Possui dados de aulas ministradas e credenciamento em programas de pós-graduação, assim como teses, TCCs e dissertações supervisionadas.

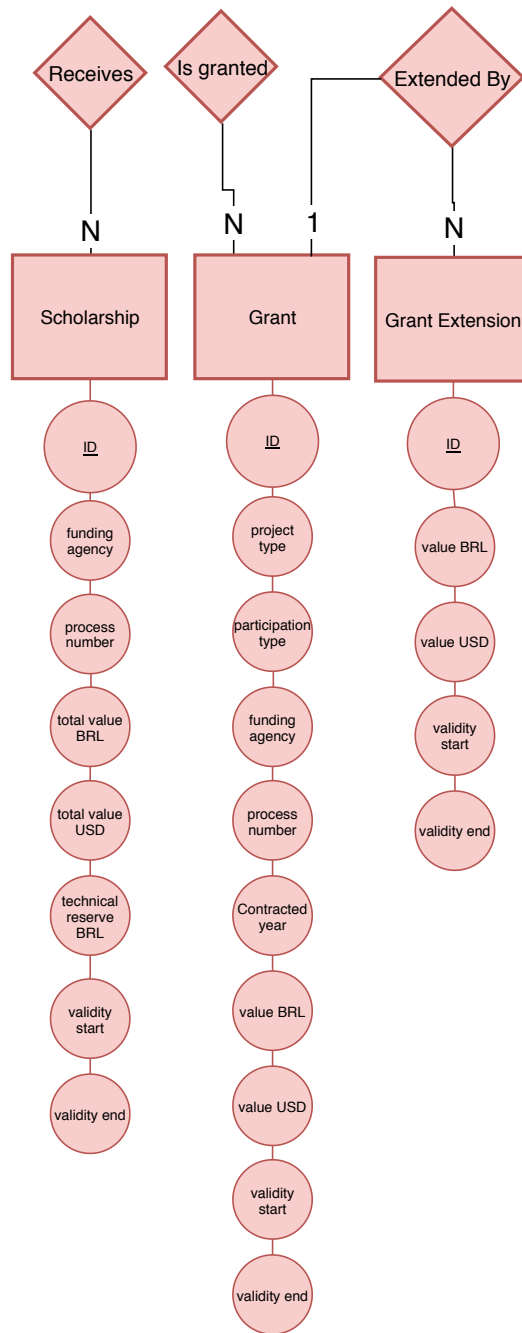


Figura A.4: Seção 4 do MER. Abriga dados de valores recebido e contratados como bolsas e auxílios.

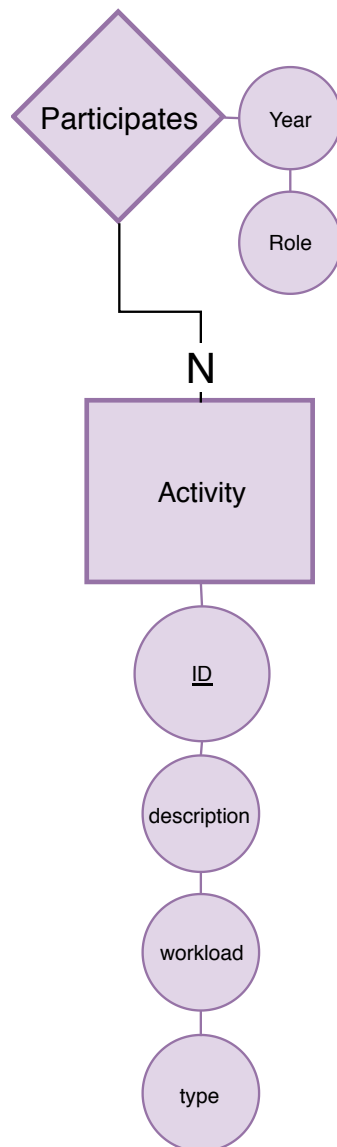


Figura A.5: *Seção 5 do MER. Conta com dados de atividades como atividades administrativas, translacionais e culturais realizadas por pesquisadores.*

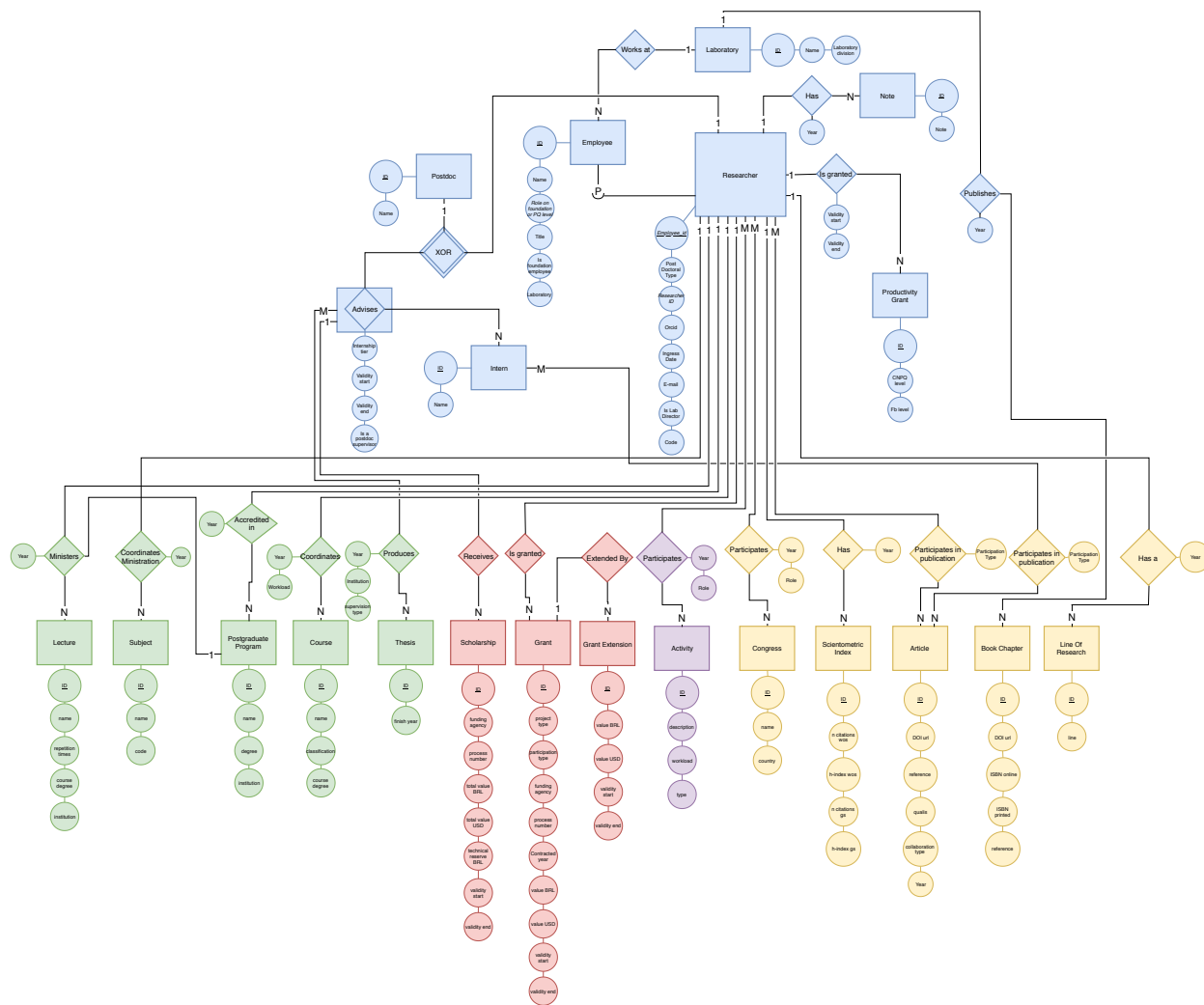


Figura A.6: Modelo Entidade-relacionamento completo do sistema. Nele é possível ver os diferentes grupos e como quase tudo está ligado com a entidade pesquisador. Algumas entidades menores foram omitidas em favor da legibilidade e estão representadas como atributos.

Referências Bibliográficas

- Bornmann(2011)** Lutz Bornmann. Scientific peer review. *Annual Review of Information Science and Technology*, 45(1):197–245. doi: 10.1002/aris.2011.1440450112. URL <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/aris.2011.1440450112>. Citado na pág. 6
- CAPES(2014)** CAPES. Qualis periódicos e classificação de produção intelectual, 2014. URL <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/qualis-periodicos-e-classificacao-de-producao-intelectual>. Acessado em 21/10/2019. Citado na pág. 9
- CAPES(2018)** CAPES. Critérios para credenciamento e credenciamento de orientadores, 2018. URL http://www.capes.gov.br/images/novo_portal/documentos/DAV/avaliacao/18072019_Esclarecimentos_Qualis2.pdf. Arquivo PDF. Acessado em 24/10/2019. Citado na pág. 9
- EP-USP(2004)** EP-USP. Critérios para credenciamento e credenciamento de orientadores, 2004. URL <https://www.poli.usp.br/pos-graduacao/stricto-sensu/185-criterios-para-credenciamento-e-credenciamento-de-orientadores.html>. Acessado em 22/10/2019. Citado na pág. 10
- Estado de São Paulo(1975)** Assembléia Legislativa Estado de São Paulo. Lei complementar nº 125, 1975. URL <http://www.pesquisador.sp.gov.br/lei.html>. Acessado em 16/10/2019. Citado na pág. 10
- FSF(2007)** FSF. Gnu general public license, version 3. <http://www.gnu.org/licenses/gpl.html>, 2007. Citado na pág. 18, 19
- Garfield(1955)** Eugene Garfield. Citation indexes for science. *Science*, 122(3159):108–111. Citado na pág. 7
- Garfield(1999)** Eugene Garfield. Journal impact factor: a brief review. *CMAJ*, 161(8): 979–980. ISSN 0820-3946. URL <https://www.cmaj.ca/content/161/8/979>. Citado na pág. 8
- HAAK et al.(2012)** Laurel L. HAAK, Martin FENNER, Laura PAGLIONE, Ed PENTZ e Howard RATNER. Orcid: a system to uniquely identify researchers. *Learned Publishing*, 25(4):259–264. doi: 10.1087/20120404. URL <https://onlinelibrary.wiley.com/doi/abs/10.1087/20120404>. Citado na pág. 11
- Hirsch(2005)** Jorge E Hirsch. An index to quantify an individual’s scientific research output. *Proceedings of the National academy of Sciences*, 102(46):16569–16572. Citado na pág. 7
- Krasner et al.(1988)** Glenn E Krasner, Stephen T Pope et al. A description of the model-view-controller user interface paradigm in the smalltalk-80 system. *Journal of object oriented programming*, 1(3):26–49. Citado na pág. 12, 13

- Marques(2015)** Fabrício Marques. O melhor de cada um, 2015. revistapesquisa.fapesp.br/2015/06/16/o-melhor-de-cada-um/. Citado na pág. 18, 19
- Marín-Martín et al.(2016)** Alberto Marín-Martín, Enrique Orduña-Malea, Juan Manuel Ayllon e Emilio Delgado López-Cózar. The counting house: measuring those who count. presence of bibliometrics, scientometrics, informetrics, webometrics and altmetrics in the google scholar citations, researcherid, researchgate, mendeley & twitter. *CoRR*, abs/1602.02412. URL <http://arxiv.org/abs/1602.02412>. Citado na pág. 11
- Mattedi e Spiess(2017)** Marcos Antônio Mattedi e Maiko Rafael Spiess. A avaliação da produtividade científica. *História, Ciências, Saúde-Manguinhos*, 24(3). Citado na pág. 1
- Moustafa(2015)** Khaled Moustafa. The disaster of the impact factor. *Science and Engineering Ethics*, 21(1):139–142. ISSN 1471-5546. doi: 10.1007/s11948-014-9517-0. URL <https://doi.org/10.1007/s11948-014-9517-0>. Citado na pág. 8
- Paskin(2010)** Norman Paskin. Digital object identifier (doi®) system. *Encyclopedia of library and information sciences*, 3:1586–1592. Citado na pág. 6, 7
- PedroII(2019)** Colégio PedroII. O que é e como é feita a avaliação capes, 2019. URL <http://www.cp2.g12.br/blog/mpcp2/conteudo/6>. Acessado em 07/11/2019. Citado na pág. 9
- Rails(2005)** Ruby on Rails, 2005. URL <https://rubyonrails.org/>. Acessado em 22/10/2019. Citado na pág. 12
- Scholl(2019a)** Bruno B. Scholl. Scientiometer back-end, 2019a. URL [COLOCARURL](#). Repositório GIT. Acessado em 06/11/2019. Citado na pág. 19
- Scholl(2019b)** Bruno B. Scholl. Scientiometer database, 2019b. URL <https://github.com/msreis/Scientiometer>. Repositório GIT. Acessado em 06/11/2019. Citado na pág. 18
- Scholl(2019c)** Bruno B. Scholl. Scientiometer front-end, 2019c. URL <https://gitlab.com/brunobbs/scientiometer-front>. Repositório GIT. Acessado em 06/11/2019. Citado na pág. 19
- SIBi-USP(2019)** SIBi-USP, 2019. URL <http://www.sibi.usp.br/apoio-pesquisador/identificacao-pesquisadores/researcher-id/>. Acessado em 15/10/2019. Citado na pág. 11
- Vine(2006)** Rita Vine. Google scholar. *Journal of the Medical Library Association*, 94(1): 97. Citado na pág. 11