

Análise da pesquisa espacial brasileira sob a ótica da Biblioteca Digital Brasileira de Teses e Dissertações

Lillian Alvares

Coordenadora-Geral de Produtos Consolidados do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Pós-Doutorado em Sistemas Informáticos, na Escola Superior de Tecnologia i Ciències Experimentals da Universitat Jaume I, Espanha, <http://lattes.cnpq.br/5541636086123721>, lillian@alvarestech.com

Maria Virgínia Alves

Pesquisadora e Chefe de Gabinete do Instituto Nacional de Pesquisas Espaciais, Doutorado em Geofísica Espacial e Mestre em Astrofísica pela mesma instituição, <http://lattes.cnpq.br/6460301165792800>, mvirginia.alves@inpe.br

Silvia Castro Marcelino

Chefe do Serviço de Informação e Documentação do Instituto Nacional de Pesquisas Espaciais (INPE), Mestre em Semiótica, Tecnologias da Informação e Educação pela Universidade Braz Cubas, <http://lattes.cnpq.br/9643135223468862>, silvia.marcelino@inpe.br

Bianca Amaro

Coordenadora do Laboratório de Metodologias de Tratamento e Disseminação da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Doutora em Linguística Aplicada pela Universidade Pompeu Fabra (2003), Espanha, <http://lattes.cnpq.br/1445782939373313>, bianca@ibict.br

Taina Batista de Assis

Coordenadora de Atendimento à Comunidade do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Mestre em Ciência da Informação pela Universidade de Brasília, <http://lattes.cnpq.br/8018556894529318>, taina@ibict.br

Submetido em: 22/08/2017. Aprovado em: 24/10/2017. Publicado em: 22/02/2018.

RESUMO

Descreve a pesquisa científica e tecnológica brasileira na área espacial desenvolvida nos programas de pós-graduação stricto sensu oferecidos em universidades e institutos de pesquisa no período de 2006 a 2016 sob a ótica da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), repositório digital concebido e mantido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) desde 2002. A metodologia utilizada é a análise automática de informação por meio da mineração de dados a partir de softwares livres. Considerando as inúmeras possibilidades de análise que emergiram da metodologia proposta, optou-se pela pesquisa espacial para estudo aprofundado. Os resultados comprovam que é possível revelar as principais instituições, áreas de concentração, pesquisadores, relacionamentos entre pares, entre outros, a partir de estudo quantitativo de metadados das teses e dissertações nacionais. De outra perspectiva, esta pesquisa demonstra que é possível mapear a ciência brasileira a partir do retrato explicitado no estudo estatístico da BDTD, revelando o que de fato integra o desenvolvimento científico e tecnológico nacional.

Palavras-chave: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Pesquisa espacial. Análise automática de informação. Mineração de dados. Pesquisa científica e tecnológica brasileira.

Analysis of Brazilian space research from the perspective of the Brazilian Digital Library of Theses and Dissertations

ABSTRACT

Describes the Brazilian scientific and technological research in the spatial area developed in the stricto sensu postgraduate programs offered in universities and research institutes from 2006 to 2016 from the perspective of the Brazilian Digital Library of Theses and Dissertations (BDTD), a digital repository conceived and maintained by the Brazilian Institute of Information in Science and Technology (IBICT) since 2002. The methodology used is the automatic analysis of information through the mining of data from free software. Considering the numerous possibilities of analysis that emerged from the proposed methodology, we opted for spatial research for in-depth study. The results show that it is possible to reveal the main institutions, areas of concentration, researchers, peer relationships, among others, from a quantitative study of theses metadata and national dissertations. From another perspective, this research demonstrates that it is possible to map Brazilian science from the explicit picture in the statistical study of BDTD, revealing what in fact integrates national scientific and technological development.

Keywords: *Brazilian Digital Library of Theses and Dissertations (BDTD). Space research. Automatic analysis of information. Data Mining. Brazilian scientific and technological research.*

Análisis de la investigación espacial brasileña bajo la óptica de la Biblioteca Digital Brasileña de Tesis y Disertaciones

RESUMEN

Describe la investigación científica y tecnológica brasileña en el área espacial desarrollada en los programas de postgrado stricto sensu ofrecidos en universidades e institutos de investigación en el período de 2006 a 2016 bajo la óptica de la Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD), repositorio digital concebido y mantenido por el Instituto Brasileño de Información en Ciencia y Tecnología (IBICT) desde 2002. La metodología utilizada es el análisis automático de información a través de la minería de datos a partir de software libre. Considerando las innumerables posibilidades de análisis que surgieron de la metodología propuesta, se optó por la investigación espacial para un estudio en profundidad. Los resultados demuestran que es posible revelar las principales instituciones, áreas de concentración, investigadores, relaciones entre pares, entre otros, a partir del estudio cuantitativo de metadatos de las tesis y disertaciones nacionales. De otra perspectiva, esta investigación demuestra que es posible mapear la ciencia brasileña a partir del retrato explicitado en el estudio estadístico de la BDTD, revelando lo que de hecho integra el desarrollo científico y tecnológico nacional.

Palabras clave: *Biblioteca Digital Brasileña de Tesis y Disertaciones (BDTD). Investigación espacial. Análisis automático de la información. Minería de datos. Investigación científica y tecnológica brasileña.*

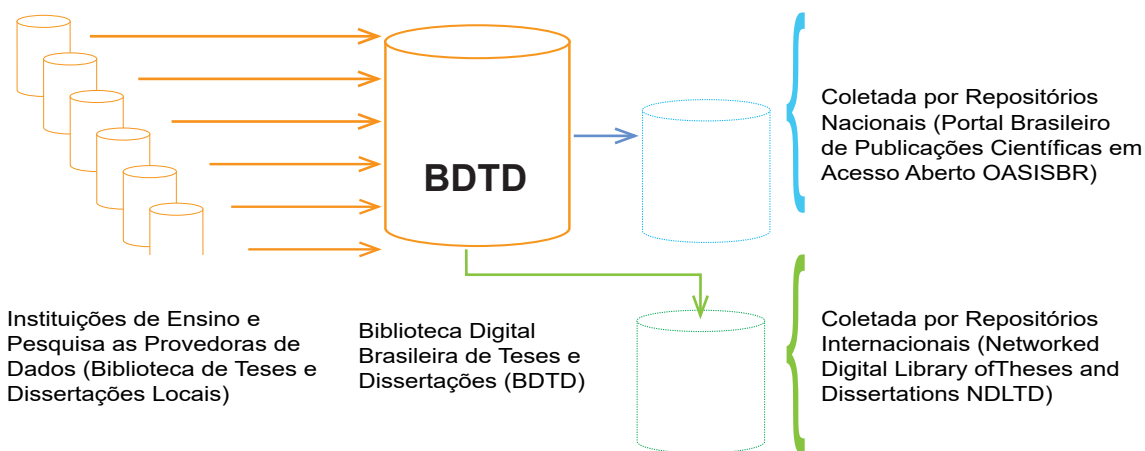
INTRODUÇÃO

Trata-se de estudo prospectivo da Biblioteca Digital Brasileira de Teses e Dissertações BDTD (<http://bdtd.ibict.br/vufind/>), banco de dados que contém os resultados da pós-graduação stricto sensu do Brasil, com objetivo de analisar a pesquisa científica e tecnológica brasileira na área espacial realizada no período de 2006 a 2016.

De acordo com Vanz e Stumpf (2010), bases de dados de produção científica são fundamentais para a geração de indicadores científicos e tecnológicos, assim como afirmam antes também Velho (1986) e Sancho (1990). No caso da BDTD, são quase 400.000 registros (100.821 teses e 276.344 dissertações), oriundos de 105 instituições brasileiras, baseados no Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), protocolo desenvolvido pela Open Archives Initiative que define um mecanismo para coleta de registros de metadados em repositórios de dados, adotando assim o modelo baseado em padrões de interoperabilidade.

A BDTD é uma rede distribuída de sistemas de informação que gerenciam teses e dissertações, com texto completo, sendo as instituições de ensino e pesquisa as provedoras de dados e o IBICT operando como agregador (TRISKA; CAFÉ 2001). Isto é, o processo abrange a coleta dos metadados das teses e dissertações dos provedores (instituições de ensino e pesquisa), organiza-os em um único repositório e os expõem para coleta para outros provedores de serviços, como, por exemplo, a Networked Digital Library of Theses and Dissertations ND LTD (<http://www.ndltd.org/>), organização internacional que promove a criação, divulgação e preservação de teses e dissertações eletrônicas de vários países. O processo está ilustrado na figura 1.

Figura 1 – Ecossistema da Biblioteca Digital Brasileira de Teses e Dissertações



Fonte: Elaborado pelos autores, 2017.

Os principais objetivos que motivaram a criação da BDTD foram estudar experiências existentes no Brasil e no exterior de desenvolvimento de bibliotecas digitais de teses e dissertações, definir padrões de metadados e tecnologias a serem utilizadas pelo sistema e desenvolver um sistema de publicação eletrônica de teses e dissertações para atender àquelas instituições de ensino e pesquisa que não possuíam sistemas automatizados para implantar suas bibliotecas digitais. A implantação de sistemas de publicação de teses e dissertações locais, em particular, foi o impulso necessário à criação dos repositórios institucionais de teses e dissertações, possibilitando assim o pleno cumprimento da missão da BDTD.

Naquele momento de criação, no entanto, não havia a intenção de explorar o grande manancial de informação reunida nesse ambiente. Ao grande volume de dados e informações científicas disponibilizadas, não havia a correspondente realização da etapa de análise da informação. Isto é, o projeto da BDTD cumpre todo o ciclo informacional, ou seja, seleção, aquisição, tratamento, armazenamento e disseminação, mas não considera a etapa final de análise.

A ciência da informação (CI) e a ciência da computação (CC), no entanto, avançaram muito nas últimas décadas nos recursos para análise automática de grandes volumes de dados, de fontes estruturadas ou não. As metodologias mais sofisticadas de análise de dados estão relacionadas ao conceito de descoberta de conhecimento que, basicamente, são metodologias que deixam distinguir ocorrências que seriam impossíveis visualizar por meio de um sistema convencional de recuperação de informação em bases de dados. Nesse contexto, está a mineração de dados, que se resume em uma técnica de análise automática para grande volume de informações. Ou ainda, num processo de extrair informação, previamente desconhecida e de máxima abrangência, a partir de grande base de dados (HAN; KAMBER; PEI, 2011).

Figura 2 – Etapas do processo de mineração de dados



Fonte: Elaborado pelos autores, 2017.

As possibilidades de análise desse processo podem ser assim relacionadas:

- classificação: organização em um conjunto em classes predefinidas;
- agrupamento (clustering): identificação de um conjunto finito de grupos ou categorias que descrevam as características comuns;
- modelagem de dependências: consiste na procura de um modelo que descreva as dependências significativas entre variáveis, como a análise de ligações, que procura relações entre campos dos registros ou análise de frequências, que procura padrões semelhantes em dados sequenciais a fim de descobrir causas e tendências associadas a um padrão específico.

No mais, importa destacar o sequenciamento das etapas de realização para mineração de dados que pode, simplificada, ser representada pela figura 2.

Os resultados alcançados sobre a pesquisa espacial estão concentrados na distribuição dos programas de pós-graduação, no perfil dos programas de pós-graduação, na quantidade de trabalhos produzidos pelos programas de pós-graduação, no número de orientadores e quantidade de trabalhos orientados, nas redes de cooperação entre programas e pesquisadores, na produtividade dos programas por comparação, entre outros.

Não menos importante, e com potencial de motivar e facilitar outros trabalhos semelhantes, destacam-se os resultados da identificação dos softwares livres utilizados para análise de informação estruturada, pois raramente são encontradas opções que satisfaçam várias soluções. Observou-se, nesses softwares livres, a tecnologia utilizada, as facilidades ofertadas, as formas de integração e de conectividade com demais ambientes web e de desenvolvimento, a função a ser realizada e a comunidade que os mantém (ALVARES, 2017).

METODOLOGIA

A metodologia utilizada foi a mineração de dados (HAN; KAMBER; PEI, 2011), a partir de dados estruturados por meio de softwares de preparação (incluindo limpeza), extração, tratamento (incluindo manipulação), processamento (incluindo visualização), interpretação e análise de dados, processo conhecido como Descoberta de Conhecimento em Bases de Dados ou em inglês Knowledge Discovery in Databases (KDD). Criou-se um referencial em software livre para tratamento automático de informação, utilizando os softwares Go JS para confecção dos gráficos, Jupyter para programação em Python, Knime para mineração de dados, Open Refine para carregamento e leitura dos dados e Tableau para limpeza, transformação e visualização dos dados.

O período de avaliação foi de 2006 a 2016, tendo em vista a confiabilidade dos dados, pois somente a partir de 2006 as instituições que oferecem programas de pós-graduação stricto sensu foram obrigadas a depositar os seus resultados de pesquisa nos repositórios institucionais. Esse fato não limitou que os registros estejam apenas nesse período de tempo, ao contrário, as instituições alimentam retrospectivamente seus repositórios institucionais e possuem registros significativos desde 1997.

Para a pesquisa, no entanto, somente o período entre 2006 e 2016 foi considerado, tendo em vista a garantia de sua completeza. Depois, notou-se a necessidade de restringir ainda mais o período, para 2015, uma vez que a extração do conjunto de dados de interesse foi realizada em 2016, e portanto uma parte das instituições ainda estava alimentando seus repositórios institucionais daquele ano. Por segurança, optou-se por finalizar o conjunto em 2015.

O processo de análise de informação exige que a fonte de dados esteja homogênea e sem ruídos informacionais, e nesse ponto teve início a primeira barreira a ser superada. Identificaram-se problemas na alimentação de dados pelas instituições e problemas na gestão do próprio sistema de coleta da BDTD.

Assim, a primeira etapa foi a realização da limpeza na base de dados em uso no projeto por meio da identificação de regras de correção que foram implementadas por meio de programação a partir de modelos estatísticos. Com essa informação, gerou-se uma relação de indicação de solução e sugestões de melhorias na entrada de dados nos repositórios de dados por instituição.

Como exemplo, criaram-se as “Regras para Autores”, a fim de uniformizar a representação dessa informação e evitar ocorrências como: Stern\ José Luiz [UNESP] (em dois campos, invertido e com o nome da instituição inserido); Leite\Carlos Eduardo\ (em dois campos e invertido); nome da instituição no lugar de nome do autor, entre outros, a fim de obter resultados da seguinte maneira: YolandaMariaFaccini (em um único campo, de forma direta).

Aqui cabe ressaltar a decisão sobre a acentuação. No campo de autores e palavras-chave isso se tornou um imenso obstáculo pela falta de padronização. Optou-se por retirar todos os acentos e padronizar a escrita das palavras, mesmo com o prejuízo à correção do idioma e fidelidade ao nome do autor.

As “regras para palavras-chave” apresentaram muitos problemas, como por exemplo, as unidades terminológicas complexas que acabaram desfiguradas, como “Beckett\,Samuel\,” (em dois campos) quando deveria ser “Samuel Beckett” ou “Crohn\,Doença de” quando deveria ser “Doença de Crohn”. Ou o uso indevido da numeração da Classificação Decimal Universal, como 8.02.10.00-7 no lugar da palavra-chave.

Outro problema eram os delimitadores dentro de um mesmo campo de palavra-chave, como o caso: “recem-nascidos - mortalidade - brasil, sistemas de recuperacao da informacao - saude publica - brasil,sistemaunico de saude (brasil),saude publica”. Observe que as palavras se repetem, criando uma multiplicação da mesma expressão ou verbete. Além disso, nesse mesmo caso, o sinal - aparece com dois usos, tanto para delimitar as palavras, como para escrever palavras como “recem-nascidos”. =

Como fazer a separação e guardar a integridade dos dados? Aqui igualmente, a programação em Python, a partir de modelos predefinidos, com o software Jupyter foi necessária para asolução dos problemas de repetição e de delimitação. Vale observar a falta de padronização da acentuação no exemplo acima, confirmando a decisão de retirá-los para garantir uma análise adequada. Da mesma maneira, apenas por indicação estética, todas as palavras, ao final, assumiram a forma maiúscula.

As “regras para rede de pesquisadores” consideraram que os campos reservados para orientador, às vezes continham dois nomes e às vezes cinco nomes. Concluiu-se (e assumiu-se) que com dois havia também o coorientador e com mais de dois, a banca examinadora. Com isso, criaram-se regras para verificar redes de pesquisadores explícitas e as palavras-chave em torno dos orientadores.

No que se refere à análise dos dados e descoberta de informação, optou-se pela mineração de dados com fins de descrição e predição da BDTD, com abordagens de classificação (mapeamento de dados para agrupamento em classes principais), regressão (descoberta de relações entre as variáveis), clustering (identificação de um conjunto de dados com características semelhantes) e sumarização (descrição do conhecimento).

Outro processo importante envolvido na construção do conjunto de dados para análise foi a classificação automática de palavras-chave de acordo com sua ocorrência (ANAYA-SÁNCHEZ; PONS-PORRATA; BERLANGA, 2011). Basicamente, as palavras-chave foram tratadas como um documento e as probabilidades condicionais entre elas foram calculadas, a fim de classificá-las em um conjunto predefinido de áreas de investigação, resultante da probabilidade entre palavras-chave e áreas predefinidas. A cada entrada no banco de dados foram atribuídas as áreas mais prováveis para a pesquisa de acordo com as probabilidades acima. Grande parte do projeto centrou-se na identificação de grandes áreas de pesquisa predefinidas, de modo que todas as entradas no banco de dados estivessem completamente cobertas.

O próximo passo, com a base de dados homogênea (e enriquecida), traduz as infinitas possibilidades de análise da pesquisa espacial brasileira: (i) decidida a área do conhecimento que será analisada (pesquisa espacial), como será a extração do grupo de dados de interesse da BDTD? Por qual palavra-chave? Em uma instituição? Em várias instituições? Quais os principais orientadores? Ou, (ii) qual instituição será analisada? Em várias subáreas? Em algumas áreas? Ou ainda, (iii) qual a trajetória de pesquisa dos orientadores? Ou também (iv) como se classifica determinado segmento, qual a taxonomia praticada? Enfim, a opção por qual perspectiva analisar é, de fato, infinita. As decisões tomadas serão apresentadas a seguir.

RESULTADOS

A PESQUISA ESPACIAL BRASILEIRA

As atividades espaciais no Brasil tiveram início logo após os lançamentos dos satélites Sputnik 1 pela então União Soviética (em 1957) e do Explorer 1 pelos Estados Unidos (em 1958). Em 1961 foi criado o Instituto Nacional de Pesquisas Espaciais (Inpe), atualmente um dos institutos de pesquisas do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) do Governo Federal.

Com o início de sua história motivado pelas expectativas criadas em torno das primeiras conquistas espaciais, o Inpe tornou-se um centro de referência do Brasil no campo das atividades espaciais e suas aplicações. As atribuições iniciais do instituto eram: propor a política espacial brasileira; executar projetos de pesquisas espaciais; desenvolver intercâmbio técnico-científico e cooperação internacional; promover a formação de especialistas e a coordenação entre as atividades espaciais e a indústria brasileira (BRASIL, 1961).

Em sua história e nos dias atuais, o instituto mantém o compromisso com a excelência na produção da ciência e da tecnologia espacial, alinhada aos padrões internacionais.

“Embora a pesquisa espacial no Brasil não esteja restrita ao Inpe, com um número crescente de pesquisadores e professores em universidades e centros de pesquisa, o instituto ainda é o principal centro nacional no tema” (CÂMARA, 2011).

As atividades atuais do Inpe são regularizadas por políticas definidas pelo governo federal em documentos específicos, como a Estratégia Nacional de Ciência, Tecnologia e Inovação (Encti) e o Programa Nacional de Atividades Espaciais (Pnae). Como um dos executores do Pnae, o Inpe é o órgão responsável pelo desenvolvimento de satélites científicos, meteorológicos e de observação da Terra.

“As missões desses satélites estão previstas no PNAE e respondem às necessidades governamentais para a implementação de diversas políticas públicas e para a solução de problemas nacionais” (INPE, 2017, p.18).

Atualmente, o Inpe tem como missão:

“Desenvolver, operar e utilizar sistemas espaciais para o avanço da ciência, da tecnologia e das aplicações nas áreas do espaço exterior e do ambiente terrestre, e oferecer produtos e serviços inovadores em benefício do Brasil” (INPE, 2016, p.13).

As competências científicas e tecnológicas do Inpe, reconhecidas nacional e internacionalmente, concentram-se nas seguintes áreas: Ciências Espaciais e Atmosféricas, Ciências Ambiental e Meteorológica e Engenharia e Tecnologias Espaciais.

Para plena execução da política espacial brasileira, o Inpe mantém cooperação com instituições nacionais e internacionais, tanto públicas quanto privadas, que, além de possibilitar intercâmbio científico e tecnológico e captação de recursos, permitem fortalecimento de suas atividades de pesquisa e desenvolvimento, acesso e fornecimento de dados, desenvolvimento de serviços, tecnologias e sistemas espaciais (INPE, 2016, p.26).

Ao longo de seus 56 anos, o Inpe adquiriu competências em atividades de desenvolvimento científico e tecnológico gerando conhecimentos, produtos, processos e serviços que são difundidos à sociedade, como: geração e distribuição de

imagens de satélites, coleta e distribuição de dados meteorológicos, monitoramento do território (desmatamento de florestas, queimadas, raios e mudanças climáticas), bem como a formação de recursos humanos.

Como o país não contava na época da criação do Inpe com a preparação de especialistas em pesquisa e desenvolvimento espaciais, o instituto iniciou a formação de recursos humanos especializados em 1968, com a criação do primeiro curso de mestrado. Dessa data até 2016, a pós-graduação do Inpe formou 2.123 mestres e 794 doutores nos diversos cursos do instituto. Atualmente o Inpe possui cursos de mestrado e doutorado em: Astrofísica, Computação Aplicada, Engenharia e Tecnologias Espaciais, Geofísica Espacial, Meteorologia, Sensoriamento Remoto e Ciência do Sistema Terrestre.

A produção acadêmica dos alunos de pós-graduação está, em geral, inserida nos projetos das instituições. No caso do Inpe, as teses e dissertações em sua maioria, são em maior ou menor grau, relacionadas à pesquisa espacial.

A CRIAÇÃO DO CONJUNTO DE DADOS

O radical selecionado para a criação do conjunto de dados foi *spac*. A hipótese é que toda a gama de temas relacionados à pesquisa espacial será recuperada como esse radical.

Ao total, foram recuperados 4.523 registros. Desses, todos aqueles associados como os seguintes temas foram retirados: agricultura familiar, agronomia, alimentos, animais, arte, aspectos ambientais, aspectos sociais, alimentacao, cidadania, ciencias agrarias, cinema, comunicacao, conservacao, consumo, crianca, cultivo, cultura, democracia, desenvolvimento sustentavel, direito, doencas, ecologia, educacao, educacao ambiental, educacaofisica, enfermagem, engenharia civil, ensino, epidemiologia, estado, farmacologia, filosofia, formacao, genero, geografia, historia, historia social, identidade, idoso, inflamacao, internet, leitura, letras, linguagem,

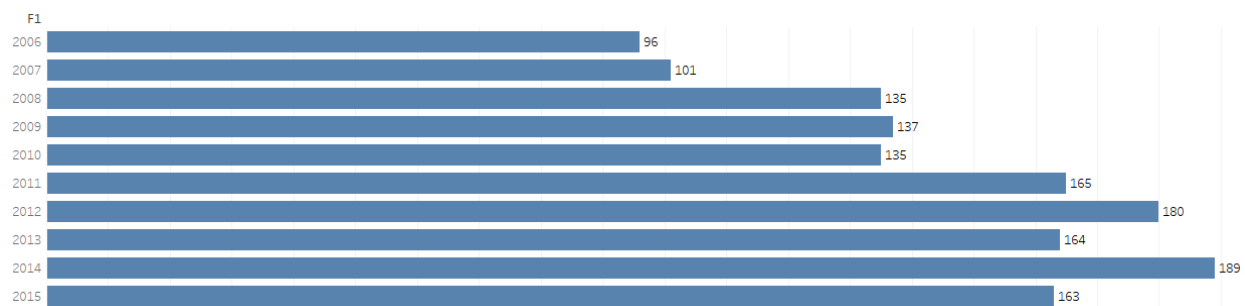
linguistica, literatura, marketing, medicina, meio ambiente, memoria, nutricao, obesidade, odontologia, plantas, psicanalise, psicologia, ratos, reproducao, rio grande do sul, santa catarina, saude publica, sociologia, solos, subjetividade, sustentabilidade, trabalho, tratamento e turismo. Cabe registrar que existe a possibilidade de haver trabalhos publicados nesses segmentos, mas seguramente não são muitos e estatisticamente não comprometerão os resultados da pesquisa.

E aqueles associados com os seguintes temas, resultando em 1.465 registros finais foram mantidos: administracao, aprendizagem, astronomia, biotecnologia, ciencia da computacao, controle, desenvolvimento, economia, energia, engenharia de materiais, engenharia de producao, engenharia eletrica, engenharia mecanica, fatores de risco, fisica, fisica nuclear, matematica, performance, planejamento, qualidade, quimica, sensoriamento remoto, tecnologia, tecnologia da informacao e tecnologia nuclear.

ANÁLISE DAS INFORMAÇÕES OBTIDAS

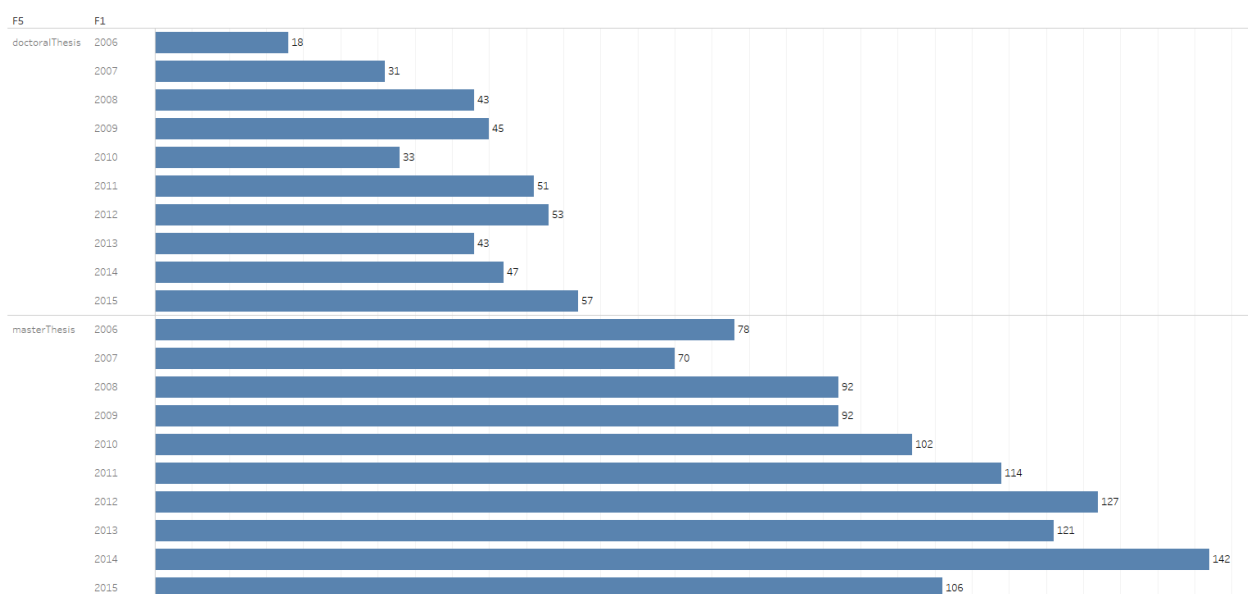
As análises apresentadas a seguir foram obtidas a partir do conjunto final de dados, 1.465 registros, conforme descrito anteriormente. O número de pesquisas por ano está apresentado no gráfico 1, sendo a média anual de pesquisas realizadas de aproximadamente 146,5 pesquisas por ano, com tendência de crescimento.

Gráfico 1 – Número de pesquisas por ano



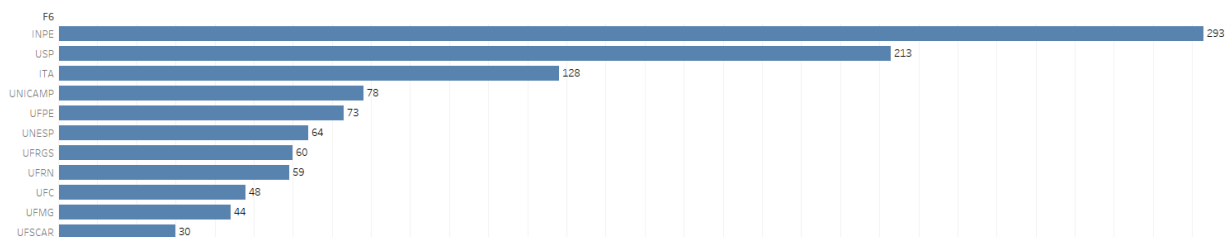
Fonte: Elaborado pelos autores.

Gráfico 2 – Número de pesquisas de mestrado e doutorado por ano



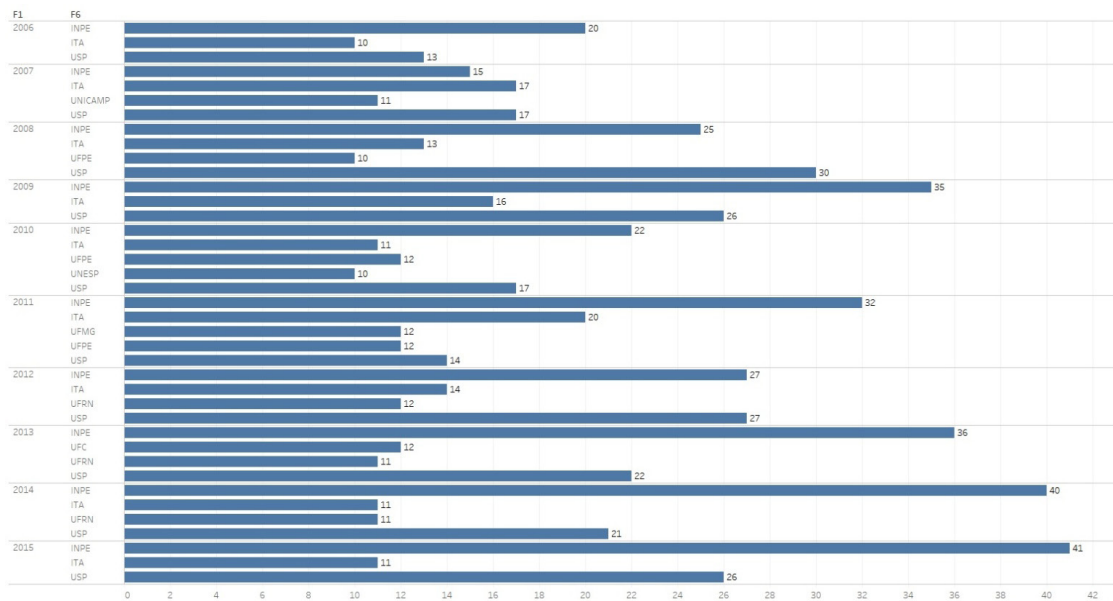
Fonte: Elaborado pelos autores, 2017.

Gráfico 3 – Ranking das instituições que desenvolveram acima de 30 pesquisas por ano



Fonte: Elaborado pelos autores, 2017.

Gráfico 4 – Número de pesquisas por ano e por instituição



Fonte: Elaborado pelos autores, 2017.

O gráfico 2 apresenta a divisão entre doutorado e mestrado do total de instituições, somando 421 doutorados e 1.044 mestrados concluídos.

O gráfico 3 mostra o ranking das instituições que desenvolveram acima de 30 pesquisas por ano no período considerado. As três instituições que apresentam o maior número de teses e dissertações defendidas, na área objeto deste trabalho, são o Instituto Nacional de Pesquisas Espaciais (Inpe), a Universidade de São Paulo (USP) e o Instituto Tecnológico da Aeronáutica (ITA). De fato, as três primeiras juntas representam 58% do total das pesquisas desenvolvidas e dessas três, o Inpe representa quase a metade do que é feito, 46%.

O gráfico 4 apresenta o número de pesquisas por instituição e por ano daquelas instituições que desenvolveram mais de 10 pesquisas no ano. Pode-se observar que, com exceção do ano de 2013, quando o ITA não aparece entre os primeiros, o Inpe, o ITA e a USP aparecem em todos os anos com mais de 10 pesquisas concluídas.

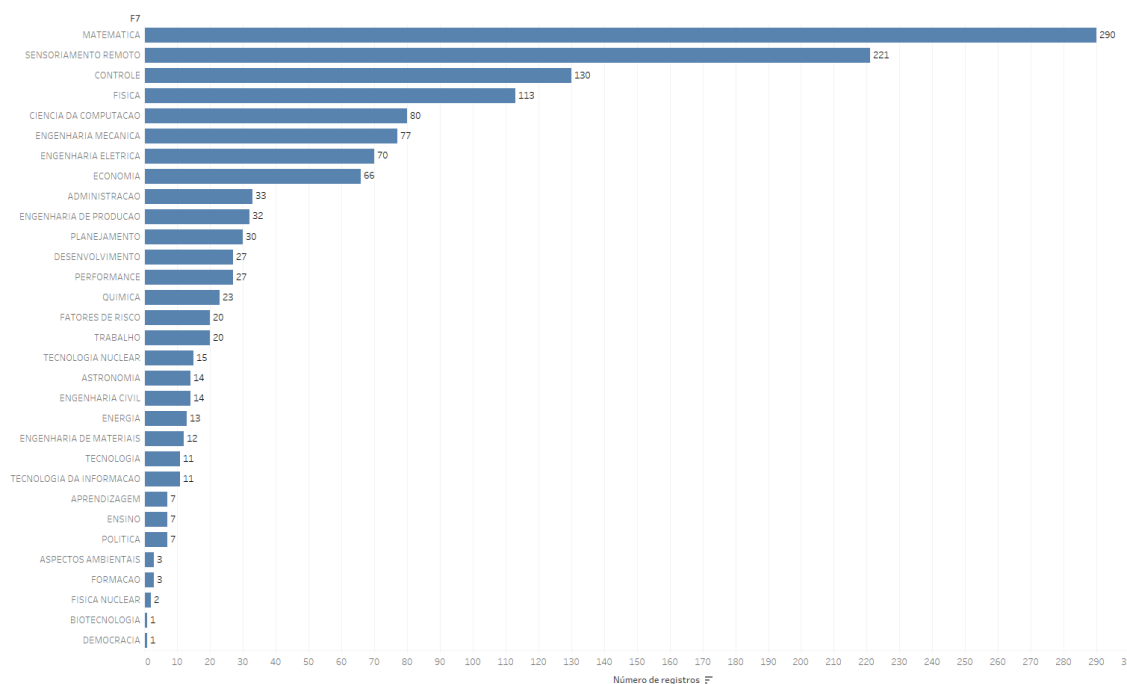
O gráfico 5 traz os principais descritores das três mais representativas instituições que atuam na pesquisa espacial, considerando o quantitativo de trabalhos desenvolvidos.

Aqui cabe destacar como tais descritores foram concebidos e suas principais diferenças com as palavras-chave. A última não obedece a nenhuma estrutura, é aleatória e retirada do texto a ser indexado em linguagem livre. Já o descritor obedece a rígido controle de sinônimos, significado e importância na classificação de determinado assunto. São organizados em estruturas hierárquicas, facilitando a pesquisa e a posterior busca e recuperação da informação.

Nessa pesquisa, os descritores foram elaborados em três fases: a primeira a partir do conjunto de palavras-chave mais frequentes em termos absolutos, isso é, independentemente da posição que ocupavam no sequenciamento das palavras-chave, cujo entendimento pode variar desde dar início com a grande área, como dar início com a área específica.

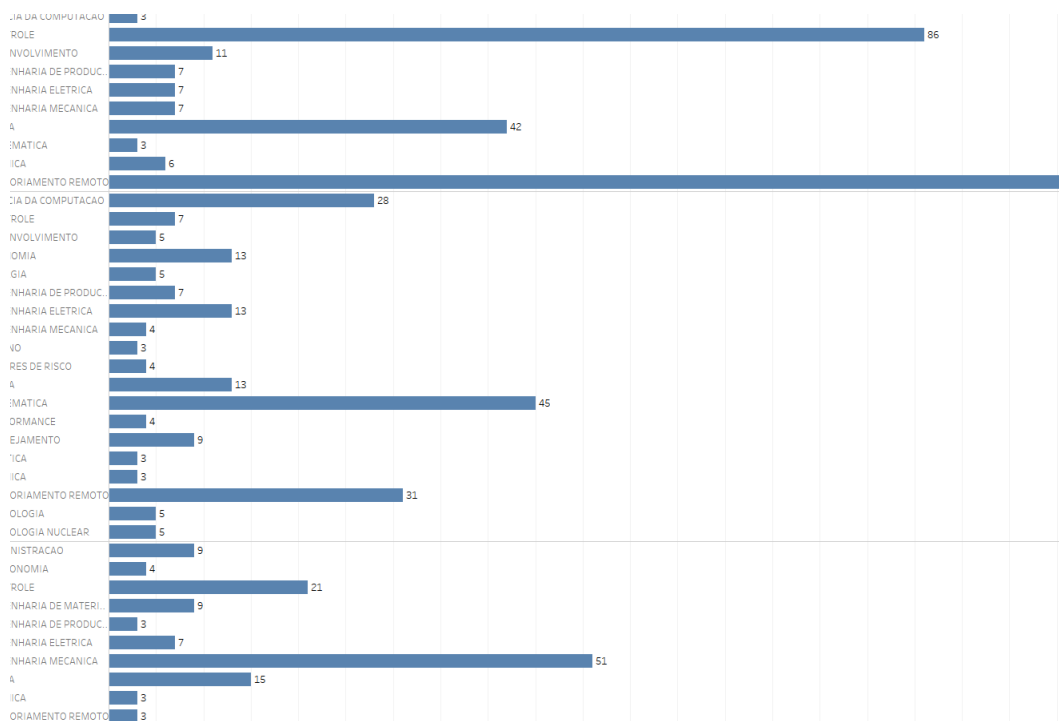
Em seguida, a avaliação dessas palavras-chave e como se relacionam hierarquicamente. E a terceira e última fase, a decisão de quais seriam os descritores da pesquisa espacial no primeiro nível hierárquico, isso é, aqueles que agrupam grandes áreas do conhecimento.

Gráfico 5 – Descritores selecionados e apresentados na ordem de ocorrência



Fonte: Elaborado pelos autores, 2017.

Gráfico 6 – Principais descritores por instituição



Fonte: Elaborado pelos autores, 2017.

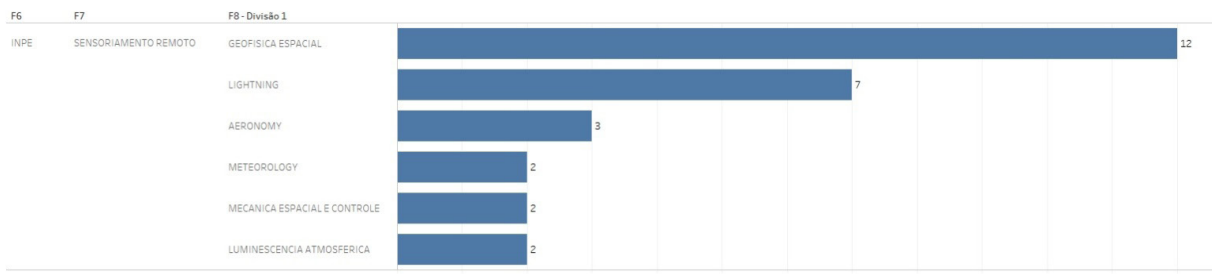
O gráfico 6 apresenta os principais descritores por instituição. Com base nesta informação, a área de sensoriamento remoto pode ser a de maior dedicação no Inpe, a de matemática pode ser a predominante na USP e a de Engenharia Mecânica pode ser o que caracteriza o ITA.

O gráfico 7 explora o tema sensoriamento remoto no Inpe. Interessante observar a relação do descritor com as palavras-chave ao qual ele está relacionado. No Inpe é assim que ele se relaciona com a primeira palavra-chave da pesquisa:

Cabe uma análise da relação do mesmo descritor, sensoriamento remoto, com as primeiras palavras-chave mais utilizadas na soma de todas as instituições.

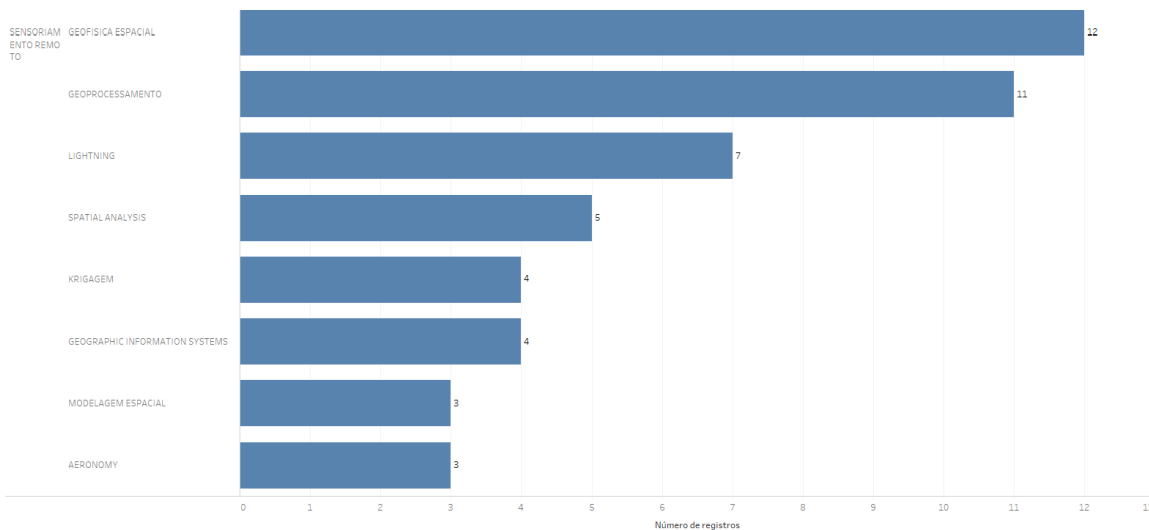
A análise das palavras-chave traduz o que se tem feito no Inpe no período analisado. A palavra-chave mais utilizada foi mecânica espacial e controle, cuja interação ocorre com as palavras-chave indicadas na tabela 1.

Gráfico 7 – Descritor sensoriamento remoto e sua relação com as primeiras palavras-chave mais utilizadas no Inpe



Fonte: Elaborado pelos autores, 2017

Gráfico 8 – Descritor sensoriamento remoto e sua relação com as primeiras palavras-chave mais utilizadas.



Fonte: Elaborado pelos autores, 2017

Tabela 1 – Relacionamento de palavras-chave no Inpe

F6	Palavra-chave 1	Palavra-chave 2	Palavra-chave 3		
INPE	MECANICA ESPACIAL E CONTROLE	ALGORITMOS DE IDENTIFICACAO DE E...	SENSOR DE ESTRELAS		
		AMBIGUITY RESOLUTION	CYCLE SLIPES		
		ATTITUDE ESTIMATION	UNSCENTED KALMAN FILTER		
		COLISAO	INCLINATION		
			SEPARACAO DE ESTAGIO		
		CONTROLE DE ATITUDE	ESTIMACAO		
			ESTRUTURA RIGIDO		
			SIMULADOR		
		CONTROLE DE PRECESSAO	SIMULACAO MATLAB SIMULINK		
		DECAIMENTO ORBITAL	ATMOSPHERIC FRAGMENTATION		
		DEVELOPMENT FLOW	CO SIMULATION EMULATION		
		ESTABILIDADE	TRIPLE SYSTEM		
		INPUT SHARING	STEP MOTOR		
		INTERPLANETARY SPACECRAFT	TRITON		
		LOG LTR CONTROL	LYAPUNOV STABILITY		
		LQR CONTROL DESIGN	CONTROLE H		
		PAINEL SOLAR	CONDUTANCE		
		PARAMETRIZACAO	ENCONTRO E ACOPLAMENTO ATRAC		
		PLANETARY MIGRATION	EVOLUCAO ORBITAL		
		PROBLEMA BI	LIPTICO INCLINADO PRECESSANDO		
		REACTION WHEEL	ATITUDES DE SATELITES		
			ELECTRONIC CONTROL		
			RODA DE REACAO		
		REAL TIME	INTERNAL SYNCHRONIZATON		
		TETRAEDRO	UMI		
		THERMAL DESIGN	PROJETO TERMICO		
		TOLERANCIA A FALHAS	FAULTS SENSORS AND ACTUATORS		
		TRAJECTORY	SIMULACAO		
		TRANS	NEPTUNIAN OBJECTS		
		UMI	MEMS		
		VELA SOLAR	SOLAR SAIL		
		WELCH	LYNCH		
		GEOFISICA ESPACIAL		AIRGLOW	IMAGING SPECTOMETER
				BRIGHT POINTS	PONTOS BRILHANTES
IONOSPHERE	EQUATORIAL ANOMALY				
IONOSPHERIC ELECTRONIC TEMPERAT.	AERO LUMINESCENT EMISSIONS				
MESOSFERA	AIRGLOW				
	METEOR RADAR				
MIDDLE ATMOSPHERE	SABER				
ONDAS DE MUITO BAIXA FREQUENCIA	BAIXA IONOSFERA				
PARAMETRIZED MODELS	FINITE FOURIER SERIES				
RELACOES SOL	TERRA				
SPORADICE	LAYERS				
TEMPESTADES MAGNETICAS	PRECIPITACAO DE PARTICULAS				
	VENTOS TERMOSEFERICOS				
VARIACAO SAZONAL	METHANE				

(Continua)

Tabela 1 – Relacionamento de palavras-chave no Inpe

F6	Palavra-chave 1	Palavra-chave 2	Palavra-chave 3
INPE	LIGHTNING	ESPECTROSCOPIA SEM FENDA	DESCARGAS ATMOSFERICAS
		GEOFISICA ESPACIAL	LIDER BI
			LIGHTNING LOCATION SYSTEMS
			THESIS
		GEOFISICA ESPACIAL CIENCIAS ATMOSFERICAS	DVAR WRF MESOSCALE MODEL
PMM		HIGH	SPEED CAMERAS
		LIDER	ATMOSPHERIC DISCHARGE
		ANALISE DE MISSAO	PROPRIEDADES DE MASSA
		MECANICA ESPACIAL E CONTROLE	AVIONICA MODULAR INTEGRADA
		RECONFIGURABLE CONTROL	MECANICA ESPACIAL E CONTROLE
SYSTEMS ENGINEERING		SISTEMAS DE CONTROLE	SUBSISTEMA DE CONTROLE DE ATIT
		SUBSISTEMA PROPULSIVO	SIMULACAO
		AERONAUTIC NAVIGATION	NVEGACAO AERONAUTICA
		ARCHITECTURE DEVELOPMENT	ENGENHARIA DE SISTEMAS
		E	ENGINEERING
MAGNETOSFERA		EFEITOS DE RADIACOES IONIZANTES	ENGENHARIA DE SISTEMAS
		ENGENHARIA SIMULTANEA	ENGENHARIA DE SISTEMAS
		GEOFISICA ESPACIAL	SPACE ELECTRODYNAMICS
		MAGNETOPOUSE	MAGNETOPOUSA
		NUMERICAL SIMULATION	MHD
MANOBRAS ORBITAIS		TRANSIENT EVENT	MAGNETIC FIELD RECONNECTION
		MECANICA ESPACIAL E CONTROLE	CONTROLE DE TRAJETORIA
			TRANSFERENCIA ORBITAL
		MINOR BODIES	MECANICA ESPACIAL E CONTROLE
		SATELITES ARTIFICIAIS	ORBIT DETERMINATION
AERONOMY		ANALISE ESPECTRAL	SOLAR ACTIVITY
		ATMOSFERA TERRESTRE	TEMPESTADES GEOMAGNETICAS
		GEOMAGNETIC ACTIVITY	ATIVIDADE SOLAR
		PREVISAO DE TEMPESTADE GEOMAG	GEOFISICA ESPACIAL
		GEOFISICA ESPACIAL	ATMOSPHERIC ELECTRIC FIELDS
CONTROLE			ATMOSPHERIC TEMPERATURE
		PROCESSAMENTO DE IMAGENS	GEOFISICA ESPACIAL
		DINAMICA ORBITAL	MECANICA ESPACIAL E CONTROLE
		MECANICA ESPACIAL E CONTROLE	SYSTEMS SIMULATION
		SISTEMAS DE CONTROLE	OTIMIZACAO
COHERENT STRUCTURES		EDGE OF CHAOS	GEOFISICA ESPACIAL E CIENCIAS DO A...
		SYNCHRONIZATION	GEOFISICA ESPACIAL
CORRENTE CONTINUA		GEOFISICA ESPACIAL	RAIOS POSITIVOS
		LIGHTNING	GEOFISICA ESPACIAL E CIENCIAS ATM...
DINAMICA ORBITAL		MECANICA ESPACIAL E CONTROLE	CONTROLE DE TRAJETORIA
		RAZAO DH	DH RATIO
LUMINESCENCIAATMOSFERICA		BOLHAS DE PLASMA	GEOFISICA ESPACIAL
		PERKINS INSTABILITY	INSTABILIDADE PERKINS
METEOROLOGY		DEPARTMENT OF AIR SPACE CONTROL	DECEA
		DIAS DE TEMPESTADE	GEOFISICA ESPACIAL CIENCIAS ATM.

(Continua)

Tabela 1 – Relacionamento de palavras-chave no Inpe

(Conclusão)

F6	Palavra-chave 1	Palavra-chave 2	Palavra-chave 3
INPE	OTIMIZACAO	MECANICA ESPACIAL E CONTROLE	ASTRODINAMICA
	SATELITES ARTIFICIAIS	ENVIRONMENTAL TESTS	M
		SPACE TECHNOLOGY AND ENGINEERING	VERIFICACAO
	SPACE PLASMAS	GEOFISICA ESPACIAL	ESTABILIDADE
			SPACE ELECTRODYNAMICS
			TEMPESTADES MAGNETICAS

Fonte: Elaborado pelos autores, 2017

CONSIDERAÇÕES FINAIS

A obtenção de conhecimento sobre a pesquisa científica e tecnológica brasileira na área espacial a partir da análise automática de informação na Biblioteca Digital Brasileira de Teses e Dissertações é importante para toda a cadeia de produção do conhecimento nesse campo. Pode contribuir tanto para o planejamento estratégico de políticas de ensino e pesquisa como para a organização de bancas examinadoras. Pode tanto oferecer o perfil dos programas de pós-graduação que atuam na área como o perfil de determinado segmento da pesquisa. Enfim, pode-se assegurar que é um manancial inesgotável de fonte de informação sobre a pesquisa espacial brasileira, mesmo sendo estes resultados iniciais e carecendo de aprofundamento em vários aspectos, sobretudo na análise de pesquisadores envolvidos e de como se dá a evolução dos vários temas de pesquisa ao longo dos anos.

A criação de conjuntos de dados com vistas à análise de informação foi possível graças à identificação de softwares livres responsáveis pela preparação (incluindo limpeza), extração, tratamento (incluindo manipulação), processamento (incluindo visualização), interpretação e análise de dados, processo conhecido como Descoberta de Conhecimento em Bases de Dados ou em inglês Knowledge Discovery in Databases (KDD).

Optou-se por excluir a longtail, isto é, os inúmeros registros que: (i) por trazerem temas emergentes, (ii) erros de digitação ou (iii) falta de padronização na entrada de dados, geram uma relação significativamente extensa, porém impossível de figurar estatisticamente.

É possível antever que o aprendizado obtido neste trabalho fornece ampla possibilidade para a investigação das atividades de pesquisa em outras áreas de conhecimento, incluindo a proposta de áreas de conhecimento e suas subáreas a partir do que efetivamente se está realizando nos programas de pós-graduação do país.

REFERÊNCIAS

- ALVARES, L. *Descoberta de conhecimento na Biblioteca Digital Brasileira de Teses e Dissertação* (Relatório de Estágio Pós-Doutoral), Castellón de la Plana, Comunidade Valenciana, Espanha. Escola Superior de Tecnologia i Ciències Experimentalsdo Departament de Llenguatges i Sistemes InformàticsUniversitat Jaume I, UJI, 2017.
- ANAYA-SÁNCHEZ, H.; PONS-PORRATA, A.; BERLANGA, R. A. document clustering algorithm for discovering and describing topics. *Pattern Recognition Letters*, v.31, n.6, p.502-510, 2010.
- BRASIL. *Decreto nº 51.133, de 3 de agosto de 1961*- cria o grupo de organização da comissão nacional de estudos espaciais e dá outras providências. Brasília, 1961. Disponível em: <<http://legis.senado.gov.br/legislacao/ListaPublicacoes.action?id=230703>>. Acesso em: 29 ago. 2017.
- CÂMARA, G. A pesquisa espacial no Brasil: 50 anos de Inpe (1961-2011). *Revista USP*, n.89, mar./maio 2011, p. 234-243. Disponível em: <<http://rusp.scielo.br/pdf/rusp/n89/16.pdf>>. Acesso em: 05 set. 2017.
- HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Series in Data Management Systems, 2011.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). *Plano diretor do INPE 2016-2019*. São José dos Campos, 2016. 88 p. Disponível em: <http://www.inpe.br/noticias/arquivos/pdf/Plano_Diretor_2016-2019.pdf>. Acesso em: 20 ago. 2017.

_____. *Relatório de gestão do exercício de 2016*. São José dos Campos, 2017. 202p Disponível em: <http://www.inpe.br/gestao/relatorio_gestao.php>. Acesso em: 28 ago. 2017.

SANCHO, R. Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología: revisión bibliográfica. *Revista Española de Documentación Científica*, v.13, n.3-4, p.842-65, 1990.

TRISKA, R.; CAFÉ, L. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. *Revista Ciência da Informação*, v.30, n.3, p.92-96, 2001.

VANZ, S.A. de. S.; STUMPF, I.R.C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. *Informação & Sociedade: Estudos*, v.20, n.2, p.67-75, 2010.

VELHO, L. A avaliação do desempenho científico. *Cadernos USP*, n.1, p.22-40, 1986.