UNIVERSIDADE DO SUL DE SANTA CATARINA
PAULO HENRIQUE GUIMARÃES FIDENCIO
VINÍCIUS ROGGIA GOMES

RECUPERAÇÃO DE INFORMAÇÕES TEXTUAIS E MULTIMÍDIA, UTILIZANDO EXPANSÃO DE CONSULTA A RECURSOS DA WEB 2.0

Palhoça
2011

PAULO HENRIQUE GUIMARÃES FIDENCIO
VINÍCIUS ROGGIA GOMES

RECUPERAÇÃO DE INFORMAÇÕES TEXTUAIS E MULTIMÍDIA, UTILIZANDO EXPANSÃO DE CONSULTA A RECURSOS DA WEB 2.0

Trabalho de Conclusão de Curso, apresentado ao Curso de Graduação em Ciência da Computação, na Universidade do Sul de Santa Catarina, - como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Flávio Ceci, M. Eng.

Palhoça
2011

PAULO HENRIQUE GUIMARÃES FIDENCIO
VINÍCIUS ROGGIA GOMES

RECUPERAÇÃO DE INFORMAÇÕES TEXTUAIS E MULTIMÍDIA, UTILIZANDO EXPANSÃO DE CONSULTA A RECURSOS DA WEB 2.0

Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Ciência da Computação, e aprovado, em sua forma final, pelo Curso de Graduação em Ciência da Computação, na Universidade do Sul de Santa Catarina.

Palhoça, 17 de novembro de 2011.

_____
Professor e orientador Flávio Ceci, M. Eng.
Universidade do Sul de Santa Catarina

_____
Prof. Aran Bey Tcholakian Morales, Dr. Eng.
Universidade do Sul de Santa Catarina

_____
Prof. Ricardo Villarroel Dávalos, Dr. Eng.
Universidade do Sul de Santa Catarina

Dedico este trabalho a meus avós e padrinhos, Affonso Paulo Guimarães e Regina Emília Guimarães, à minha querida bisavó Vitória Aluiza Schreiber Guimarães (in memoriam) e, principalmente, a meu querido bisavô, Jorge Paulo Guimarães (in memoriam).

Paulo Henrique Guimarães Fidencio

Dedico este trabalho a meus pais, João e Rosana, que foram os maiores incentivadores à minha formação de nível superior, - tornando-se, assim, responsáveis por esta conquista.

Vinícius Roggia Gomes

AGRADECIMENTOS

Paulo Henrique Guimarães Fidencio agradece

Este trabalho foi feito, com dedicação e força de vontade, por meu amigo Vinícius e por mim; porém, muitas pessoas participaram dele, de uma forma direta ou indireta. Nesse espaço, - quero agradecer a todas as pessoas que me ajudaram a concluí-lo.

Primeiramente, agradeço a meu colega e amigo Vinícius, que me vem acompanhando desde o início da faculdade, em meus trabalhos acadêmicos, e mesmo na área profissional. Tenho certeza que este trabalho não estaria concluído sem a ajuda e o empenho do Vinícius.

Meu agradecimento a meu amigo e orientador Flávio, excelente profissional, sempre apresentando novas ideias e diferentes caminhos, auxiliando-me na elaboração deste trabalho, e incentivando meu crescimento, não só na área profissional, mas em minha vida.

Agradeço, penhoradamente, a meu querido bisavô, Jorge Paulo, conhecido como vô Fone, por ter cuidado de mim, quando criança, sempre me mantendo forte e animado, nos momentos bons e ruins. Figura que guardo, carinhosamente, como pai, - preocupando-se com meu bem estar, até os últimos momentos de sua vida.

Não poderia deixar de consignar meu agradecimento à minha querida bisavó Vitória, que me ensinou a seguir, sempre, o caminho da verdade, da justiça e do bem. Deixou para mim a sua imagem de mãe.

Meu muito obrigado a meu avô Affonso, que, sempre, me inspira a prosseguir, na conquista de um ideal mais alto. É uma pessoa que admiro e tenho como exemplo, por ser um homem batalhador e responsável, tendo assumido diferentes papéis, em toda a sua vida, como professor, escritor, advogado e um pai carinhoso e dedicado.

Agradecimentos, também, à minha avó Regina, tenho-a como mãe, jamais poupando paciência e dedicação para com a minha pessoa. Mulher forte que me apoia, visando, sempre, o meu sucesso e a minha felicidade.

Obrigado à minha namorada, Jéssica Luana, pelo carinho, atenção e paciência, nos dias em que trabalhei, no desenvolvimento desse trabalho, não tendo condições de visitá-la.

Gratidão a meus pais, Orli e Patrícia, que, mesmo não estando presentes no meu dia a dia, me apoiaram na escolha do curso de Ciência da Computação.

Agradeço, também, a todos os professores e amigos do curso de Ciência da Computação e Sistemas da Informação, pois todos fizeram parte e contribuíram, de uma maneira ou de outra, para a minha formação. Agradecimentos pelos conselhos, pelos conhecimentos e pela experiência que compartilharam comigo.

Vinícius Roggia Gomes agradece

Agradeço a meus pais, João Pereira Gomes Neto e Rosana do Carmo Roggia Gomes, pelo amor, carinho e exemplo de vida.

Obrigado a meu irmão, Fernando Roggia Gomes, pela amizade incondicional desde nossa infância até os dias de hoje.

Meus agradecimentos à minha namorada, Eduarda Lencina Mattos, que esteve ao meu lado e me apoiou durante toda esta jornada, mesmo naqueles dias de trabalho e estudo, em que não pudemos ficar juntos.

Gratidão aos amigos e demais familiares que influenciaram, de forma direta e indireta, na formação da pessoa que sou hoje.

Ao meu amigo e parceiro, neste trabalho, Paulo Henrique Guimarães Fidencio, meus agradecimentos, pela amizade, paciência, atenção e colaboração, nesta empreitada, que nos custou um ano de dedicação.

Obrigado ao amigo e orientador, Flávio Ceci, que colocou à minha disposição sua atenção e tempo, possibilitando o desenvolvimento deste trabalho, auxiliando e orientando, através da disponibilização de materiais importantes, através da troca de centenas de e-mails, nas reuniões na faculdade e em sua própria casa, - sempre visando melhorar o trabalho.

E, também, meus agradecimentos penhorados aos professores do curso de Ciência da Computação, da Universidade do Sul de Santa Catarina, pela dedicação, profissionalismo, e, em especial, ao Prof. Dr. Engº. Aran Bey Tcholakian e ao Prof. Dr. Engº. Ricardo Vilarroel Dávalos, por terem aceito o convite para compor a banca examinadora.

“A imaginação é mais importante que a ciência, porque a ciência é limitada, ao passo que a imaginação abrange o mundo inteiro.” (Albert Einstein).

RESUMO

Atualmente, nota-se o grande crescimento de informações, geradas no meio digital, sendo que estas informações apresentam-se de diferentes maneiras, podendo ser um arquivo em formato de texto, som, imagem ou vídeo. O grande desafio é encontrar a informação que nos interessa, de uma maneira rápida e fácil, pois, ao realizar uma busca no ambiente web, a maioria dos resultados, normalmente, não tem relação alguma com o que estamos procurando. Tendo como base esse cenário, o objetivo desse trabalho é mostrar um sistema que realize buscas em documentos textuais. Os resultados são os documentos que mais se relacionam com o termo da busca, além de retornar imagens, vídeos e sites que se relacionam com o tema dos documentos encontrados. Para a validação do sistema, foi criado um protótipo, tendo como foco um caso de uso, envolvendo o meio acadêmico. O ambiente se caracteriza pelos perfis de alunos e professores, onde os professores publicam artigos, e alunos os recuperam, através da busca. A validação do protótipo foi feita através de uma pesquisa com usuários de diferentes perfis. O resultado foi satisfatório, sendo que os documentos publicados foram recuperados, trazendo, na maioria das vezes, sites, imagens e vídeos que se relacionam com o documento retornado.

Palavras-chave: Recuperação de Informação, WEB 2.0, Expansão de Buscas.

ABSTRACT

Nowadays, there is the rapid growth of information generated in the digital environment and this information presents in different ways and can be a file in text format, sound, image or video. The great challenge is finding the information that interests us in a quick and easy way because most of the results when we usually performing a search on the web environment have no relation to with what we are looking for. Based on this scenario, the objective of this study is show a system to conduct searches in text documents. The results are the documents that most closely related to the search term in addition to return images, videos and sites that relate to the theme of the documents found. A prototype was created in order to validate the system focusing on one use case, involving academia. The environment is characterized by the profiles of students and teachers, where teacher publish articles and students through the search recover these articles. The validation of the prototype was done through a survey with users of different profiles. The result was satisfactory being that the published documents were recovered, bringing the most part, websites, images and videos that relate to the document returned.

Keywords: Information Retrieval, WEB 2.0, Search Expansion.

LISTA DE ILUSTRAÇÕES

SUMÁRIO

1 INTRODUÇÃO

Tendo em vista o crescimento de dados de diferentes tipos, no meio digital, e o grande problema para se encontrar informações de nosso interesse, dentro desse meio, - é necessário o uso de tecnologias que facilitem uma busca mais inteligente, trazendo resultados corretos, ou que se relacionem diretamente com o termo da busca.

De acordo com Santarém e Vidotti (2011), a Recuperação de Informação (RI) tem sido muito discutida na Ciência da Informação, e a busca por informação de qualidade com a necessidade do usuário se tornou pesquisa constante.

1.1 PROBLEMA DE PESQUISA

Dada a grande quantidade de informações disponíveis em documentos eletrônicos, espalhados nas organizações e pela Web, - é despendido um grande tempo para encontrar as informações necessárias. A falta de interpretadores inteligentes, nos sistemas de recuperação de informação, dificulta o encontro com uma solução mais próxima do ideal, fazendo com que, muitas vezes, seja despendido tempo à procura de uma resposta correta. Assim sendo, a maioria dos resultados são desnecessários, ou não se relacionam com os dados da busca.

Dennis, Bruza e McArthur (tradução nossa, 2002) citam uma pesquisa, realizada no primeiro semestre de 2000, com 33.000 usuários da Web, onde eles responderam à pergunta: "Quantas vezes você encontra o que procura?". O resultado foi que quase 60% dos pesquisados responderam que “a maior parte das vezes” encontram o que procuram. Já 21% afirmam que sempre encontram. E 2.6% nunca encontram.

Rijsbergen (1979, apud SEIBEL 2007) e Yates et al. (1999, apud SEIBEL 2007) afirmam que, “os métodos de recuperação de informação tradicionais se baseiam essencialmente na contagem da frequência em que as palavras aparecem em um documento; sem apresentar soluções para que o conteúdo semântico do discurso seja interpretado”. Por não processarem o documento, adequadamente, podem-se perder importantes informações,

“retornando” assim resultados não relevantes o que acaba despendendo mais tempo do usuário.

Segundo Wives e Loh (1998), muitas das informações disponíveis para acesso rápido e fácil não estão em formatos que possam ser tratados por meios computacionais (imagens, textos, vídeos, gráficos). Wives e Loh (1996 apud CHEN, 1998) citam que 80% das informações que uma empresa utiliza não estão armazenadas em banco de Dados na forma de números e caracteres.

Batista e Schwabe (2009) asseveram que é difícil o desenvolvimento de aplicações que necessitam capturar e manipular informações, diretamente, do conteúdo digital, como, por exemplo, a busca de vídeos, sem utilizar seus metadados descritivos (trechos de informação textual associado aos vídeos).

Outro problema que se encontra na busca de informações é com relação ao próprio usuário. Este, muitas vezes, utiliza argumentos que se relacionam com o seu interesse, de forma indireta, como, verbi gratia, quando se busca o melhor lutador de boxe da atualidade, desconhecendo seu nome, ou os campeonatos que ganhou.

Têm-se também a Web 2.0, conhecida como a web colaborativa, esta é conceituada como o momento em que as pessoas pararam de apenas consumir na internet para também contribuir com conteúdo. Esse conteúdo podendo ser de qualquer tipo, desde texto até vídeo. Encontra-se ai outro problema, como encontrar o conteúdo que é gerado a todo o momento?

A questão é como se pode recuperar documentos textuais relevantes, a partir de buscas, e sugerir novos conteúdos, a fim de acarretar mais valor ao resultado apresentado?

1.2 OBJETIVOS

Esta seção é reservada a apresentar os objetivos gerais deste projeto, bem como os seus objetivos específicos.

1.2.1 Objetivo geral

Desenvolver um sistema de busca em documentos textuais completos que expanda o resultado a partir de recursos da Web 2.0, acrescentando conteúdos multimídia.

1.2.2 Objetivos específicos

Os objetivos específicos são:

a) Construir um protótipo para web que irá cadastrar documentos para, posteriormente, serem recuperados;
b) Desenvolver um módulo para realizar a busca, através de um SRI;
c) Estudar e apresentar extensão de consulta, com o objetivo de expandir os resultados e obter informações multimídia (vídeos, imagens e textos) relevantes relacionados com os resultados;
d) Validar o protótipo a partir de um estudo de caso.

1.3 JUSTIFICATIVA

De acordo com Murugesan (2007), a Web 2.0 transforma a Web tradicional, ou Web 1.0, onde os usuários são simples consumidores da informação, em uma plataforma social para trocas de informações, através da colaboração. Nesta nova versão da Web, os usuários podem se encontrar, colaborar e interagir, através de aplicações para criar e compartilhar o conhecimento.

O trabalho de Jovanovic et al. (2008) mostra que, através da união da Web semântica e da Web Social, é possível criar e compartilhar conteúdo de forma colaborativa e, automaticamente, representar esta informação de forma explícita e com semântica para que os

computadores possam compreendê-la e utilizá-la. A Web que une as tecnologias da Web Semântica e da Web 2.0 é conhecida como Web 3.0.

Com o passar do tempo houve um grande crescimento de documentos que se espalharam pela Web, sendo necessário criar algo para "facilitar" a forma de se encontrar as informações na rede, então surgiram os diretórios, com objetivo indexar a "Web" manualmente, e na sequência foram desenvolvidos os sistemas de buscas automáticos, como exemplo, o Google.

Ramalho e Robin (2004) explicam que uma das causas da ineficiência das buscas é o fato de que documentos e consultas são indexados por palavras-chave, independentemente do seu significado. Assim, ao comparar os termos do documento e da consulta, podem não ser recuperados os documentos relevantes. Para se resolver esse problema, tem-se utilizado da técnica de expansão de consulta.

Segundo Isotani (2008), na Web 3.0, uma nova classe de sistemas, os chamados sistemas de conhecimento coletivo, está em fase de desenvolvimento. Estes sistemas serão capazes de auxiliar na produção de conhecimento coletivo através de análise da contribuição colaborativa humana.

Com a necessidade de utilização de sistemas de busca mais eficazes é necessário antes disso analisar as formas de se organizar esse conteúdo de uma forma inteligente. Essa pesquisa mostra o desenvolvimento de um sistema de recuperação de informação (RI). O objetivo do sistema é armazenar os arquivos, indexar, buscar e por fim, exibir um resultado mais próximo do esperado pelo usuário, além de exibir informações multimídia através de extensão de consulta a recursos da Web 2.0.

1.4 ESTRUTURA DO TRABALHO

Este trabalho esta estruturado da seguinte forma: o primeiro capítulo apresenta uma introdução ao assunto, descrevendo o problema de pesquisa, o objetivo geral, os objetivos específicos e a justificativa.

O segundo capítulo trata do referencial bibliográfico, mostrando algumas opiniões e conceitos em relação à recuperação de informações, tendo como tópicos gerais: recuperação de informação, anotação de documentos e recuperação de informação multimídia.

O terceiro capítulo apresenta o método de pesquisa, caracterizando o tipo de pesquisa do trabalho, mostrando as próximas etapas para a conclusão do mesmo, o esquema da solução proposta e as delimitações.

O quarto capítulo mostra o projeto de solução, definindo as técnicas e a metodologia utilizada. Será abortado nesse capítulo o conceito do Iconix, da UML e da Orientação a Objetos, assim como será apresentada a modelagem do sistema seguindo a metodologia de desenvolvimento ágil do Iconix.

O quinto capítulo apresenta o modelo proposto, descrevendo o protótipo desenvolvido com base no estudo de caso escolhido e os resultados encontrados no desenvolvimento do problema.

Por fim, o sexto e último capítulo apresenta as conclusões, assim como idéias e sugestões para futuros trabalho.

2 REFERENCIAL BIBLIOGRÁFICO

O seguinte capítulo descreve alguns elementos necessários para o entendimento deste trabalho. São abordados, primeiramente, conceitos básicos, relacionados à Recuperação de Informações: extração de informação, stopwords, stemming, tesauros, transformação de dados e indexação. Em seguida, foram levantadas informações, referentes à anotação de documentos, descrevendo cada tipo de anotação, ressaltando alguns pontos positivos e negativos. Para concluir, são descritos tópicos, com relação à Recuperação de Informação Multimídia: expansão de consulta, MPEG-7 e extensão de busca para web.

2.1 RECUPERAÇÃO DE INFORMAÇÃO

Conforme Leite (2009, p.7), "o escopo da recuperação de informação pode considerar os recursos de uma área de conhecimento específica como, por exemplo, agricultura, artes, leis e saúde, até os recursos de toda a WWW (World Wide Web)".

Ceci (2010 p.35) descreve o principal objetivo de RI, que se resume a:

> Tornar o acesso mais fácil aos documentos de maior relevância conforme a necessidade de informação do usuário. Essa necessidade normalmente é simbolizada por meio de um busca por palavra-chave. A recuperação de informação, nesse contexto, consiste basicamente na determinação de quais documentos de uma coleção contêm as palavras-chave da consulta realizada pelo usuário. A dificuldade está não somente em extrair a informação, mas também em decidir a sua relevância.

Para Beppler (2002, p.11), a ideia da RI é simples e objetiva: tendo um armazenamento de informações relevantes, devemos recuperar apenas a informação desejada, porém o foco é a maneira como será feito o armazenamento, pois existem muitos documentos que são difíceis de serem armazenados de forma correta, causando problemas de interpretação no formato digital, um exemplo, são os textos em linguagem natural. Para um ser humano é fácil verificar em um documento o que seria relevante, porém para um computador fazer esta interpretação é um problema, pois ele terá de saber que tipo de informação possui cada documento, o que nos leva a ter que criar um modelo de decisões relevantes que possam ser determinadas de maneira dinâmica.

Segundo Ferneda (2003, p. 12) “os primeiros sistemas de recuperação de informação baseavam-se na contagem de frequência das palavras do texto e na eliminação de palavras reconhecidamente de pouca relevância”.

2.1.1 Extração de informação

Pal (tradução nossa, 2002) define Extração de Informação (EI) como o procedimento que identifica fragmentos específicos de um único documento, e que constituem o núcleo de seu conteúdo semântico.

Para Silva (2003, p. 20) a EI tem como objetivo analisar um documento e extrair dados relevantes relacionados a um determinado tema, sendo posteriormente convertido para uma estrutura tabular. Esta estrutura irá permitir uma melhor visualização do conteúdo, tanto para usuários como também para aplicações que irão utilizar essas informações.

Segundo Bittencourt (2002), a EI apresenta o seguinte objetivo:

> Extrair de um texto dados específicos que caracterizem o contexto abordado baseando-se num conjunto de atributos pré-definidos que representem a informação a ser extraída. Estas características podem ser exploradas para prover a usuários informações menos ruidosas e mais focadas em seus reais interesses.

Tendo em vista o contexto de Web, a EI surgiu como uma forma de aprimorar e organizar os resultados oferecidos pela Recuperação de Informações (RI), extraindo as informações relevantes de algum contexto. (SILVA, 2003, p.2).

Não se pode esquecer que Extração de Informação e Recuperação de Informação possuem conceitos distintos. Em vez de extrair a informação, o objetivo da RI é selecionar um subconjunto relevante de documentos de uma maior coleção de dados, com base em uma consulta realizada por um usuário. (EIKVIL, 1999, p.5). Enquanto, a EI é utilizada para filtrar esse resultado da consulta realizada pela RI. (SILVA, 2003, p.23).

Para realizar a tarefa de extrair as informações de um documento, a EI utiliza métodos que, geralmente, envolvem a escrita de códigos, conhecidos como *wrappers*, e que mapeiam o documento para algum modelo de representação do conhecimento. (MARINHO, 2003).

> Os *wrappers* extraem informações de textos baseando-se em delimitadores como palavras-chave, *tokens*, posição no texto, etc. Um problema desta abordagem é que o domínio tratado é bastante restrito, exigindo dos textos uma formatação padrão de

> extração. Outro problema é que os dados extraídos não recebem uma conotação semântica. (BITTENCOURT, 2002).

2.1.2 Stopwords

Stopwords são palavras que não possuem muita importância, quando se esta analisando um texto. Segundo Rizzi (1983 apud SALTON, 2000), "estas palavras são bastante frequentes e sua eliminação pode prover uma redução de 40 a 50% dos textos dos documentos a serem analisados".

Para um melhor entendimento do conceito de *stopwords* segue a definição, proposta por Moraes (2008, p.4):

> As palavras que não são passíveis de serem representantes de alguma categoria são conhecidas como *stopwords* ou palavras negativas e podem ser representadas por artigos, pronomes, preposições, advérbios e outras palavras que se apresentem com elevada ou baixa frequência nos textos.

Com o objetivo de comprimir o texto para análise, é feita a eliminação das *stopwords*, pois desta forma é reduzido o número de palavras a serem analisadas, além de que se reduz também o número de palavras que serão armazenadas na base de dados. (DIAS, 2004, p.28). Contudo Gonzalez (2003) alerta que: "com tal eliminação, corre-se, entretanto, o risco de perder a estrutura composicional de expressões".

2.1.3 Stemming

Para Geraldo (2009, p.14) *stemming* "consiste no processo de redução de uma palavra a sua raiz morfológica". Para realizar esse processo são aplicados algoritmos chamados de *stemmers*, que auxiliam Sistemas de Recuperação de Informações (SRI) a melhorar seu desempenho. (COELHO, 2007, p.13).

Martins (2004, p.4) explica que, algoritmos de *stemming* consistem em uma normalização linguística, onde palavras que derivam de um termo são reduzidas a outra palavra, através da eliminação de prefixos, sufixos ou até mesmo a transformação de um

verbo para sua forma no infinitivo, o resultado será uma forma comum para o termo, o qual é denominado *stem* (raiz). "O *stem* resultante não precisa ser uma palavra válida do idioma, porém deve conter o significado base original de suas palavras". (FLORES, 2009, p. 10). "Portanto, um algoritmo de *stemming* é fortemente dependente do idioma no qual os documentos estão escritos". (MARTINS, 2004, p.4).

Frakes e Yates (tradução nossa, 1992, p.131) afirmam que uma técnica para melhorar o desempenho da RI é encontrar derivações morfológicas dos termos utilizados pelos pesquisadores no momento da busca. Se, por exemplo, um usuário inserir um termo (ex: mentiroso), como parte de uma consulta, é provável que ele, também, esteja interessado em derivações ou palavras similares ao termo pesquisado (ex: mentira ou mentir). É usado, também, o termo conflação, no sentido de fundir ou combinar termos, resultando em um termo mais geral, que é utilizado para relacionar as derivações morfológicas. Essa conflação pode ser realizada de maneira manual, utilizando algum tipo de expressão regular, ou automática, através dos *stemmers*.

Segundo Frakes e Yates (tradução nossa, 1992, p.131), *Stemmings,* também, são usados na RI para reduzir o tamanho dos índices de arquivos. Um único *stem,* geralmente, corresponde a vários outros temos. Armazenando os *stems* em vez de termos, podemos ter mais de 50% de aproveitamento do espaço de armazenamento.

Apesar dos *stemmings* trazerem benefícios para os sistemas de RI, Coelho (2007, p.15) mostra que:

> O uso de *stemming* traz também algumas desvantagens, dentre as quais destacam-se a perda de precisão na recuperação da informação (uma vez que não temos mais os termos exatos para a consulta, e sim seus radicais), e a perda do contexto da informação, que induz a produção de *stems* iguais para termos com sentidos diferentes mas de mesma escrita (homônimos), ocasionando a conflação de termos não relacionados. Como exemplo deste último caso, temos as palavra "verão", que pode ser o verbo "ver" conjugado na 3ª pessoa do plural do futuro do indicativo, ou um substantivo (uma das estações do ano).

Para testar a qualidade dos resultados de uma busca que utilize técnicas de *stemming*, pode-se utilizar duas medidas segundo Flores (2009, p.14): a revocação, que seria a razão entre o número de informações relevantes, retornadas por uma consulta, e o número de informações relevantes, na base de dados para aquela consulta, e a precisão, que seria a razão entre o número de informações relevantes retornadas e o número total de informações retornadas (relevantes e não relevantes).

Além das duas técnicas citadas, Flores (2009, p.14) explica que existe uma medida chamada de *precisão média* ou AvP, que têm um maior foco nos documentos

relevantes retornados nas primeiras posições do ranking, sendo esta a medida mais utilizada. A figura 1.1 mostra a equação que calcula a AvpP.

$$AvP = \frac{\sum_{r=1}^{k}(P(r) \times \mathrm{Re}l(r))}{\sum_{r=1}^{k}(\mathrm{Re}l(r))}$$

Figura 1 - Equação da precisão média
Fonte: Flores (2009, p.14)

> Na Figura 1 – Equação de precisão média, *k* é o número de documentos retornados, *r* é uma posição do ranking, Rel( ) é uma função binária de uma posição de ranking dada (1 se o elemento retornado é relevante ou 0, caso contrário) e P( ) é a precisão para uma dada posição do ranking. Quando mais de uma consulta é usada, é necessário calcular a média das precisões médias, a qual é chamada de *mean average precision* (MAP). O MAP é bastante usado para avaliar os resultados obtidos em um sistema de RI. Para computar as medidas de qualidade de RI, é necessário usar um conjunto de teste composto por documentos, consultas e listas que indiquem quais os documentos relevantes de cada consulta. (FLORES, 2009, p.14)

2.1.4 Tesauros

Segundo Dodebei (1960 apud Vickery, 2002, p.64), a palavra tesauro (latim = *thesauru*, grego = *thesauros*) teve origem na Grécia com o significado de *Treasury or Storehouse* (tesouro ou armazenagem/repositório), entretanto, em 1936, o *Oxford English Dictionary* deu a palavra tesauro a definição de dicionário, enciclopédia ou similares.

Dodebei (2002, p.66) afirma, também, que, a partir de 1940, o termo tesauro começou a ser utilizado pelo SRI, como sendo uma ferramenta capaz de armazenar conceitos e suas relações mútuas, como definidos na língua documentada, porém, regular, com um controle de sinônimos e sintaxe simplificada.

Para Grobmam (2009, p.318), “Tesauro são bibliotecas de dados que podem ser relacionadas entre si”.

Atualmente, tesauros são simples listas de sinônimos, com o objetivo de mostrar as mínimas diferenças entre as palavras, ajudando escritores a escolherem termos mais exatos. Desta maneira, pode-se dizer, também, que são dicionários com idéias que se relacionam,

contudo, não possuindo a característica de definições que um dicionário normal utiliza. Em outras palavras, tesauros são listas de palavras com significado semelhante, dentro de uma determinada área específica de conhecimento. Ressalta-se que tesauros não são apenas listas de palavras-chaves (ou termos), com seus respectivos sinônimos, mas é a constituição de uma hierarquia global de termos que se relacionam. (SOUZA, 2008, p.604).

Murakami (2005, p.11) descreve o surgimento dos tesauros na RI, da seguinte maneira:

> Eles surgiram na década de 50 com o propósito de servir de ajuda para ampliar o vocabulário de indexadores e devido às combinadas pressões de surgimento de novas áreas de assuntos e coleções, de novos modelos no uso da informação e expansão de aplicações de armazenamento e de processamento e recuperação da informação em computadores, foram aperfeiçoados para promover o controle terminológico de sistemas de informação e se tornar uma estrutura conceitual de um determinado campo do conhecimento.
> Desde então, são principalmente utilizados para promover o controle de vocabulário em sistemas de recuperação da informação (SRI). Para isso, são utilizados pelos indexadores no momento da indexação e devem ser disponibilizados para o usuário no momento da recuperação.

Dodebei (2002, p.67) afirma que nas tarefas de indexação e RI, o tesauro tenta resolver o problema de organização de documentos em classes de assuntos, não apenas por possuir a capacidade de controle do vocabulário, mas também por ser um instrumento que relaciona termos de uma forma mais precisa e consistente, criando uma estrutura simples e uma rede de referências cruzadas complexa, apresentando uma relação lógica e hierárquica dos descritores. Dessa forma, o especialista pode localizar, mais facilmente, qualquer termo ou palavra-chave.

No ambiente da Web, os tesauros assumiram a responsabilidade de organizar as informações da própria web. Esta se desenvolve e cresce, constantemente, porém, para que os tesauros possam ser utilizados, estes devem estar representados em um formato compatível com os padrões vigentes na Web, sendo necessário acompanhar sua evolução. (MURAKAMI, 2005, p.11).

Souza (2008, p.608) lista alguns problemas encontrados na Internet, onde o uso de tesauros é recomendado:

> 1) Deficiência nos pontos de acesso à informação;
> 2) Falta de conhecimentos da área pelos usuários para utilizar o sistema;
> 3) Falta de padronização;
> 4) Falta de aprofundamento na construção da Arquitetura da informação do Website;
> 5) Tradução automática de textos;
> 6) Expansão de resultados de busca.

Souza (2008, p.607) cita, também, os três modos mais importantes de representação de tesauros baseados em tecnologias Web:

1) **Tesauro no formato de texto estático, mas navegável**: é o formato mais comum encontrado, sendo que resume a tecnologia HTML. É considerado navegável por utilizar hiperlinks para se locomover entre textos.
2) **Tesauro com interface gráfica**: utiliza tecnologias como Java ou Flash, permitindo uma melhora, com relação à visualização de mapas de redes de relacionamentos entre termos e uma facilidade maior na navegação. É um aperfeiçoamento do formato de texto estático, mas navegável.
3) **Tesauro em formatos legíveis por máquina**: facilita a recuperação automática de informação, porém, ainda não existe uma implementação completa deste modo, podendo ser um avanço para os tesauros na Web.

2.1.5 Transformação de dados

Souza (2006) afirma que "Os modelos clássicos de recuperação são três: o modelo booleano, o modelo vetorial e o modelo probabilístico. Para cada um deles, há modelos alternativos que visam estendê-los em funcionalidade e o desempenho.".

Aqui, serão abordados os modelos clássicos e o modelo fuzzy que serve para conjuntos não convencionais, ou seja, não representados pela lógica clássica.

2.1.5.1 Modelo booleano

Segundo Baeza-Yates e Ribeiro-Neto (1999), esse modelo é baseado na teoria dos conjuntos e na álgebra booleana.

Jackson e Moulinier (2002) definem a pesquisa booleana como aquela em que o usuário procura num banco de dados com uma consulta que conecta palavras com os operadores, como E, OU e NÃO.

As tabelas de verdade do AND (E), OR (OU) e NOT (NÃO) são apresentados na Figura 2.

AND	TRUE	FALSE
TRUE	TRUE	FALSE
FALSE	FALSE	FALSE

OR	TRUE	FALSE
TRUE	TRUE	TRUE
FALSE	TRUE	FALSE

NOT	
TRUE	FALSE
FALSE	TRUE

Figura 2 - Tabela da verdade
Fonte: Figura nossa, baseado em Jackson e Moulinier (2002, p.29)

"O mecanismo de busca retorna os documentos que possuem combinações dos termos que satisfazem à construção lógica da consulta [...]" (GONZALES et al., 2003).

Poder-se-ia, por exemplo, fazer uma consulta para nos retornar os arquivos de áudio, imagens e bibliografia de certo artista, exceto os arquivos de certo ano, ficando da seguinte maneira: *(audio OR imagem OR bibliografia) AND artista AND NOT ano*.

Conforme Sparck-Jones (1997), os principais problemas do modelo Lógico são: (a) normalmente, o usuário não possui treinamento apropriado, tendo dificuldade em formular consultas usando operadores lógicos; (b) há pequeno controle sobre o tamanho da saída produzida por uma determinada consulta; e (c) a recuperação lógica resulta em uma simples partição da coleção de documentos em dois subconjuntos discretos: os registros que satisfazem a consulta e os que não a satisfazem.

2.1.5.2 Modelo espaço vetorial

O modelo espaço-vetorial (*vector-space model*) foi desenvolvido por Gerard Salton, que, durante anos contribuiu com estudos relevantes para a área. Salton desenvolveu esse modelo para poder ser utilizado em um SRI, chamado SMART que ele criou, enquanto trabalhava na universidade de Cornell (WIVES, 2002).

O modelo de espaço vetorial (Salton & McGill, 1983) é uma variação do modelo booleano. Diferentemente deste, onde apenas a frequência absoluta de ocorrência de uma palavra é considerada, o modelo de espaço vetorial busca privilegiar palavras que ocorrem de forma concentrada em alguns textos (mesmo que a frequência absoluta destas palavras seja elevada em relação ao conjunto de documentos).

Este modelo representa uma coleção de documentos como vetores de termos com seus respectivos pesos de relevância, em um mesmo ambiente vetorial, com objetivo de viabilizar operações como o cálculo de relevância, a classificação e o agrupamento de documentos semelhantes. (MANNING, RAGHAVAN e SCHÜTZE, 2008).

Assim, cada documento é representado por um vetor associado, que é constituído por pares de elementos na forma {(termo1, peso1), (termo2, peso2), (termo3, peso3)}. Cada elemento do vetor é considerado uma coordenada dimensional. Desta forma, os documentos podem ser colocados em um espaço Euclidiano de "n" dimensões (onde "n" é o número de termos) e a posição do documento em cada dimensão é dada pelo seu peso. (IGARASHI, 2005).

Para Manning et al. (2008), os termos mais relevantes de um documento são calculados normalmente pelo TF-IDF, em que DF é a frequência dos documentos (ou *Document Frequency*) que contêm o termo a ser pesquisado. Outro dado que precisa ser extraído é o IDF, que é calculado a partir do próprio DF. Trata este da quantidade de vezes, em que o termo é encontrado, no conjunto de documentos.

O valor TF é proporcional à frequência das palavras no documento, enquanto que o valor do IDF é inversamente proporcional à frequência do documento no corpus (ZHANG; GONG; WANG, 2005, p. 49-55).

Segundo Igarashi (2005), neste espaço Euclidiano, a consulta do usuário, também, é representada por um vetor. Dessa forma, os vetores dos documentos podem ser comparados com o vetor da consulta e o grau de similaridade entre cada um deles pode ser identificado.

Para isso, se faz uso das medidas de similaridade entre vetores (*Vector-Based Matching*), as quais podem ser: medida do co-seno, índice de Jaccard, índice de Dice, índice "N", medidas de sobreposição (*overlap measures*). Dentre estas medidas, uma das mais utilizadas para o processo de comparação vetorial é o co-seno, devido ao seu grau de estabilidade (EGGHE, 2002, p. 845).

Segundo Ferneda (2003), o cálculo de similaridade permite estabelecer o grau de semelhança entre dois documentos, ou ainda, entre os documentos com os termos a serem

pesquisados. Abaixo, na Figura 3 – Fórmula medida do co-seno, temos a fórmula onde “w” é o peso do elemento “i” nos vetores “x” e “y”.

$$sim(x,y) = \frac{\sum_{i=1}^{t}(w_{i,x} \times w_{i,y})}{\sqrt{\sum_{i=1}^{t}(w_{i,x})^2 \times \sum_{i=1}^{t}(w_{i,y})^2}}$$

Figura 3 - Fórmula medida do co-seno
Fonte: Ceci (2010, p.40)

O grau de similaridade obtido, através da medida do co-seno, representa as distâncias entre documentos, ou seja, documentos que possuem os mesmos termos são colocados em uma mesma região do espaço e, em teoria, tratam de um assunto similar, sendo nomeada esta característica de espaço-vetorial. (IGARASHI, 2005)

Abaixo, Figura 4, ilustrando o modelo espaço-vetorial.

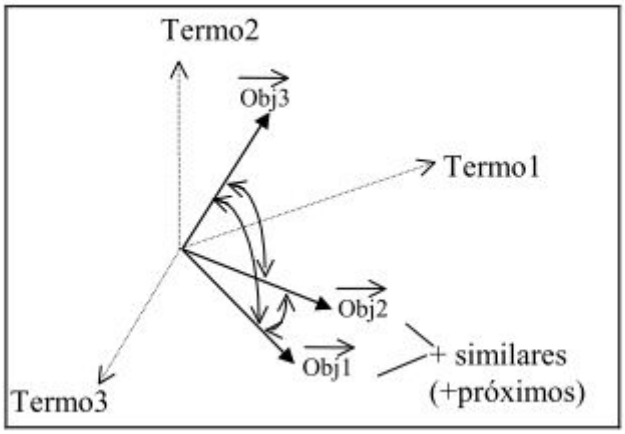

Figura 4 - O modelo espaço-vetorial
Fonte: Wives (2002, p.39)

2.1.5.3 Modelo Fuzzi

A lógica fuzzy, definida por Zadeh (1973), serve, na área de recuperação de informações, para diminuir as incertezas do uso de termos linguísticos e para detalhar a importância dos termos em relação à consulta, a relevância dos documentos para a consulta e o grau de cada termo em um documento.

Segundo Leite (2009):

> Para lidar coma limitação de pertinência binária, imposta pelo modelo booleano, surgiu o modelo fuzzy. A estrutura básica para representação formal dos relacionamentos neste modelo é a relação fuzzy a qual estende o conceito matemático de relação. Enquanto as relações matemáticas clássicas descrevem apenas a presença (1) ou ausência (0) de associação entre elementos de dois conjuntos, as relações fuzzy permitem expressar o grau da relação.

Souza (2006) explica que, nesse modelo, se busca entender o conceito da representação dos documentos, por palavras chave, sendo que cada *query* determina um conjunto difuso, e cada documento pertence, de algum modo, a esse conjunto. A modalidade de pertencer a este conjunto pode ser determinada pela ocorrência de palavras na *query*, como no modelo booleano, mas também, pode utilizar um tesauro para determinar termos, com relação semântica aos termos índice, para se verificar o grau da modalidade de pertencer ao conjunto difuso.

2.1.5.4 Modelo probabilístico

O modelo probabilístico para a recuperação foi apresentado, pela primeira vez, por Maron e Kuhn (1960). Desde então, tem sido elaborado de diferentes maneiras, testado e aplicado.

Jones (2000) explica que o modelo probabilístico tenta buscar a resposta, através da probabilidade de que um documento com certa descrição é relevante ou não para a consulta feita.

Van (1979) afirma que a principal ferramenta matemática do modelo probabilístico é o teorema de Bayes.

Dada uma consulta q e um documento dj, Leite (2009) afirma que:

> No modelo probabilístico o peso dos termos de indexação é binário, isto é, wij ∈ {0, 1} e wiq ∈ {0, 1}. Uma consulta é um subconjunto dos termos de indexação (ki).
>
> Seja R o conjunto de documentos considerados relevantes e R o complemento de R, ou seja, o conjunto de documentos não relevantes. P (ki|R) é a probabilidade do termo ki estar presente em um documento selecionado aleatoriamente do conjunto R. P ki|R é a probabilidade do termo ki estar presente em um documento selecionado aleatoriamente do conjunto R.

Assim, a relevância do documento dj para a consulta q, levanto em conta t termos de indexação da coleção, é dada pela equação apresentada na Figura 5 – Equação de Modelo Probabilistico.

$$\mathrm{r}(d_j,\, q) \;\approx\; \sum_{i=1}^{t} w_{iq}\, w_{ij} \;\left(\log \frac{P\left(k_i|R\right)}{1 \,-\, P\left(k_i|R\right)} + \log \frac{1 \,-\, P\left(k_i|\overline{R}\right)}{P\left(k_i|\overline{R}\right)} \right)$$

Figura 5 - Equação de Modelo Probabilístico
Fonte: Leite (2009, p.13)

De acordo com Cardoso (2000), além do bom desempenho prático, o princípio probabilístico de ordenação resulta em um comportamento ótimo do método. Entretanto, a desvantagem é que esse comportamento depende da precisão das estimativas de probabilidade. Além disso, o método não explora a frequência do termo no documento.

2.1.6 Indexação

Conforme Lancaster (2004, p.20):

> Indexação [...] refere-se à representação do conteúdo temático de partes de itens bibliográficos inteiros [...] O processo pelo qual o conteúdo temático de itens bibliográficos é representado em base de dados publicadas – em

> formato impresso ou eletrônico [...]quer se estejam examinando itens total ou parcialmente.

Complementando a afirmação de Lancaster, a indexação: "é uma combinação metodológica altamente estratégica entre tratamento do conteúdo de documentos e sua recuperação por um usuário, demonstrando uma relação estreita entre o processo e a finalidade da indexação." (FUJITA, 2003, p. 61).

Para Lancaster (1993, p.75):

> Define-se de um modo muito pragmático a 'boa indexação' como a indexação que permite que se recuperem itens de uma base de dados durante buscas para as quais eles sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis.

Dentre as formas de indexação, temos a indexação manual e a indexação automática.

Segundo Lancaster (1993), a indexação manual envolve duas etapas principais: a análise conceitual e a tradução. A análise conceitual é a atividade de definição dos assuntos que são tratados no documento; e a tradução corresponde, por sua vez, à atividade de conversão dos conceitos identificados na análise para uma linguagem de indexação.

Para Pinheiro (1978), a indexação envolve julgamento e, consequentemente, oscila muito no seu nível de concordância, e apresenta discrepâncias.

Pode-se citar, como problemas no processo manual, o tempo restrito do indexador, a quantidade cada vez maior de documentos passíveis de tratamento, a falta de conhecimento do indexador sobre o domínio do documento, a subjetividade, a inconsistência interindexadores, e a falta de domínio do idioma do documento.

Já a indexação automática destaca dois tipos de processos: a indexação por extração automática e a indexação por atribuição automática. O primeiro tipo, conforme Lancaster (2004) é feito a extração de palavras ou expressões do texto para representar seu conteúdo. Pode-se usar um software para extrair os termos a partir dos princípios utilizados por seres humanos (frequência, posição e contexto de palavra no texto). O segundo tipo é mais complexo, em relação ao anterior. Necessidade de controle terminológico para a representação do conteúdo temático. Desenvolve-se, para cada termo atribuído, um 'perfil' de palavras ou expressões associativas ao termo e que ocorrem nos documentos.

Para Lancaster (1978), não há dúvida sobre a eliminação progressiva das práticas de indexação manual frente às técnicas de indexação automática.

Para Carneiro (1985), os objetivos de uma política de indexação são as definições das variáveis que afetam o desempenho do serviço de indexação, o estabelecimento dos princípios e critérios que servirão de guia na tomada de decisões para otimização do serviço, a racionalização dos processos e a consistência das operações envolvidas.

De acordo com Cesarino (1978, p.271):

> Todo o procedimento de recuperação de informações é ligado à manipulação de 'classes'. Quando indexamos um documento, estamos colocando-o em uma classe determinada. Para facilitar o processo, cada classe recebe 'um nome', que é chamado 'termo indexador'. Ao conjunto de termos indexadores chamamos Linguagens de Recuperação de Informações ou Linguagens de Indexação.

Segundo Jackson e Moulinier (2002), um índice para a pesquisa de texto completo de documentos eletrônicos é, geralmente, mais exaustivo do que o índice de qualquer livro. Alguém poderia querer consultar documentos, combinando os termos da busca com termos que, de fato, ocorrem no texto desses documentos. Isso requer que um documento seja indexado com todas as palavras que ocorrem nele, ao invés de que sejam indexados, somente, por palavras-chave ou cabeçalhos de assunto prestado por um editor ou um bibliotecário.

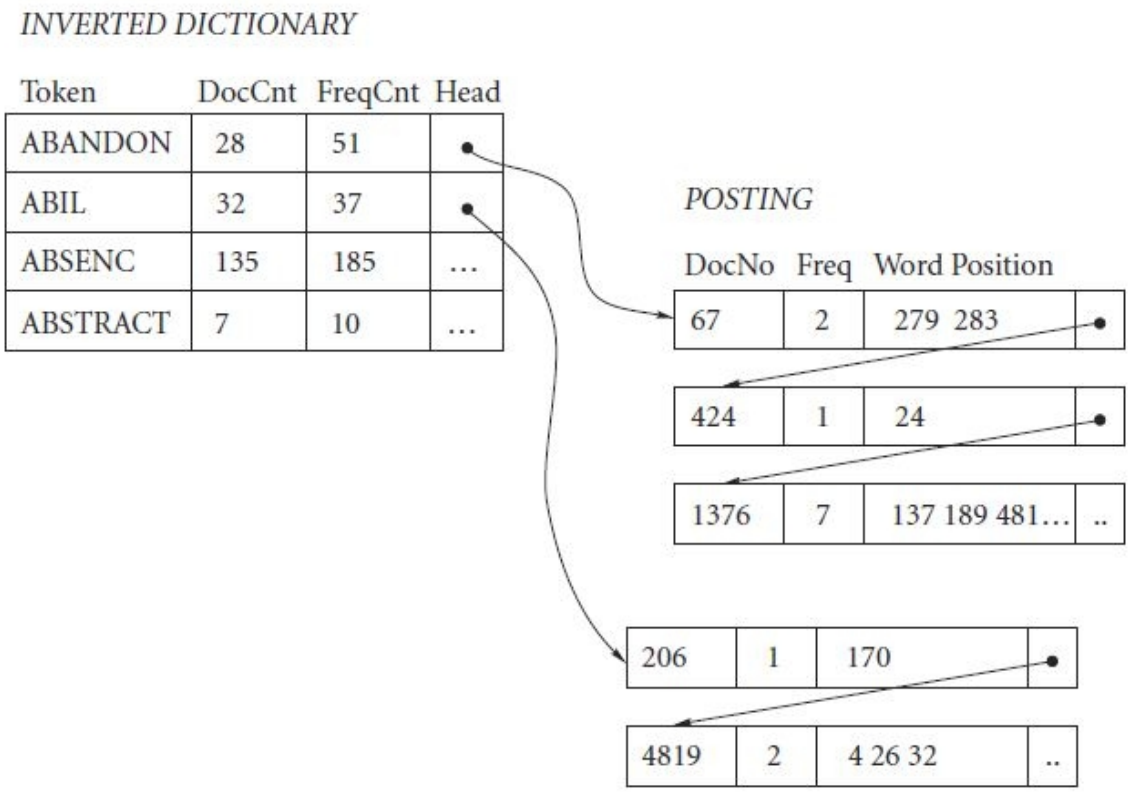

Figura 6 - Dicionário invertido

Fonte: Jackson et al(2002, p.28)

Conforme explica Igarashi (2005) em relação a Figura 6, **DocCnt** é o contador de documentos, informa em quantos documentos o termo ocorre. Isto permite computar estatísticas úteis, com o propósito de criar uma pontuação de relevância, chamada de DF (*Document Frequence* ou Freqüência de Documentos); **FreqCnt** é o contador de frequência total, informa qual é o total de ocorrências de um termo em todos os documentos, esta é uma medida básica de quão comum é o termo; **Freq** é a frequência**,** informa quantas vezes o termo ocorre no documento; **Word position** registra as localizações no texto em que as palavras foram encontradas.

Igarashi (2005, p.33) afirma que:

> É necessário destacar que durante o processo de indexação, as palavras devem sofrer algum tipo de tratamento, desse modo as variantes de uma palavra são indexadas apenas uma única vez.

Para Frakes (2003, p26), uma das formas mais usadas para tratar uma palavra é a técnica de *stemming*. Esta técnica visa reduzir a palavra a sua raiz. Conseguindo através desta, promover uma redução no tamanho do dicionário, ou seja, redução do número de termos distintos necessários para representar um conjunto de documentos.

2.2 ANOTAÇÃO DE DOCUMENTOS

Tendo como base o trabalho de Ramiro (2005, p. 636), pode-se afirmar que as anotações têm como principal objetivo facilitar o acesso e a recuperação de documentos na web, servindo como um índice que se relaciona com o conteúdo do documento.

Para Eller (2008, p. 39), pode-se representar anotações de duas formas, intrusiva e não intrusiva, sendo que a forma intrusiva ocorre quando as anotações são armazenadas no próprio documento e a forma não intrusiva, quando são guardadas em repositórios de anotações, onde as anotações apontam para os documentos que foram anotados.

2.2.1 Anotação automática

Jesus (2009, p. 25) cita em seu trabalho que, para realizar anotação automática, no caso de imagens, é necessário uma extração de características do conteúdo visual, o que para a recuperação de informações é algo complexo, pois, tratam de cores, texturas e formas. Porém, também, se pode utilizar de metadados contidos no cabeçalho EXIF (*Exchangeable Image File Format*) do ficheiro JPEG (*Joint Photographic Experts Group*) da imagem, como instante de captura, informação de GPS ou distância ao sujeito.

Outra técnica citado por Jesus (2009, p. 37) baseia-se na classificação binária para anotar as imagens, consistindo em detectar a presença ou não de objetos e lugares na imagem através de classificadores binários.

Para Eller (2008, p. 69), "ferramentas de anotação automática são as mais indicadas para grandes volumes de documentos. Contudo, geralmente, estas ferramentas ainda estão sujeitas a falhas, ocasionando anotações errôneas e indesejadas".

2.2.2 Anotação semiautomática

A anotação semiautomática ocorre quando uma parte da anotação é feita de forma automática e a outra é realizada por um ser humano. (JESUS, 2009, p. 29).

A anotação semiautomática, para Reeve (2005), tem o objetivo de auxiliar o processo de anotação, sendo que permite identificar e classificar, automaticamente, documentos, porém, com maior precisão devido à intervenção humana dentro desse processo.

Jesus (2009, p. 29) aborda dois tipos de sistemas que utilizam da anotação semiautomática, considerando imagens:

- Retroação de relevância: mecanismo, onde o utilizador valida os resultados obtidos, através de uma interrogação, adicionando esta informação escolhida pelo usuário e apresentando novos resultados para validação. O processo se repete, fazendo com que o sistema inclua as palavras escolhidas na interrogação que são associadas às imagens consideradas relevantes;

- Anotação através de áudio: é utilizado um microfone para que o usuário faça comentários sobre a imagem, e, posteriormente, essa informação é transcrita para o formato de texto, por uma aplicação de reconhecimento de voz.

2.2.3 Anotação manual

Com base nas palavras de Jesus (2009, p. 19), a anotação manual é a descrição de um documento, feita de forma manual por um ser humano. Sua maior desvantagem se encontra quando temos um grande volume de documentos, o que despende muito tempo para realizar toda a anotação, tornando essa tarefa cansativa. De outro ponto de vista, a anotação manual pode trazer benefícios para algoritmos automáticos, pois são construídos, utilizando exemplos anotados manualmente. (YAN, 2007, p. 14).

Yan (tradução nossa, 2007, p. 14) mostra dois tipos de abordagem com relação a anotações manuais de imagens, que podem, de certa forma, servir para outros tipos de documentos:

- *Browsing* (navegação): quando é escolhida uma palavra ou termo, sendo que o usuário deverá percorrer a base de dados e anotar apenas documentos que se relacionem com o termo escolhido, fazendo com que o usuário não necessite lembrar todas as possíveis palavras-chave durante um longo período de tempo;
- *Tagging* (etiquetagem): ao contrário da abordagem de *browsing*, se apresenta quando se escolhe um documento e o usuário adiciona palavras para descrevê-lo. Essa abordagem torna as anotações mais flexíveis, pois o usuário poderá escolher livremente palavras para a descrição do documento, porém essa flexibilidade pode trazer problemas se muitos usuários utilizarem palavras diferentes para conceituar o mesmo documento.

Jesus (2009, p. 26) afirma que a anotação manual é atualmente a maneira mais eficiente para realizar a associação das imagens com as palavras que irão descrevê-la. Segundo ele, essa técnica é normalmente utilizada por pessoas para a organização de suas fotos pessoais, e cita algumas aplicações comerciais como o iPhoto, o Picassa e o Adobe

Photoshop Album, sendo estas, aplicações que utilizam de anotação manual. Esses aplicativos permitem uma organização e recuperação de fotos em coleções.

2.2.4 Anotação semântica

Segundo Eller (2008, p. 38), a anotação semântica é fundamental para a Web Semântica, pois permite a criação de novos documentos com um conteúdo semântico definido ou então acrescenta semântica aos documentos existentes.

Através da anotação semântica "é possível correlacionar termos (conceitos, instâncias ou propriedades) da ontologia a palavras, simples ou compostos, do texto que passou pelo processo de anotação semântica." (ELLER, 2008, p. 38).

Eller (2008, p. 39) explica que o processo de anotação semântica possui várias etapas como:

- Análise léxica: onde ocorre a separação das palavras em tokens;
- Análise sintática: onde ocorre a identificação das palavras e eliminação dos conectores;
- Extração dos radicais das palavras;
- A associação entre as palavras relevantes encontradas e suas definições semânticas na ontologia;
- Armazenamento das anotações.

2.2.5 Folksonomia

Segundo Fernades (2010, p.33), o termo folksonomia foi utilizado por Thomas Vander Wal em 2004 como sendo uma analogia ao termo taxonomia, onde sua principal função é criar anotações a partir da linguagem de pessoas que a utiliza, sendo assim folksonomia é uma maneira de classificar qualquer tipo de informações disponíveis na Web.

Tendo em vista o conceito de web 2.0, a folksonomia é sem dúvidas um dos recursos que mais caracteriza a condição de construção coletiva de inteligência informacional, sendo considerado um elemento fundamental para a evolução dessa tecnologia. (VIDOTTI, 2011, p.286).

Aquino (2007) conceitua folksonomia como um sistema de indexação de informações que tem como objetivo permitir a adição de tags ou etiquetas que descrevem o conteúdo dos documentos armazenados, sendo que essas tags são criadas pelos próprios usuários da web, que de uma forma coletiva representam, organizam e recuperam as informações no ambiente web.

> A finalidade da Folksonomia seria ordenar o caos existente na web. Embora sua característica de liberdade para classificar aponte para a idéia de uma falta de estrutura organizacional, o resultado para quem pesquisa é uma maior facilidade para encontrar termos que as demais linguagens de indexação não conseguem acompanhar em suas tabelas hierárquicas. (BLATTMANN, 2007, p.207).

Têm-se, como exemplo, sites como a Wikipédia e Youtube que utilizam da folksonomia para permitir que seus próprios usuários criem cabeçalhos (anotações), representando os documentos ou objetos que serão classificados. Dessa forma, após a classificação, essas anotações podem ser compartilhadas por um conjunto de serviços, relacionando as informações não importando onde elas se localizem. O usuário simplesmente cria a anotação usando o critério que acha mais relevante para quando for buscar o conteúdo. (FERNANDES, 2010, p.34).

Para ter uma melhor visualização do funcionamento desse conceito, Fernandes (2010, p.62) mostra uma representação de um sistema de classificação que utiliza da folksonomia (Figura 7), mostrando a vantagem de se usar as informações geradas, permitindo um aumento na precisão de consultas, uma vez que vários usuários podem adicionar diversas anotações a um mesmo objeto.

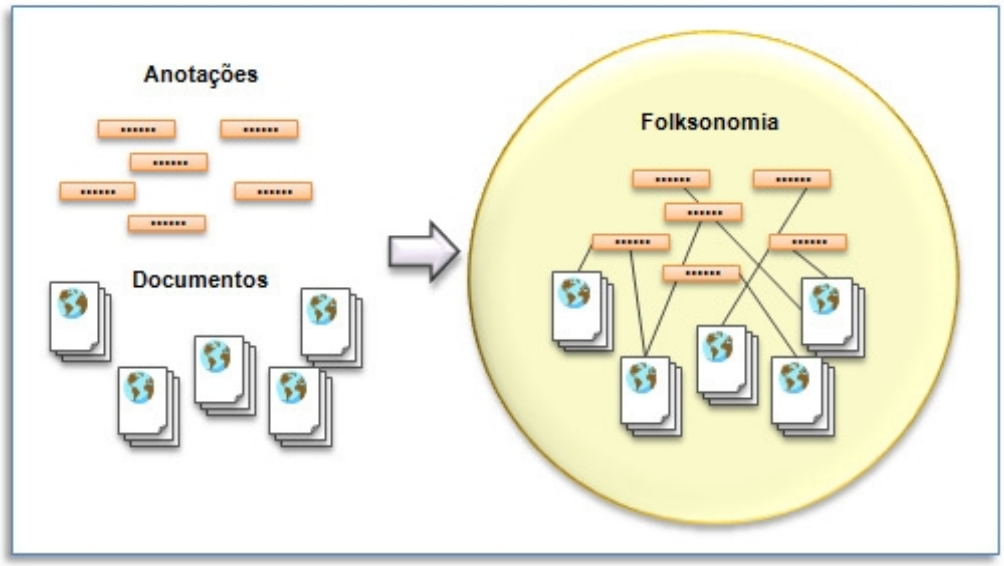

Figura 7 - Folksonomia de documentos
Fonte: Fernandes (2010, p.62)

Entretanto, para Aquino (2001, apud DREYFUS, 2007), a folksonomia apresenta problemas em relação à disposição das informações na Web, sendo que tudo pode ser linkado, não havendo hierarquias nem limites, permitindo os links se proliferarem descontroladamente.

Vidotti (2011, p.286) apresenta sua conclusão, em relação à Web e à folksonomia:

> A Folksonomia mudou o paradigma em relação à recuperação da informação em ambientes *web*, tanto que é comum ver sites apresentando buscas baseadas em palavras-chave que foram inseridas pelo próprio usuário dentro do ambiente. Portanto trata-se de um recurso rico, que contribui de forma acentuada para o fortalecimento e solidificação da Internet como plataforma para construção de informação coletiva.

2.3 RECUPERAÇÃO DE INFORMAÇÃO MULTIMÍDIA

Tendo em vista o grande volume de informações em formato texto, gráfico, áudio e vídeo, distribuídas em diferentes base de dados, se faz necessário uma forma de buscar e recuperar estas informações. Entre as aplicações deste tipo de recuperação, pode-se citar diagnósticos médicos (*Medical Image Database*), reconhecimento de padrões como impressões digitais, e a própria pesquisa multimídia.

O tipo de anotação disponível para a recuperação condiciona a forma como o utilizador pode formular a consulta e, por consequência, como são recuperados os documentos. Em geral, esta recuperação consiste na construção de uma lista ordenada de documentos que satisfazem à consulta. Para tal, é necessário medir a relevância que cada documento tem para a consulta, e utilizar um sistema de indexação que permita aceder aos documentos, de forma eficiente. (JESUS, 2009)

2.3.1 Expansão de consulta

Segundo Xu (1996), um problema fundamental em recuperação de informação é que os autores nem sempre usam as mesmas palavras que os usuários para descrever o mesmo conceito.

Cardoso (2002) afirma que:

> A importância deste problema tende a diminuir com o aumento do tamanho da consulta. Entretanto, em muitas aplicações, as consultas podem possuir uma pequena quantidade de termos. Um caso extremo ocorre no contexto da Web, onde as consultas possuem tipicamente duas palavras.

A expansão de consulta é uma técnica na qual se busca aumentar a quantidade de termos que devem ser buscados, sendo que estes devem possuir certo grau de equivalência entre si para aumentar a probabilidade de se encontrarem documentos relevantes. (CARDOSO, 2002)

Para Leite (2009) “a expansão de consulta consiste em adicionar novos termos semanticamente relacionados como os termos presentes na consulta inicial em função do conhecimento contido em uma base de conhecimento [...]”.

Segundo Cardoso (2002), os vários métodos de expansão de consulta podem ser divididos em dois principais grupos, métodos iterativos e métodos automáticos.

Métodos iterativos, *User FeedBack Relevance*, também, conhecidos por expansão semi-automática é, provavelmente, a técnica mais comum de expansão de consulta. Conforme afirma Bettio (2007), esta técnica requisita que o usuário atribua relevância a um conjunto de documentos trazidos através de uma busca inicial, ou seja, é necessário interação do usuário com o sistema.

Segundo Christopher (2008), os métodos automáticos, diferentes dos iterativos não necessitam de interação com o usuário, o que os torna uma técnica mais interessante, uma vez que o processo é transparente para o usuário.

Conforme afirma Fernandes (2010), o mecanismo de expansão de consultas é essencial no processo de recuperação da informação. Com a ajuda deste, as pesquisas nessa área permanecem com abordagens mais voltadas para o usuário, que possui um papel central nesse contexto.

2.3.2 MPEG-7

De acordo com Barros (2010) o MPEG-7 é formalmente conhecido como Multimedia Content Description Interface, ou interface para descrição de conteúdos multimídia, desenvolvido para que possa ser usado em grandes bases de dados multimídia.

O MPEG-7 começou a ser desenvolvido em 1996 pelo Moving Picture Experts Group-MPEG, um grupo de trabalho ISO/IEC, que tem o objetivo de desenvolver padrões para representação codificada de dados digitais de áudio e vídeo. O MPEG é aprovado pela International Organization for Standardization - ISO, sendo que os padrões anteriores (MPEG-1, MPEG-2 e MPEG-4) tinham foco na codificação e compactação dos conteúdos (VICENTE, 2005).

O objetivo final do MPEG-7 é promover interoperabilidade entre sistemas e aplicações usados na geração, gerenciamento, distribuição e consumo de descrições de conteúdos áudio-visuais. (CHANG et al, 2001, p. 688, tradução nossa)

O MPEG-7 é uma ferramenta que, desde sua criação, vem se consolidando, sendo utilizado nas práticas de descrição de conteúdo de objetos multimídia em sistemas de gerenciamento e recuperação da informação que servem a diferentes grupos de usuários, e que têm apresentado resultados práticos satisfatórios quanto à eficiência do sistema. (DALLACOSTA et al, 2004; CAO et al, 2009; KANNAN et al, 2009; LUX, 2009; REY-LOPEZ et al, 2009; KAPELA et al, 2010)

O MPEG-7 possui uma estrutura mais complexa que as versões anteriores, prevê ferramentas de descrição (Description Tools) representadas por elementos de metadados. Esses elementos são empregados “para criar descrições que serão utilizadas por ferramentas

com funções para pesquisar, filtrar e navegar de forma eficiente em conteúdos multimídia" (CHELLA, 2004).

Com essa ferramenta é possível descrever as características dos conteúdos multimídia para que os usuários possam pesquisar, recuperar e até mesmo navegar por estes conteúdos; as descrições podem estar, fisicamente, armazenadas em conjunto com o material áudio-visual, como também podem estar em outro lugar da rede (DALLACOSTA et al, 2004).

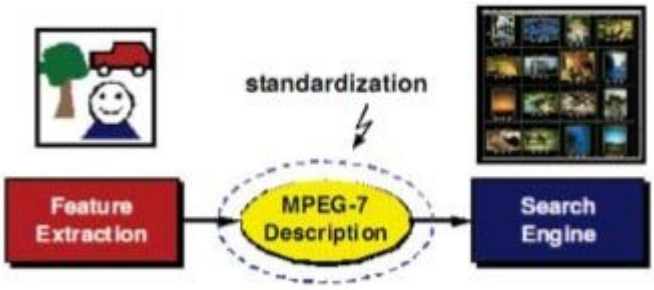

Figura 8 - MPEG-7 e sistemas de busca de informações
Fonte: Vicente (2005, p.4)

Vicente (2005) ressalta que o MPEG-7 não é um padrão para busca e recuperação de conteúdos, mas sim uma ferramenta para descrição de conteúdo multimídia com o objetivo de proporcionar uma linguagem comum de forma a aperfeiçoar tais processos.

2.3.3 Extensão de busca para Web

Nesta seção, serão abordados os tópicos que englobam a busca na Web, apresentando conceitos como Web 2.0 e metadados.

2.3.3.1 Metadados na Web

O conceito central da Web Semântica é a busca pela "Web de Dados", ou seja, uma rede que enxergue todos os dados disponíveis, independente de quais aplicações sejam proprietárias destes dados (W3C, 2008).

Dias et al. (2003) afirma que a heterogeneidade da informação dificulta a integração de conteúdo na Web e explica que uma das formas de descrição homogênea da informação é através de metadados.

Segundo Souza e Alvarenga (2004):

> Um documento na Web é composto por uma mistura de dados e metadados. "Meta" é um prefixo de auto-referência, de forma que "metadados" sejam "dados sobre dados". Os metadados em documentos na Web têm a função de especificar características dos dados que descrevem, a forma com que serão utilizados, exibidos, ou mesmo seu significado em um contexto.

De acordo com Iannela et al. (1997), no contexto da Web, três aspectos devem ser considerados no desenvolvimento de metadados: descrição de recursos, produção e uso de metadados.

Souza e Alvarenga (2004) explicam que o primeiro aspecto refere-se a informações que estarão sendo consideradas nos metadados e que estes metadados têm que ser suficientemente flexíveis, para capturar informações de diversas fontes distintas. O segundo aspecto se refere à construção de metadados, onde os metadados nada mais são do que sumários sobre uma determinada informação. Por fim, o terceiro e último aspecto trata de como os metadados serão acessados e utilizados, que estes têm que estar disponibilizados de maneira que possam ser processados, preservando seu conteúdo semântico.

2.3.3.2 Web 2.0

O termo Web 2.0 surgiu em 2004 durante uma conferência promovida pelas empresas de mídia Media-Live e O'Reilly Media. Nesta conferência, discutiu-se a idéia de que a Web deveria ser mais do que apenas uma plataforma, - deveria ser dinâmica e interativa, e colocar o usuário no centro disto.

De acordo com O'Reilly (2005):

> Não há como delimitar fronteiras para a web 2.0, pois trata-se de princípios e práticas para que diversos sites sigam. Um dos princípios fundamentais é a web como plataforma, ou seja, o usuário poder realizar atividades online que antes só eram possíveis com programas rodando em seu computador.

O'Reilly (2005) enfatiza o desenvolvimento do que chama de "arquitetura de participação", ou seja, o sistema informático que incorpora recursos de interconexão e compartilhamento. Por exemplo, nas redes peer-to-peer, as quais tem como objetivo a troca de arquivos digitais, cada computador conectado à rede se torna cliente e servidor ao mesmo tempo, pois pode fazer downloads de arquivos disponíveis na rede como pode prover arquivos. Assim, quanto mais pessoas na rede, mais arquivos disponíveis. Isso comprova, segundo O'Reilly, um princípio da Web 2.0, os serviços se tornam melhores quanto mais pessoas o usem.

Pita e Paixão (2010) afirmam que na Web 2.0, os usuários tomam um papel mais ativo, publicando conteúdo em vez de apenas consumir, exemplo de aplicações web que ajudaram a construir a Web 2.0 que conhecemos hoje são as redes sociais.

Coutinho e Bottentuit (2007) concluem que a Web 2.0 é uma forma de utilização colaborativa da internet, em que o conhecimento é compartilhado de maneira coletiva e descentralizado de autoridade para utilizá-lo e reeditá-lo.

2.4 CONSIDERAÇÕES FINAIS

Neste capítulo, foi explanado conceitos com base referencial para o desenvolvimento do trabalho. Pode-se dar destaque ao conceito de RI e a diferença sobre o conceito de EI, a simplicidade dos conceitos dos modelos de transformação de dados, a importância do processo de indexação, a forma que as anotações facilitam o acesso à recuperação de documentos nas buscas, a folksonomia com o intuito de criar de tags com a linguagem dos usuários e ordenar o caos da web, e, por fim, a expansão de consulta, que no contexto de buscas na web se faz mais do que necessário.

3 MÉTODO

Neste capítulo, define-se o tipo de pesquisa do trabalho proposto e o porquê do mesmo ser classificado neste grupo, definimos a lista de etapas, que serão os próximos passos para a conclusão do trabalho. Apresenta-se, também, o esquema de solução, assim como as delimitações do trabalho, e por fim, mostra-se o cronograma criado com base na lista de etapas definidas.

3.1 CARACTERIZAÇÃO DO TIPO DE PESQUISA

O que se chama de método científico consiste na percepção de uma estrutura lógica de ações frequentemente utilizadas na pesquisa científica, mas que, por si só, não é suficiente para garantir o êxito desse empreendimento. Os resultados satisfatórios de uma pesquisa dependem de amplo conjunto de fatores, que abrange desde a natureza do problema a ser pesquisado até os recursos materiais aplicados na pesquisa e depende, sobretudo, da criatividade e da inteligência do pesquisador (COTRIN et. al, 2002, p. 241).

Este trabalho pode ser classificado como uma pesquisa aplicada. De acordo com Cervo (2002, p.65) "na pesquisa aplicada o investigador é movido pela necessidade de contribuir para fins práticos mais ou menos imediatos, buscando soluções para problemas concretos".

Após reunir o referencial teórico que deu apoio para o desenvolvimento da pesquisa, foi construído um esquema que apresenta uma proposta de solução. A fim de que a viabilidade do mecanismo seja atestada, foi construído um protótipo que será aplicado num determinado cenário. Este cenário propõe a recuperação de artigos e informação multimídia relacionadas com o mesmo através de um aplicativo web de busca. A partir deste protótipo, será validada a proposta de solução, apresentando ao fim do trabalho os resultados, conclusões e possíveis evoluções na investigação na seção de trabalhos futuros.

3.2 ESQUEMA DA SOLUÇÃO

A figura 9 ilustra o esquema do sistema proposto, envolvendo um cenário acadêmico, onde são armazenados artigos, para posteriormente serem recuperados pelo sistema, porém retornando outros tipos de arquivos relacionados ao artigo, como vídeos, textos ou imagens.

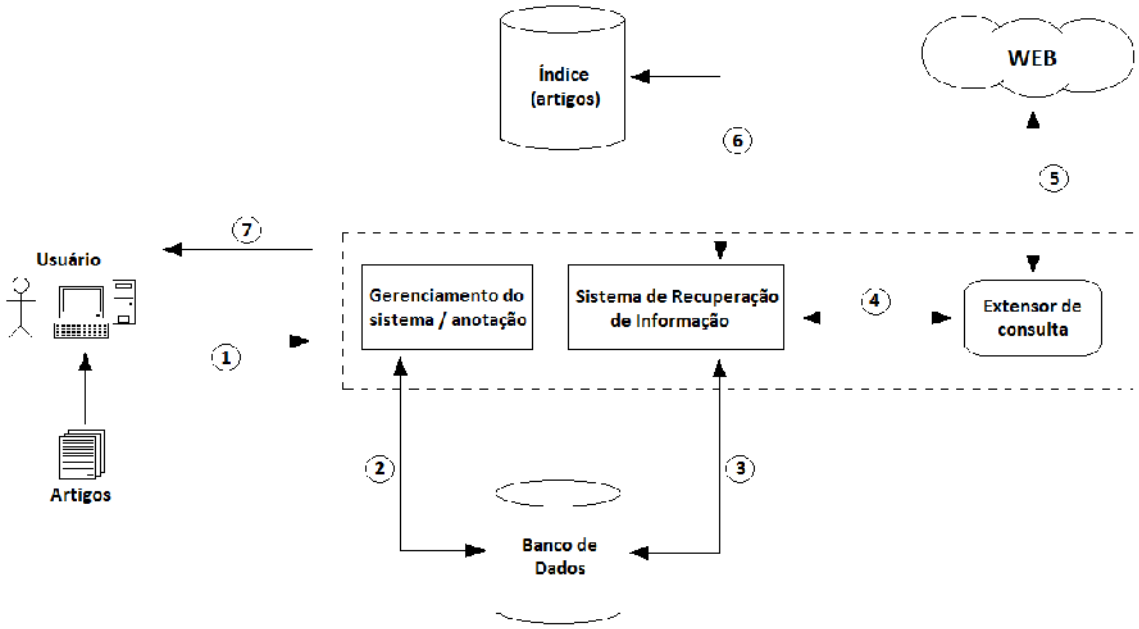

Figura 9 - Esquema do sistema proposto
Fonte: Autores

O esquema do sistema, como exibido na imagem acima, é formada, basicamente, das seguintes camadas:

a) Camada de apresentação: é a interface web a qual o usuário irá interagir, sendo responsável pelo envio (1) das informações ao sistema, como o cadastro dos dados do usuário e os documentos (artigos) que serão gravados no banco. Através dessa camada o usuário também poderá visualizar as informações recuperadas (7).

b) Camada de negócio: é o núcleo do sistema, sendo formado pelos módulos de gerenciamento do sistema, sistema de recuperação de informação e extensão de consulta. O gerenciamento do sistema irá gerenciar os usuários, realizar o cadastro dos documentos, além de ser responsável pelas anotações que serão criadas. O SRI deve indexar os documentos

cadastrados e recuperá-los através do termo de busca. E a extensão irá trazer informações multimídia da web, como vídeos, imagens e textos.

c) Camada de dados: é representada pelo banco de dados e informações retornadas da web. O banco de dados irá armazenar as informações cadastradas pelo usuário (2), além de armazenar os documentos que irão ser indexados pelo Sistema de Recuperação de Informação (3). As informações retornadas da web (5) serão mostradas ao usuário (7) através do Extensor de consulta (4), que irá realizar a busca em alguns sites pré-determinados.

3.3 DELIMITAÇÕES

O trabalho é limitado à modelagem e desenvolvimento de um protótipo, tendo o objetivo de realizar buscas através de termos literais informados pelo usuário. A busca irá retornar somente arquivos de texto (txt, doc, pdf), vídeo (link referencial) e imagens (jpeg, gif), ficando de fora o padrão MPEG-7 para resultados de vídeo.

Os arquivos serão armazenados, manualmente, pelo usuário através de um módulo de upload de arquivos, além de que no processo de indexação serão considerados os termos (tags) que o usuário professor atribuir ao artigo, sendo assim não será utilizado nenhuma técnica de anotação automática para os documentos.

Outro fator será a limitação de extensão de consulta, que ficará limitada à busca de sites, imagens e vídeos relacionados a informações da busca, através de sites de busca, por exemplo, o Google, o Bing e o Youtube. Além disso o trabalho terá como foco a Web 2.0. Também não serão utilizados fatores que envolvem busca semântica devido ao curto prazo para o desenvolvimento do trabalho.

4 PROJETO DE SOLUÇÃO

Nesta seção serão apresentadas as definições de técnica e metodologia, o modelo Iconix e seu processo de desenvolvimento, alguns conceitos básicos como UML e Orientação a Objeto (OO), o estudo de caso para validar a proposta de solução e o esquema físico e lógica do projeto de solução.

4.1 DEFINIÇÃO DE TÉCNICA E METODOLOGIA

Fachin (2001, p. 29), afirma que métodos e técnicas se relacionam, mas são distintos.

O método, segundo Garcia (1998, p. 44), representa um procedimento racional e ordenado, constituído por instrumentos básicos, que implica utilizar a reflexão e a experimentação, para proceder ao longo do caminho e alcançar os objetivos preestabelecidos no planejamento da pesquisa.

Já a técnica é o "modo de fazer de forma mais hábil, mais seguro, mais perfeito, algum tipo de atividade, arte ou ofício" (GALLIANO, 1986, p. 6).

4.1.1 Unified modeling language (UML)

"O UML (*Unified Modelling Language*) é uma linguagem diagramática, utilizável para especificação, visualização e documentação de sistemas de software." (SILVA et. al, 2001)

Fowler (2005) ainda complementa, afirmando que o UML ajuda principalmente na descrição e no projeto de sistemas de software construídos utilizando o estilo orientado a objetos (OO).

Silva et. al (2001) afirma que o UML:

> É promovido pelo Object Management Group (OMG), com contribuições e direitos de autoria das seguintes empresas: Hewlett-Packard, IBM, ICON Computing, i-Logix, IntelliCorp, Electronic Data Services, Microsoft, ObjecTime, Oracle, Platinum, Ptech, Rational, Reich, Softeam, Sterling, Taskon A/S e Unisys.

De acordo com Fowler (2005), a UML nasceu da unificação das muitas linguagens gráficas de modelagem orientadas a objetos que floresceram entre os anos oitenta e noventa, surgindo assim o UML em 1997.

Bell (tradução nossa, 2003) afirma que os diagramas UML mais utilizados são: diagrama de casos de uso, diagrama de classe, diagrama de sequência, diagrama de estados, diagrama de atividade, diagrama de componentes e diagrama de implantação.

- Diagrama de caso de uso - Segundo Furlan (1998) tem como objetivo principal descrever os requerimentos funcionais do sistema.
- Diagrama de classe - São usados para mostrar as classes de um sistema e as relações entre elas. (KIMMEL, tradução nossa, 2005)
- Diagrama de sequência - De acordo com Eriksson e Penker (2000), eles são usados para explorar e visualizar a sequência de objetos em interações uns com os outros.
- Diagrama de estado - A idéia é de estudar certos tipos de lógicas que envolvem transições possíveis entre diferentes estados. (FURLAN, 1998)
- Diagrama de atividade – Mostra o fluxo processual de controle entre dois ou mais objetos de classe durante o processamento de uma atividade. (BELL, tradução nossa, 2003)
- Diagrama de componentes – Eles analisam e gerenciam dependências entre componentes ou entre interfaces de componentes, estes podem ser arquivos de código fonte, bibliotecas ou programas executáveis. (ERIKSSON; PENKER, 2000)
- Diagrama de implantação – Furlan (2008) explica que esse diagrama tem como propósito mostrar a organização de hardware e a ligação do software aos dispositivos físicos.

4.1.2 Iconix

O ICONIX é um processo de desenvolvimento de software cuja metodologia utilizada é simples e prática, tendo um componente de análise e representação de problemas sólido e eficaz. (ROSSINI, 2007).

Segundo Rosenberg (tradução nossa, 2005, p. 57) o ICONIX é um processo leve e altamente interativo, focando-se em obter o código fonte o mais rápido possível, porém não descartando os benefícios de uma análise inicial e do processo de design.

Rosenberg (tradução nossa, 2005, p. 58) também afirma que dentro do ICONIX tudo tem um propósito primordial e explica alguns elementos que compõem o ICONIX:

- Modelo de casos de uso: Define os requisitos de comportamento;
- Modelo de domínio: Descreve os objetos do mundo real e os relacionamentos;
- Diagrama de Robustez: Explica as exigências de comportamento relacionando-as ao modelo de objeto;
- Diagrama de Sequência: Atribui funções para as classes, mostrando seu comportamento.

"A abordagem do ICONIX permite que sejam usados outros recursos da UML, para complementar os recursos usados nas fases do ICONIX, caso seja necessário." (ROSSINI, 2007).

Com a figura 3.2, podemos visualizar melhor as etapas existentes no ICONIX e como as se relacionam dentro desse processo:

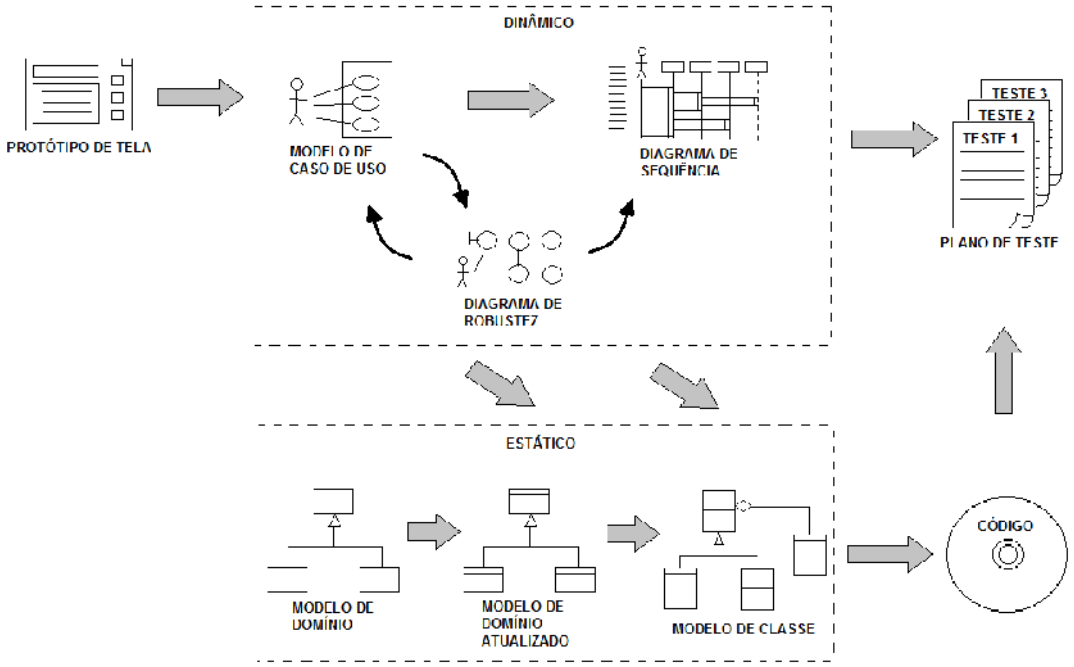

Figura 10 - Visão geral do ICONIX
Fonte: Baseado em Rosenberg (2005, p. 45).

Por fim Rosenberg (tradução nossa, 2005, p.60) apresenta todos os passos do processo ordenadamente:

- 1º Passo: Identificar os objetos do mundo real de domínio (modelagem de domínio);
- 2º Passo: Definir os requisitos comportamentais (casos de uso);
- 3º Passo: Desenvolver o diagrama de robustez com o objetivo de eliminar os casos de uso ambíguos e identificar falhas no modelo de domínio;
- 4º Passo: Definir o comportamento das classes e objetos (diagrama de sequência);
- 5º Passo: Concluir o modelo estático (diagrama de classes);
- 6º Passo: Escrever o código;
- 7º Passo: Executar o sistema e realizar os devidos testes para aceitação do usuário.

4.1.3 Orientação a objeto (OO)

Segundo Jones (1997, p. 52), a orientação a objetos não surgiu de um momento para o outro, mas sim com o decorrer do tempo, através de contribuições de várias pessoas, tanto na parte teórica, como na prática.

Pode-se afirmar que, "a condição de orientação a objetos desencoraja o desenvolvedor a pensar em uma aplicação da forma hierárquica (ou seja, de cima para baixo, funcionalmente decomposta), mas incentiva a pensar em componentes de forma plana e reutilizável". (FEDELI, 2002, p. 17).

A seguir são listadas as principais vantagens da orientação a objetos no desenvolvimento segundo Meyer (1996, p.8):

- Abstração de dados: não é necessário conhecer os detalhes da implementação de uma determinada classe para que possamos utiliza-la dentro de outra classe;
- Compatibilidade: podemos combinar componentes de software facilmente devido as heurísticas para construção de classes e suas interfaces;
- Flexibilidade: um conjunto de classes pode delimitar unidades naturais para realizar determinada tarefa dentro do desenvolvimento de software;
- Reutilização: podemos reutilizar códigos através do encapsulamento de métodos e representação de dados na construção de classes;
- Extensibilidade: é mais fácil de entender um software que foi construído utilizando técnicas de orientação a objetos devido a herança e as classes que formam uma estrutura fracamente acoplada, facilitando alterações;
- Manutenção: através da modularidade natural da estrutura de classe e a facilidade de compreensão do código, a manutenção se torna um processo mais simples.

Mesmo tendo várias vantagens, a orientação a objetos, também, apresenta algumas desvantagens como, problemas com o trabalho em equipe no momento do desenvolvimento, dificuldade para se familiarizar com a abordagem, os métodos tradicionais de análise e projeto são inapropriados para esta condição, além de não serem suficientemente abstratos quando expressos em termos de programação, e quando se tem uma introdução de

novas sistemáticas de trabalho, necessita-se de novas formas de gerenciamento. (MEYER, 2002, p. 9).

4.2 MODELAGEM DO SISTEMA PROPOSTO

Nesta seção, é apresentada a modelagem do sistema proposto, onde inicialmente são apresentados os atores envolvidos na utilização do sistema em questão. Depois de ilustrar os atores são listados os requisitos funcionais e não funcionais do protótipo.

Como foi afirmado na seção anterior, este trabalho utiliza como metodologia o Iconix para a modelagem do sistema, tendo em vista isso, após identificar os atores e requisitos do sistema são respeitadas todas as interações previstas na metodologia, sendo estas: a confecção dos protótipos de interface, seguido da criação dos casos de uso, gerando assim subsídios para o diagrama de domínio.

Após as interações iniciais é construído para cada caso de uso o seu respectivo diagrama de robustez e de sequencia, dando subsídios para a construção do diagrama de classes concreta.

4.2.1 ATORES

1. Administrador: terá permissão de buscar, criar, alterar e excluir os perfis de professor ou aluno, podendo também excluir e alterar artigos publicados pelo usuário professor, porém não terá acesso a funcionalidade de criar um artigo, pois seu foco é a administração geral do sistema.

2. Professor: terá a permissão de publicar, alterar e excluir um artigo, além de poder adicionar *tags*/termos aos artigos cadastrados com o fim de otimizar sua busca.

3. Aluno: terá somente a permissão de poder realizar a busca por artigos cadastrados, não podendo criar artigos ou adicionar tags aos artigos criados.

4.2.2 REQUISITOS

Segundo Sommerville (2007, p. 79), “os requisitos de um sistema são descrições dos serviços fornecidos pelo sistema e as suas restrições operacionais. Esses requisitos refletem as necessidades dos clientes de um sistema que ajuda a resolver algum problema”.

4.2.2.1 REQUISITOS FUNCIONAIS

Os requisitos funcionais “são as declarações de serviços que o sistema deve fornecer, como o sistema deve reagir a entradas específicas e como o sistema deve se comportar em determinadas situações”. (SOMMERVILLE, 2007, p. 80).

- RF001 - O acesso ao sistema deve ser feito através da validação de login e senha;
- RF002 – Os usuários do sistema devem ser classificados através dos perfis administrador, professor e aluno;
- RF003 - O sistema deve permitir que um usuário administrador altere o perfil de outro usuário (professor/aluno);
- RF004 - O sistema deve permitir que um usuário administrador localize um outro usuário utilizando os filtros de nome, perfil e e-mail;
- RF005 - O sistema deve permitir que o administrador edite dados dos usuários cadastrados;
- RF006 - O sistema deve permitir que um usuário administrador exclua um usuário cadastrado;
- RF007 - O sistema deve permitir que um usuário administrador visualize, edite ou exclua um conteúdo criado por um usuário professor;

- RF008 - O sistema deve permitir que um usuário professor cadastre um conteúdo;
- RF009 - O sistema deve permitir que um usuário professor altere ou exclua um conteúdo criado;
- RF010 - O sistema deve permitir que o usuário professor adicione tags aos conteúdos cadastrados;
- RF011 - O sistema deve permitir que o usuário crie sua própria conta (cadastro), porém o usuário criado terá o perfil de aluno, podendo ser alterado apenas por um usuário administrador;
- RF012 - O sistema deve permitir que todos os usuários realizem buscas por conteúdos cadastrados pelo usuário professor, tendo como parâmetros termos literais;
- RF013 - As buscas por conteúdo devem apresentar como resultado arquivos de vídeo (mpeg e links referenciando vídeos do youtube) referente à consulta efetuada;
- RF014 - As buscas por conteúdo devem apresentar como resultado arquivos de imagem (jpeg e gif) referente à consulta efetuada;
- RF015 - As buscas por conteúdo devem apresentar como resultado arquivos de texto referente à consulta efetuada, assim como o próprio conteúdo cadastrado;
- RF016 - O extensor de busca do sistema deve ser limitado a varrer uma lista de sites pré-determinados;
- RF017 - O sistema deve apresentar de forma separada os tipos de dados que foram retornados no resultado da busca por conteúdo.

4.2.2.2 REQUISITOS NÃO FUNCIONAIS

Requisitos não funcionais “são restrições sobre os serviços ou as funções oferecidas pelo sistema. Eles incluem restrições de timing, restrições sobre o processo de desenvolvimento e padrões”. (SOMMERVILLE, 2007, p. 80).

- RNF001 - O sistema deve apresentar uma interface limpa tendo como página inicial a tela de login;

- RNF002 - Em qualquer operação de busca do sistema o resultado deve ser apresentado em menos de 3 segundos;
- RNF003 - Em qualquer operação de busca do sistema, se o resultado da busca não retornar nenhum dado, uma mensagem deve ser mostrada ao usuário;
- RNF004 - O sistema deve permitir acesso simultâneo de usuários;
- RNF005 - Na operação de busca de artigos, ao adicionar uma tag ao arquivo retornado, o mesmo deve também ficar disponível por este termo;
- RNF006 - O sistema deve permitir que usuários selecionem qualquer arquivo retornado e possam executar/visualizar este;
- RNF007 - A listagem de arquivos deve ser feita por ordem alfabética da descrição do arquivo;
- RNF008 - O sistema deve apresentar interface amigável;
- RNF009 - O sistema deve permitir isolar visualização de resultados da busca por tipo;
- RNF010 - Antes de excluir qualquer registro deverá ser apresentada uma mensagem de confirmação ao usuário.

4.2.2.3 REGRAS DE NEGÓCIO

Segundo Ross (2003), regras de negócio (RNs) são diretivas cujo objetivo é influenciar ou guiar o comportamento de um negócio. Já no ponto de vista de Wiegers (2006) as RNs são políticas da companhia, padrões de indústria ou leis e regulamentações governamentais que definem, restringem ou governam alguns comportamentos de um negócio.

- RN001 - Poderá existir apenas uma única conta de usuário por e-mail;
- RN002 - Qualquer usuário só poderá alterar sua senha após informar a sua senha atual;
- RN003 - Somente usuários com o perfil professor ou administrador poderão alterar ou excluir os conteúdos cadastrados no sistema;
- RN004 - Os usuários administradores podem alterar um perfil aluno ou professor para o perfil administrador, porém não é permitida a operação inversa;

- RN005 - O usuário administrador não pode editar ou excluir uma conta de outro usuário administrador;
- RN006 - Ao trocar os dados da conta o usuário não poderá alterar o e-mail de cadastro.

4.2.3 PROTÓTIPOS DE TELA

Esta seção serão apresentas todos os protótipos de tela do sistema, permitindo visualizar melhor onde cada operação ocorre, além de auxiliar no desenvolvimento da interface com o usuário no decorrer do desenvolvimento.

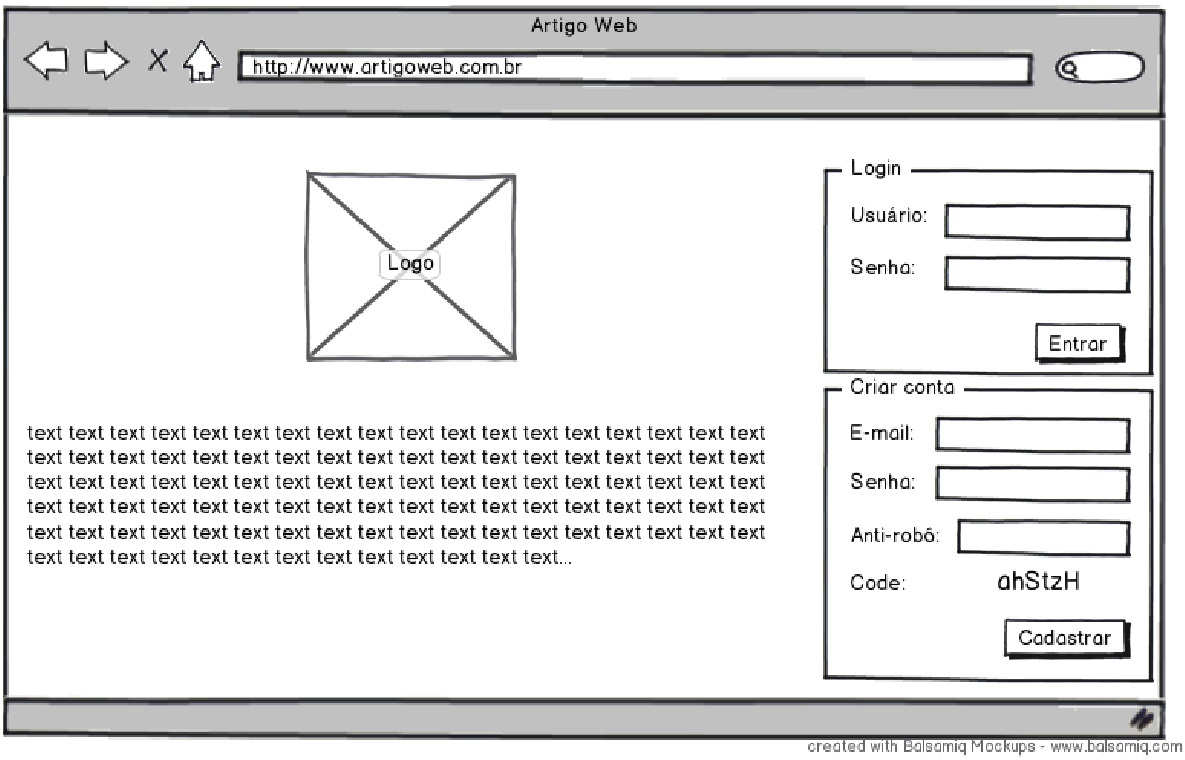

Figura 11 - Tela inicial
Fonte: Autores

Na tela inicial o usuário irá realizar o cadastro no sistema, além de efetuar o login no mesmo, através do e-mail e senha cadastrados.

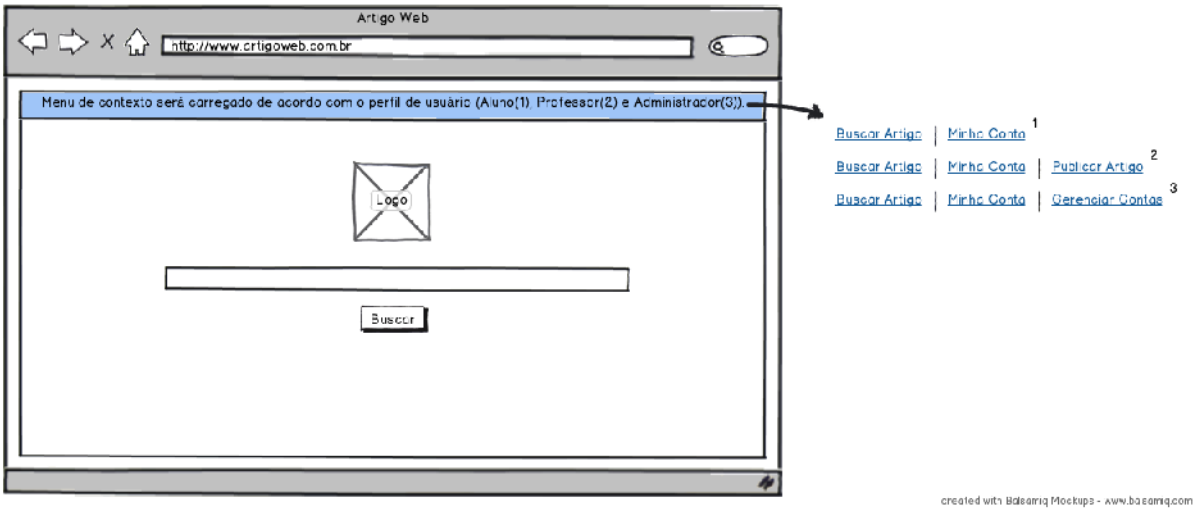

Figura 12 - Tela de busca
Fonte: Autores

Na tela de busca o usuário irá informar o termo para realizar a busca. Nota-se a existência de um menu na parte superior da tela. Nesse menu irá servir para a navegação dentro do sistema, sendo que as opções que irão aparecer menu dependem do perfil do usuário logado no sistema. O perfil aluno terá as opções de ir para a tela de busca ou ir para a tela de configuração de conta, já o perfil de professor terá as opções ir para a tela de busca, ir para a tela de configuração de conta e ir para a tela de publicação de artigos, e por fim o perfil administrador terá as opções de ir para a tela de busca, ir para a tela de configuração de conta e ir para a tela de gerenciamento de contas.

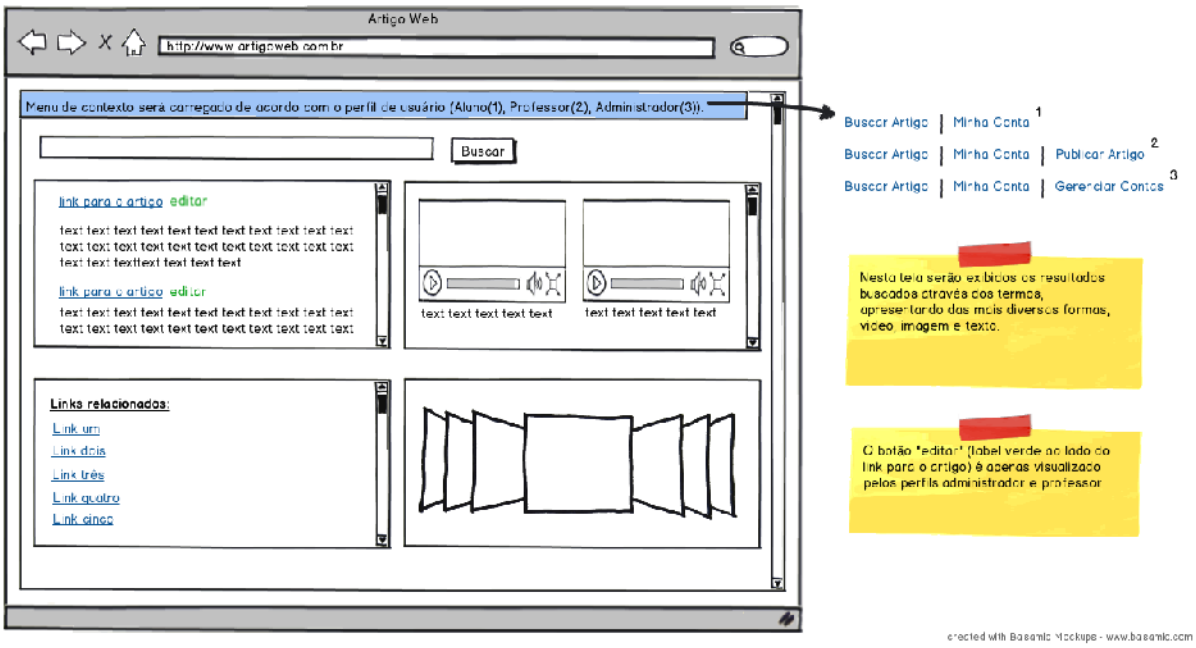

Figura 13 - Tela de resultados
Fonte: Autores

A tela de resultados irá apresentar os resultados textuais e multimídia da busca, além de permitir uma nova consulta através do campo de busca na parte superior da tela. A localização e descrição de cada tipo de resultado são apresentadas a seguir:

- Superior esquerdo: são apresentados os documentos que foram indexados e um breve resumo deles, além da opção de edição para os perfis professor e administrador;
- Superior direito: são apresentados vídeos relacionados aos documentos retornados da busca;
- Inferior esquerdo: são apresentados links de páginas da WEB relacionados aos documentos retornados da busca;
- Inferior direito: são apresentadas imagens relacionadas aos documentos retornados da busca.

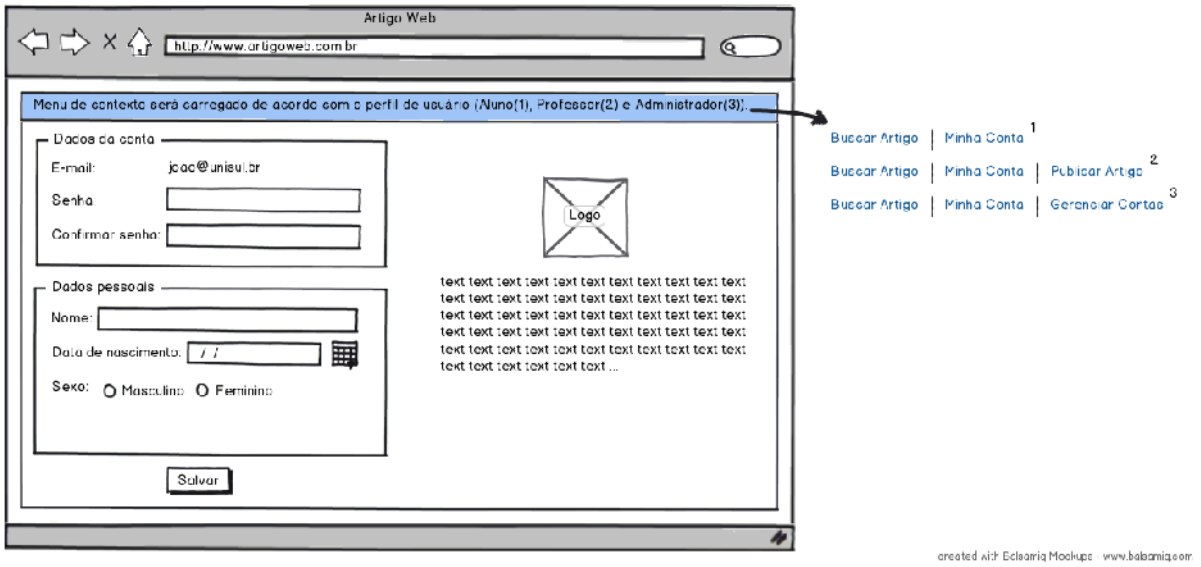

Figura 14 - Tela de configuração de dados da conta
Fonte: Autores

Na tela de configuração de conta o usuário poderá definir outra senha para sua conta e adicionar seus dados pessoais.

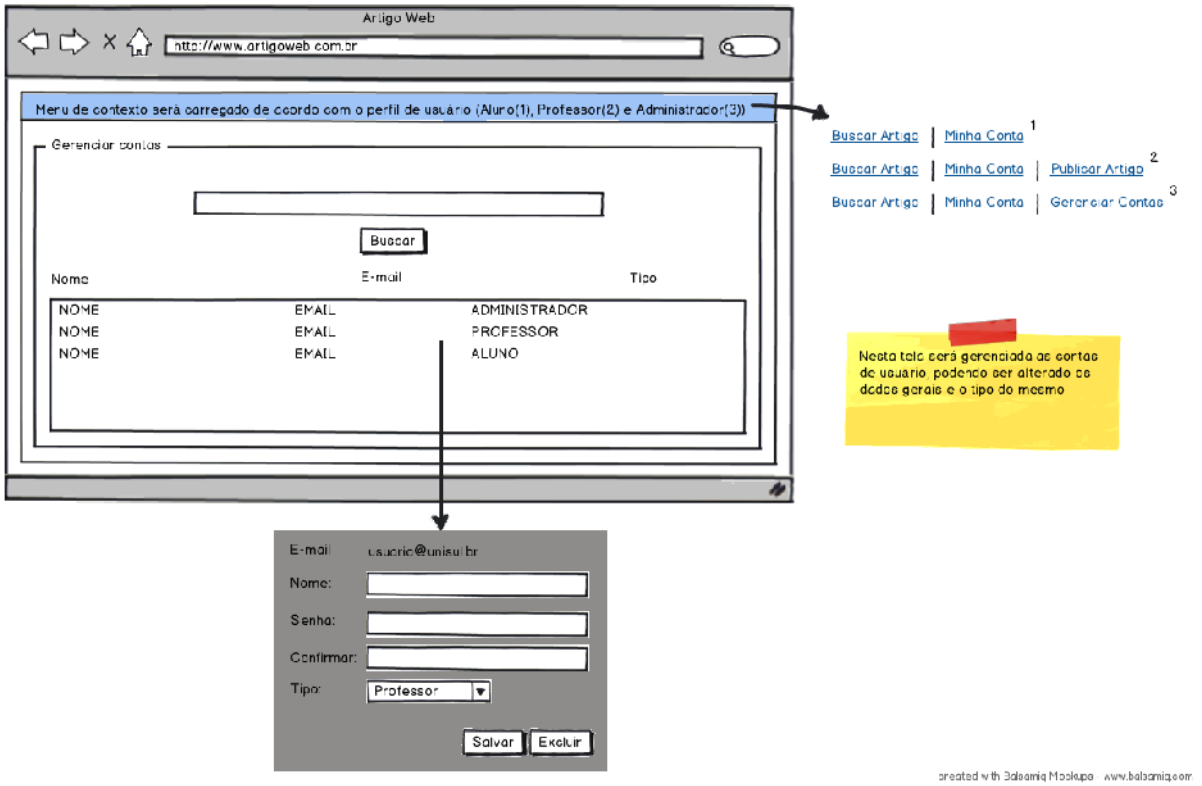

Figura 15 - Tela de configuração de contas (Administrador)
Fonte: Autores

Na tela de configuração de contas o usuário administrador poderá visualizar todos os usuários do sistema, além de poder alterar os dados dos usuários com perfil aluno ou professor, assim como alterar seus perfis, lembrando que um usuário administrador não poderá altear os dados ou o perfil de outro usuário administrador.

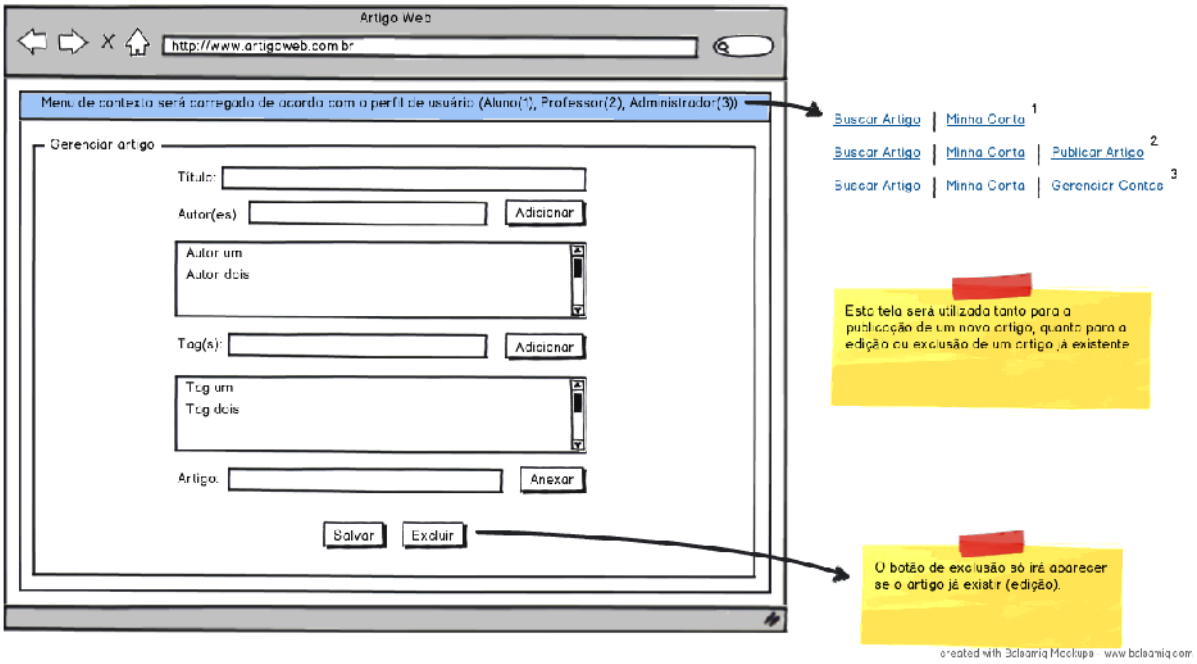

Figura 16 - Tela de gerenciamento de artigos (Professor e Administrador)
Fonte: Autores

A tela de gerenciamento de artigos será responsável pela publicação e edição dos artigos, sendo que apenas os usuários com perfil professor e administrador poderão visualizar esta tela, lembrando que o usuário administrador poderá apenas editar um artigo. Nesta tela o usuário poderá realizar o upload do artigo, além de poder adicionar e remover tags e autores ao artigo.

4.2.4 CASOS DE USO PRIMÁRIO

Nesta seção serão apresentados os casos de uso primário do sistema, mostrando os possíveis fluxos principais e alternativos das operações para cada perfil de usuário.

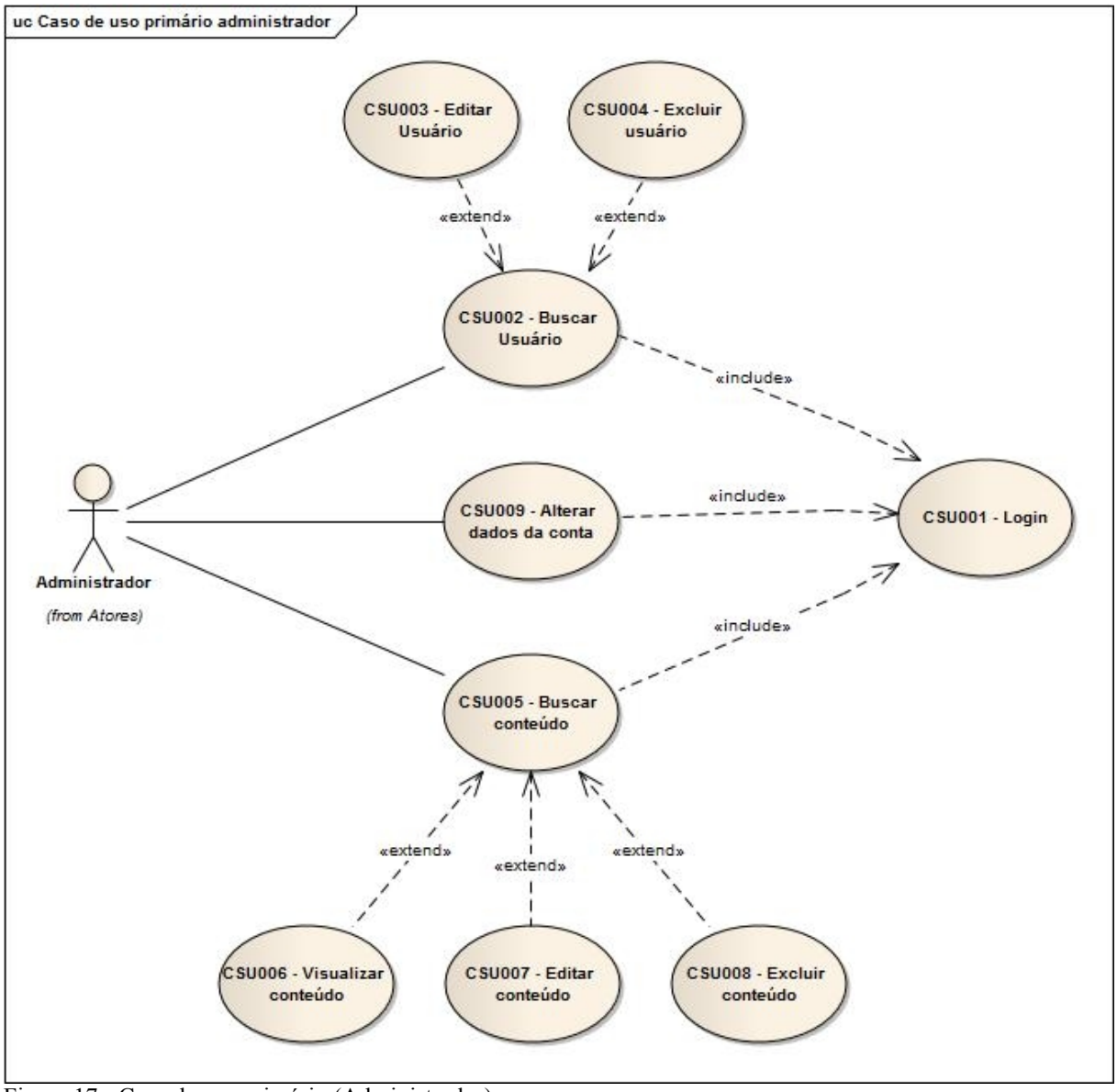

Figura 17 - Caso de uso primário (Administrador)
Fonte: Autores

CSU003 - Fluxo principal (Editar Usuário)

1. Escolhe a opção “gerenciar contas”;
2. Realiza a busca de um usuário;
3. Seleciona o usuário desejado (RN004, RN005);
4. Edita os dados do usuário (RN006);
5. Seleciona a opção “salvar”;
6. O sistema mostra uma mensagem de confirmação.

CSU004 - Fluxo alternativo (Excluir Usuário)

1. Escolhe a opção de gerenciamento de usuários;
2. Realiza a busca de um usuário;
3. Seleciona o usuário desejado (RN005);
4. Seleciona a opção “excluir”;
5. O sistema mostra uma mensagem de confirmação.

CSU009 - Fluxo alternativo (Alterar dados da conta)

1. Escolhe a opção “minha conta”;
2. Edita os dados da conta (RN002, RN006);
3. Seleciona a opção “salvar”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU007 - Fluxo alternativo (Editar conteúdo)

1. Informa o termo para a busca;
2. Seleciona o artigo desejado na tela de resultados e escolhe a opção de editar (RN003);
3. Edita os dados do artigo;
4. Seleciona a opção “salvar”;
5. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU008 - Fluxo alternativo (Excluir conteúdo)

1. Informa o termo para a busca;
2. Seleciona o artigo desejado na tela de resultados e escolhe a opção de editar (RN003);
3. Seleciona a opção de “excluir”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU006 - Fluxo alternativo (Visualiza conteúdo)

1. Informa o termo para a busca na tela de busca;
2. O sistema “retorna” os resultados na tela de resultados;
3. Escolhe um tipo de visualização, ou seleciona um link na própria página de resultados.

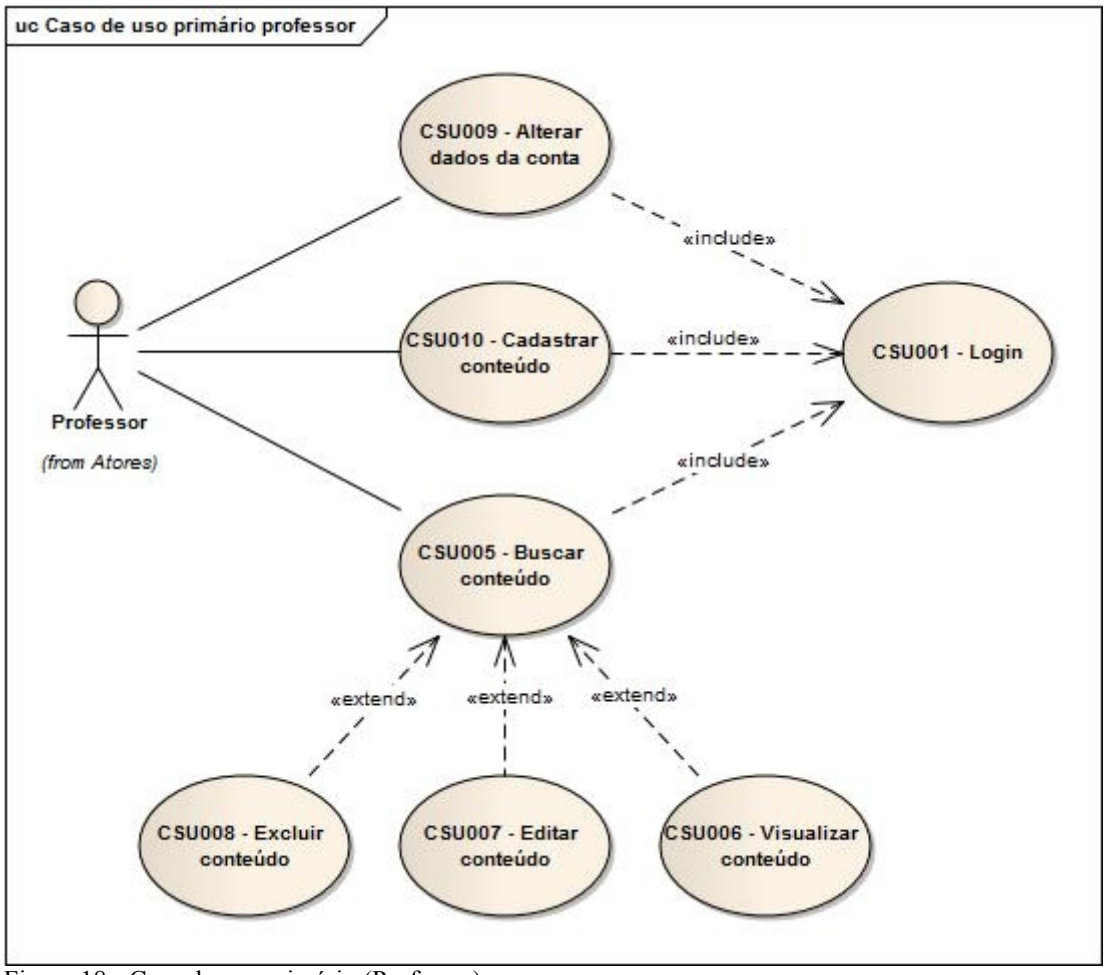

Figura 18 - Caso de uso primário (Professor)
Fonte: Autores

CSU010 - Fluxo principal (Cadastrar conteúdo)

1. Seleciona a opção “publicar artigo”;
2. Preenche os campos solicitados e realiza o upload do arquivo;
3. Seleciona a opção “salvar”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU007 - Fluxo alternativo (Editar conteúdo)

1. Informa o termo para a busca;
2. Seleciona o artigo desejado e escolhe a opção de editar (RN003);
3. Edita os dados do artigo;
4. Seleciona a opção “salvar”;
5. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU008 - Fluxo alternativo (Excluir conteúdo)

1. Informa o termo para a busca;
2. Seleciona o artigo desejado e escolhe a opção de editar (RN003);
3. Seleciona a opção de “excluir”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

CSU006 - Fluxo alternativo (Visualiza conteúdo)

1. Informa o termo para a busca;
2. O sistema retorna os resultados na tela de resultados;
3. Escolhe um tipo de visualização, ou seleciona um link na própria página de resultados.

CSU009 - Fluxo alternativo (Alterar dados da conta)

1. Escolhe a opção “minha conta”;
2. Edita os dados da conta (RN002, RN006);
3. Seleciona a opção “salvar”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

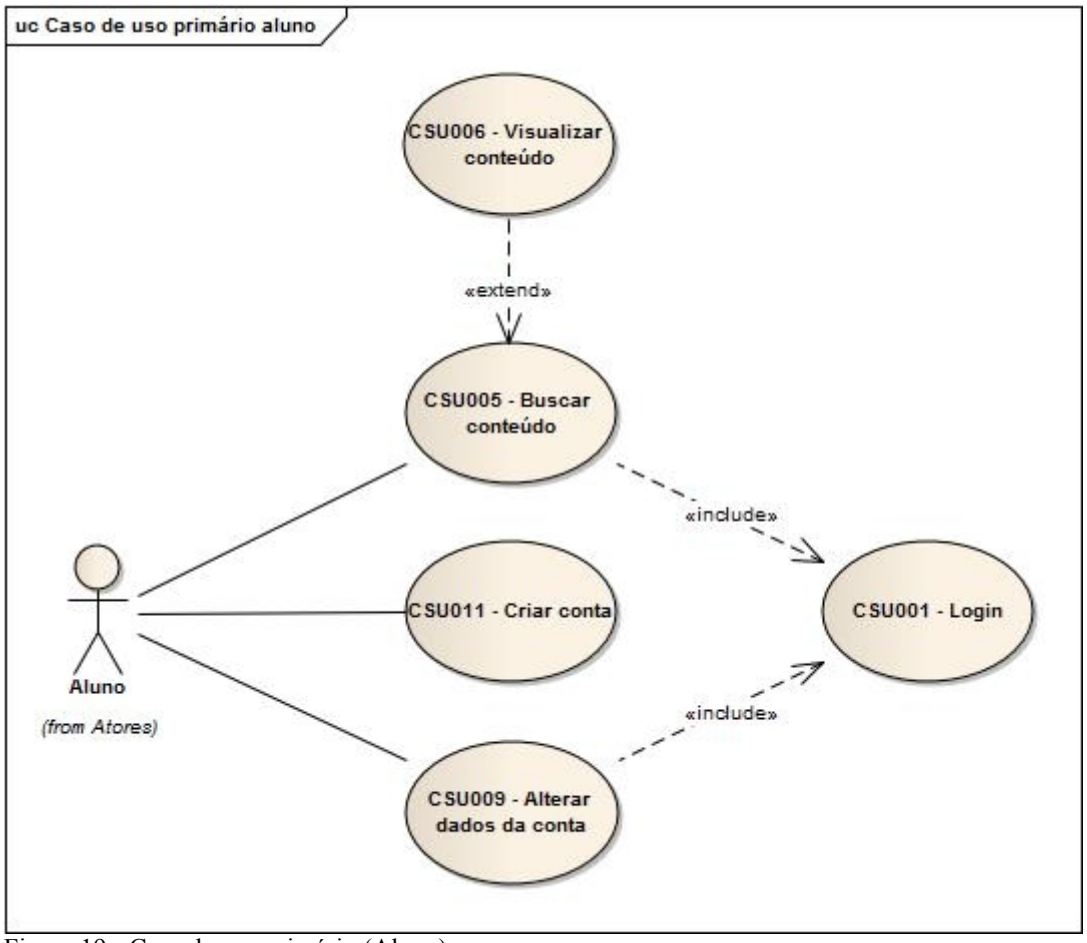

Figura 19 - Caso de uso primário (Aluno)
Fonte: Autores

CSU006 - Fluxo principal (Visualiza conteúdo)

1. Informa o termo para a busca;
2. O sistema “retorna” os resultados na tela de resultados;
3. Escolhe um tipo de visualização, ou seleciona um link na própria página de resultados.

CSU011 – Fluxo alternativo (Criar conta)

1. Informa os dados do cadastro;
2. Seleciona a opção “salvar”;
3. O sistema mostra uma mensagem de confirmação, e envia um e-mail de confirmação para o endereço de e-mail cadastrado.

CSU009 - Fluxo alternativo (Alterar dados da conta)

1. Escolhe a opção “minha conta”;
2. Edita os dados da conta (RN002, RN006);

3. Seleciona a opção “salvar”;
4. O sistema volta para a tela de busca, e mostra uma mensagem de confirmação.

4.2.5 MODELO DE DOMÍNIO

Nessa seção o modelo de domínio é apresentado, permitindo visualizar os dados do protótipo de uma maneira geral.

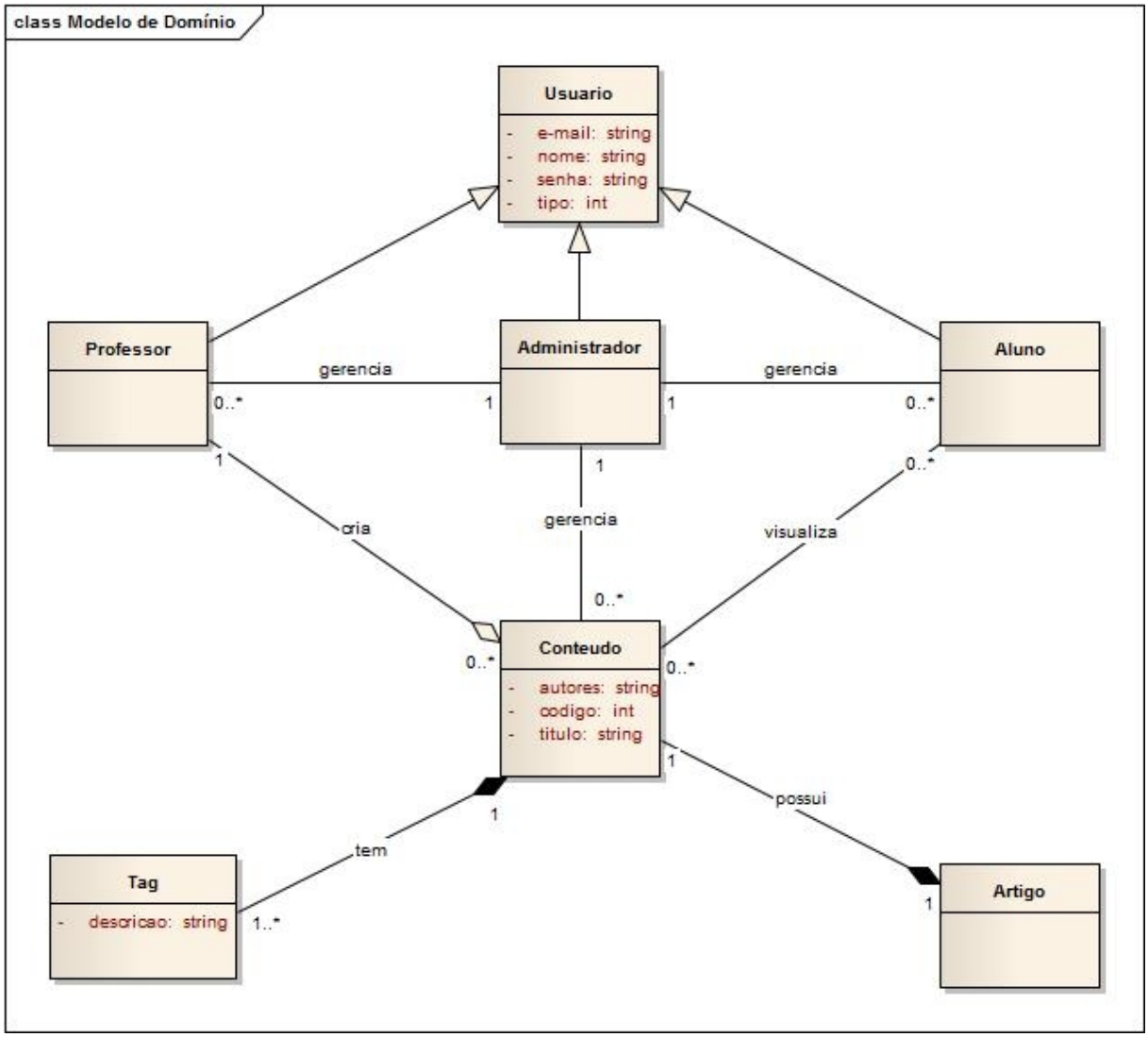

Figura 20 - Modelo de domínio
Fonte: Autores

4.2.6 DIAGRAMA DE ROBUSTEZ

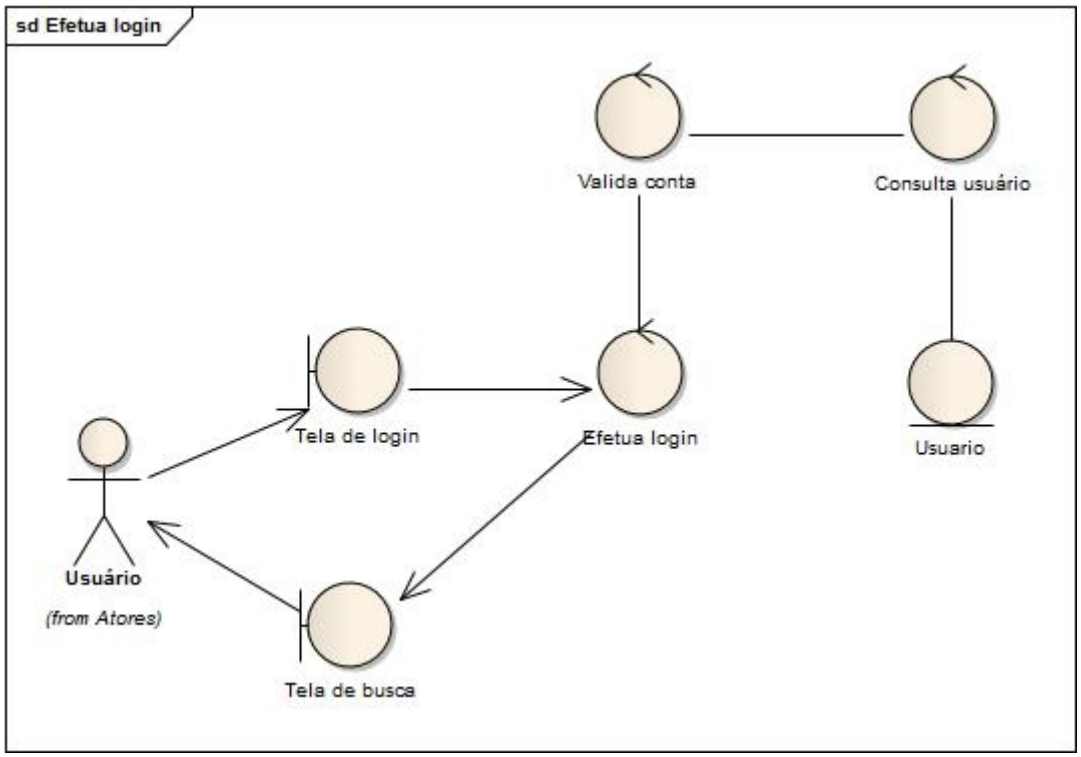

Figura 21 - Robustez: Efetua login
Fonte: Autores

Na operação de login o usuário irá informar login e senha para o sistema, que irá fazer a validação do mesmo. O sistema busca o usuário cadastrado no banco, e faz a verificação para, posteriormente, realizar a autenticação e o redirecionamento para a tela de busca.

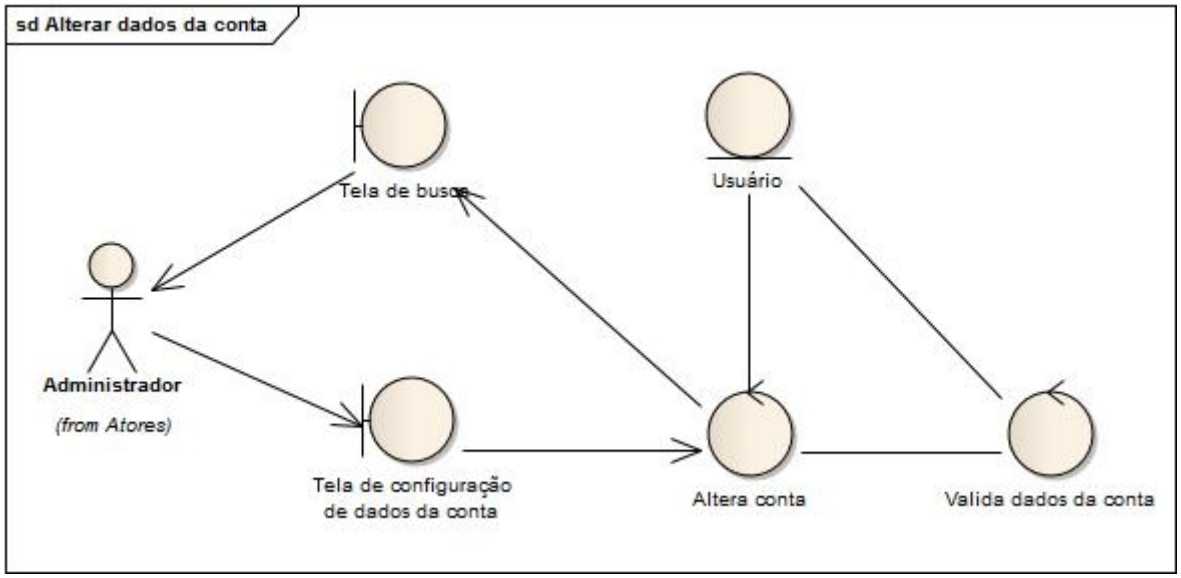

Figura 22 - Robustez: Alterar dados da conta (Administrador)
Fonte: Autores

Na operação de alterar dados da conta, o usuário irá informar novas informações, ou trocar as informações já existentes na tela de configuração de conta. Ao confirmar as alterações, o sistema irá realizar as validações das informações e alterar os dados da conta, buscando o usuário, que está logado no sistema, salvando, novamente, as informações e realizando o redirecionamento para a tela de busca.

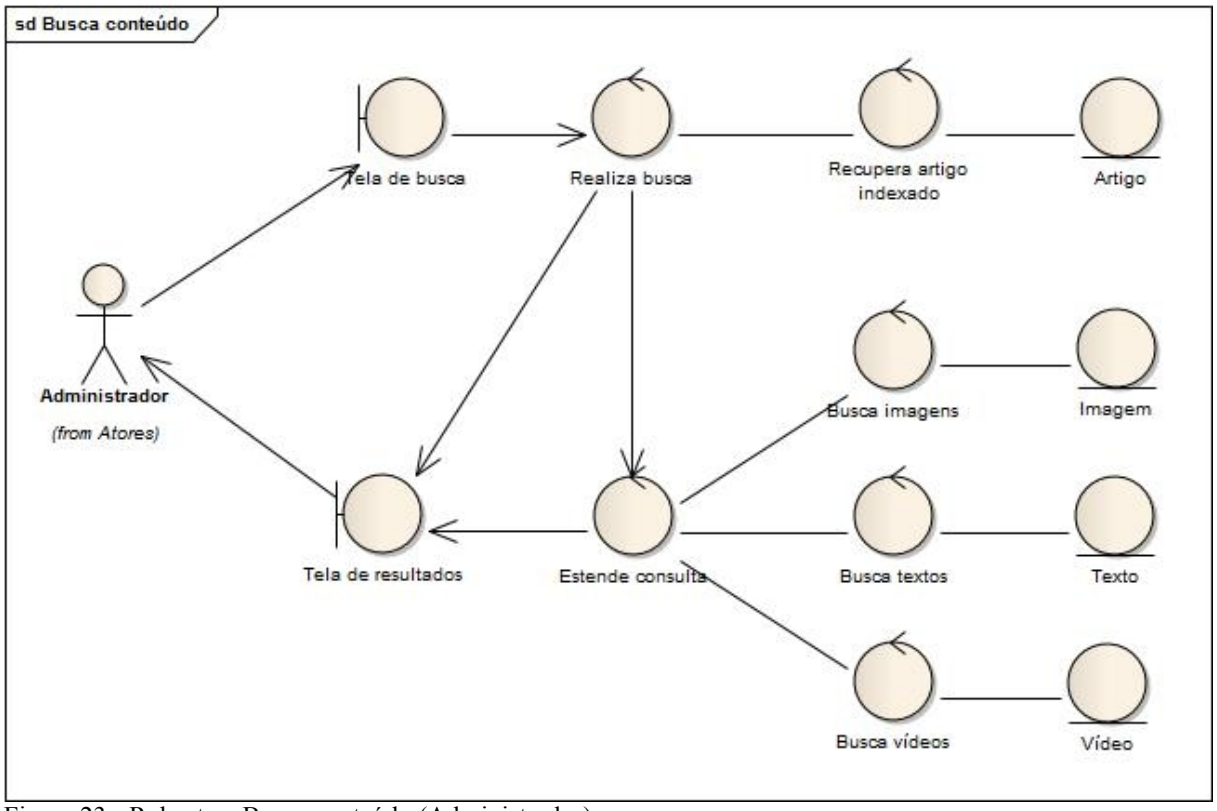

Figura 23 - Robustez: Busca conteúdo (Administrador)
Fonte: Autores

Na operação de busca do sistema, o usuário irá informar um termo na tela de busca. O sistema irá recuperar o artigo indexado, além de trazer imagens, textos e vídeos referentes ao termo informado, sendo que estas informações serão mostradas na tela de resultados.

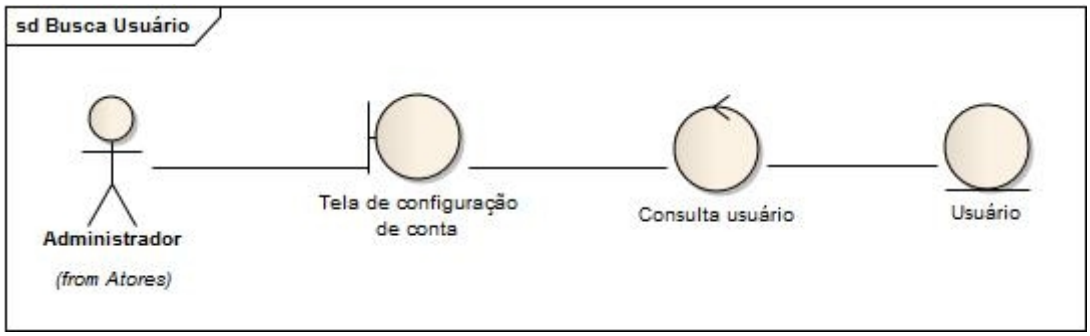

Figura 24 - Robustez: Busca usuário (Administrador)
Fonte: Autores

Na operação de busca de usuários, o usuário administrador deve informar o nome ou parte do nome de um usuário para o sistema, na tela de configuração de conta. O sistema irá realizar a busca, "retornando" uma lista contendo os usuários desejados.

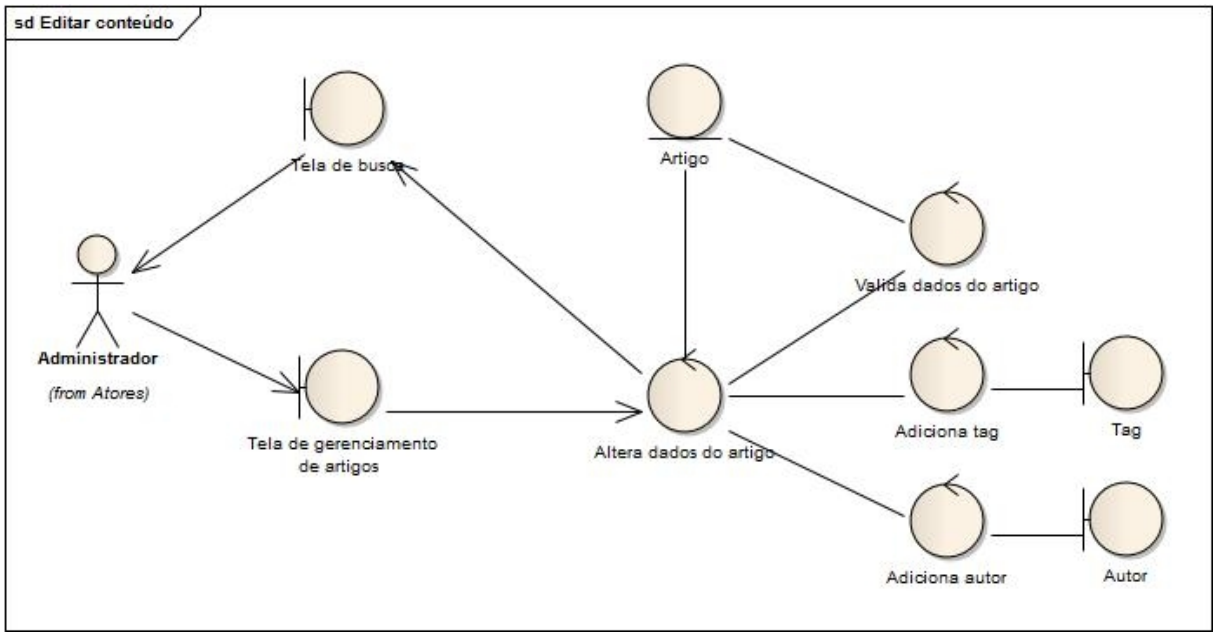

Figura 25 - Robustez: Editar conteúdo (Administrador)
Fonte: Autores

Na operação de edição do conteúdo, o usuário apresenta novas informações ou troca as informações já existentes do conteúdo (artigo). Ao confirmar as alterações, o sistema irá realizar as validações das informações e alterar os dados do conteúdo (artigo), buscando o conteúdo que foi selecionado na tela de resultados do sistema, salvando, novamente, as informações, e realizando o redirecionamento para a tela de busca.

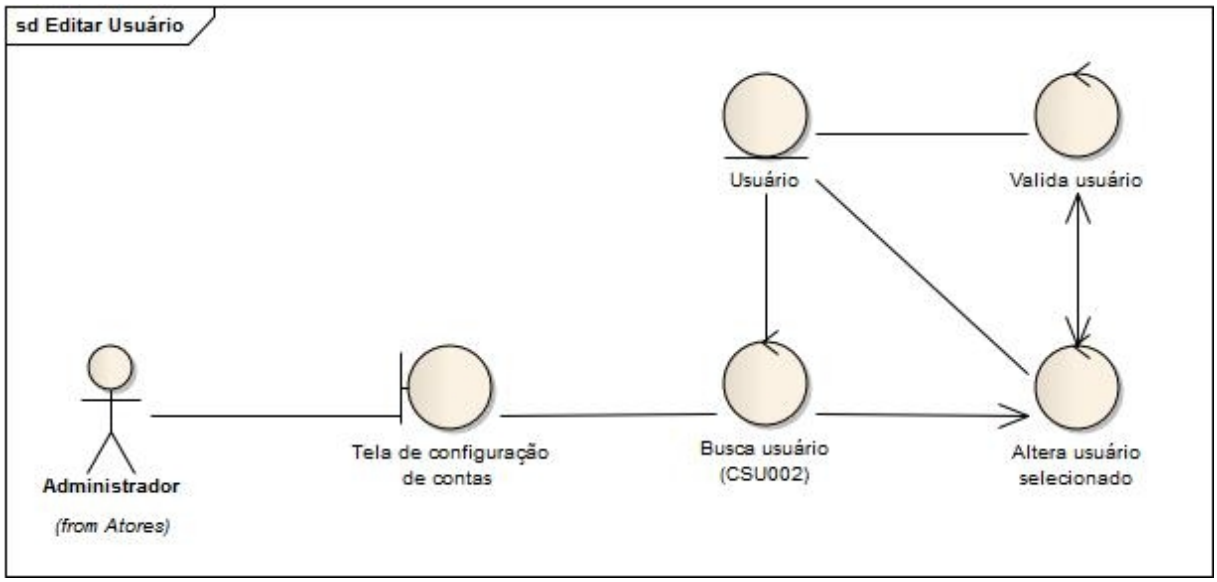

Figura 26 - Robustez: Editar usuário (Administrador)
Fonte: Autores

Para a edição de um usuário, o administrador deverá selecionar um usuário na lista dos usuários retornados pela busca na tela de configuração de contas. O administrador

poderá alterar os dados da conta selecionada, e o sistema irá validar as informações alteradas, e caso a validação esteja correta, o sistema irá salvá-las.

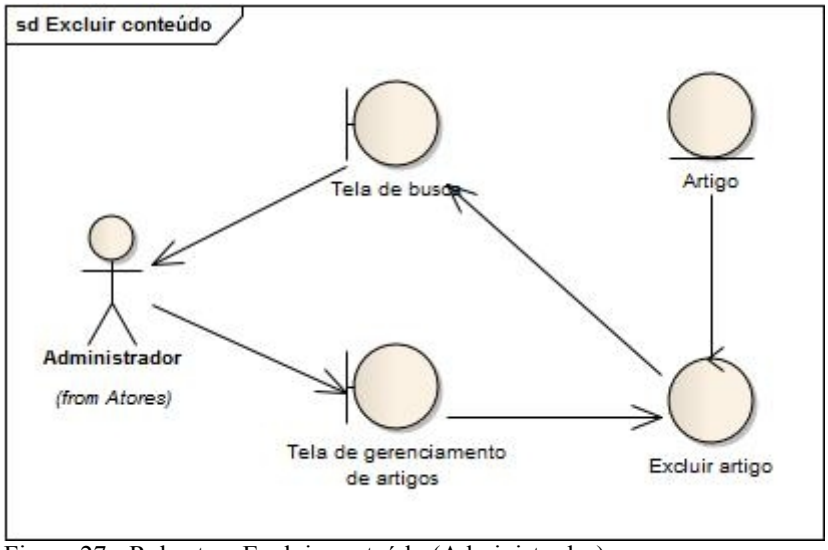

Figura 27 - Robustez: Excluir conteúdo (Administrador)
Fonte: Autores

Na operação de exclusão de conteúdo, o usuário irá excluir um conteúdo (artigo). O sistema irá excluir o artigo selecionado, e irá redirecionar para a tela de busca.

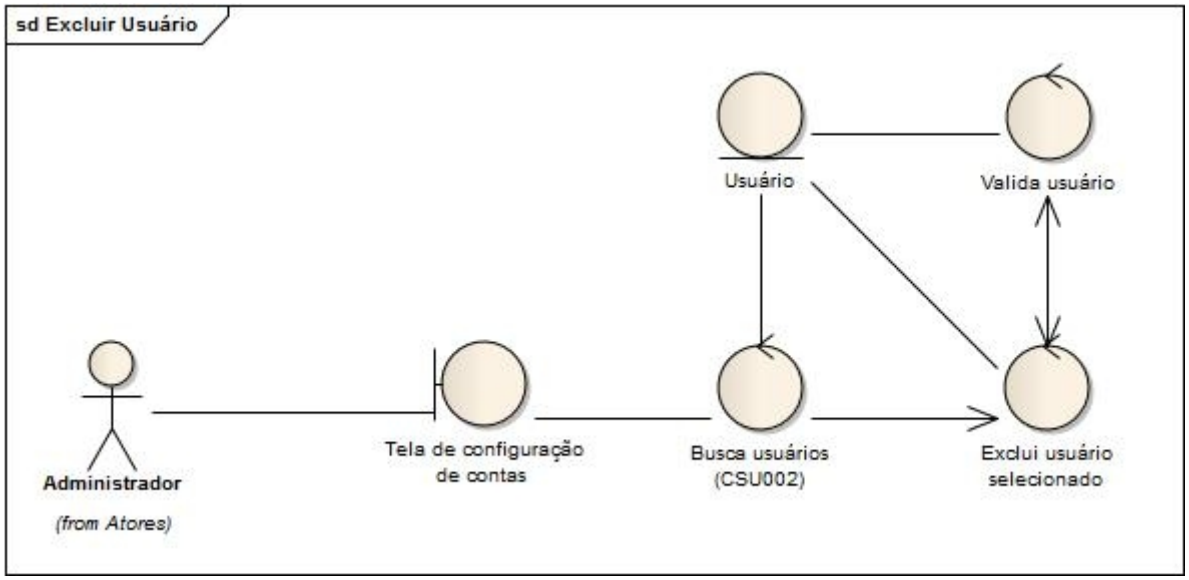

Figura 28 - Robustez: Excluir usuário (Administrador)
Fonte: Autores

Para a exclusão de um usuário, o administrador deverá selecionar um usuário, na lista dos usuários retornados pela busca na tela de configuração de contas. O sistema, então, irá verificar se o usuário poderá ser excluído, para que a exclusão seja feita.

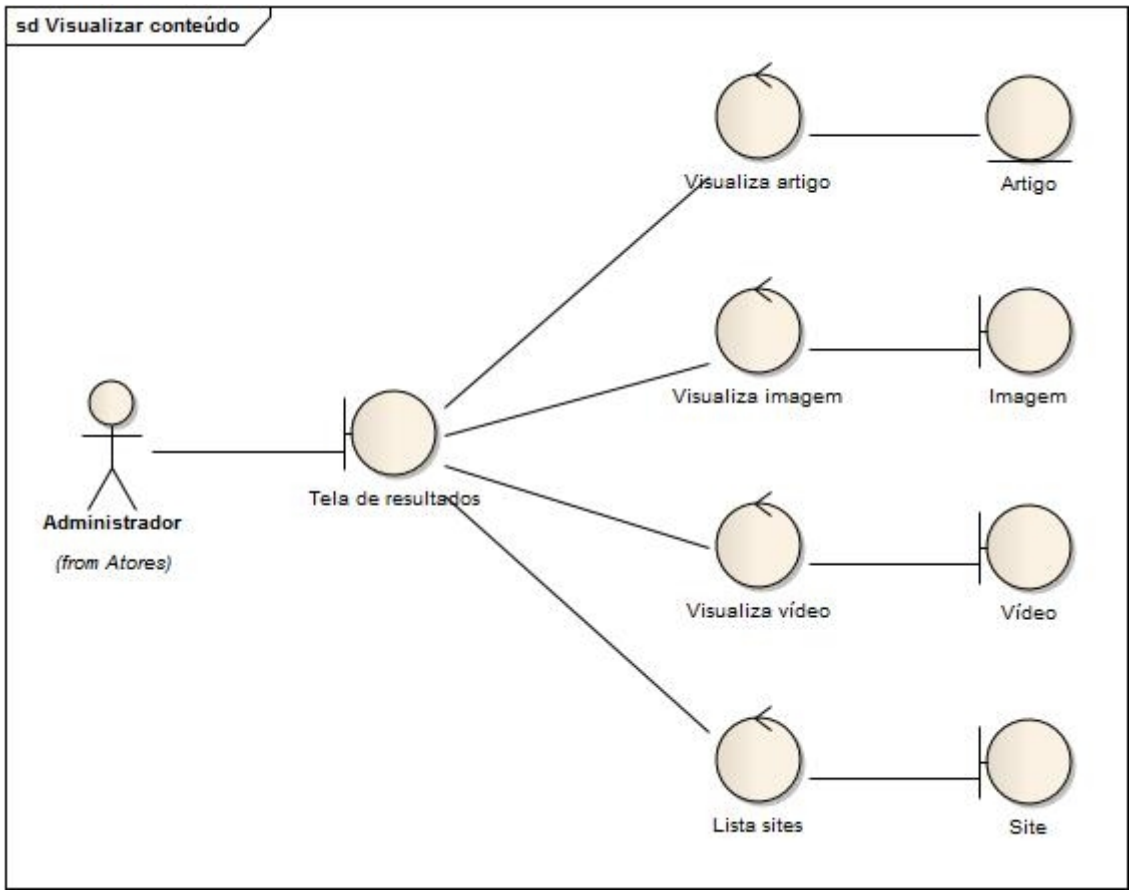

Figura 29 - Robustez: Visualizar conteúdo (Administrador)
Fonte: Autores

Após a operação de busca, o usuário poderá visualizar as informações referentes ao termo informado na busca. A tela de resultados irá mostrar imagens, textos e vídeos referentes aos artigos encontrados, além de trazer links para os mesmos artigos encontrados.

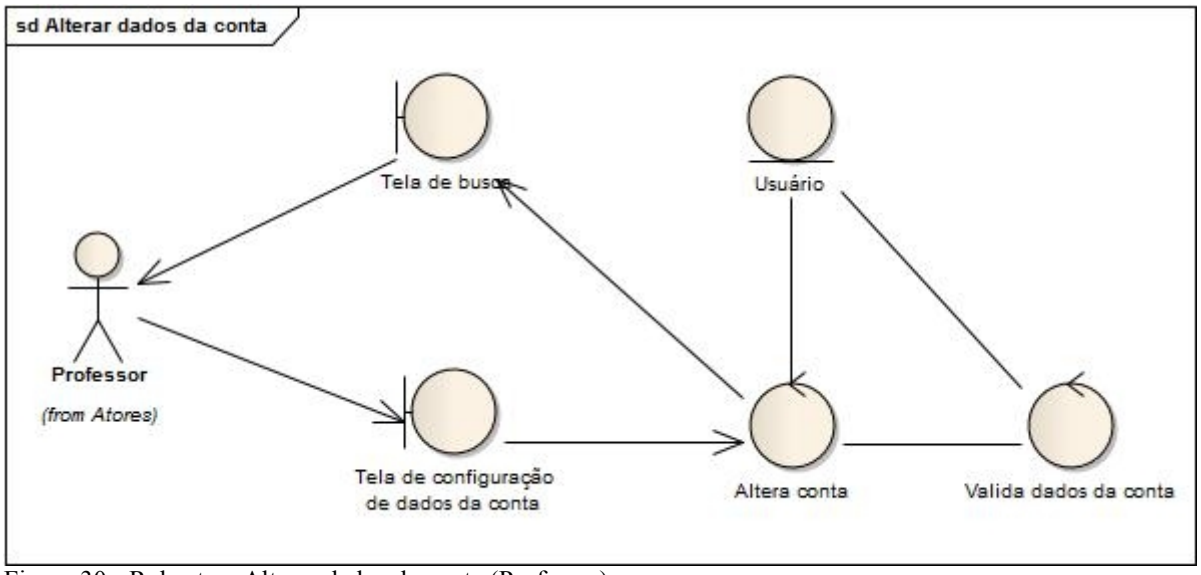

Figura 30 - Robustez: Alterar dados da conta (Professor)
Fonte: Autores

Na operação de alterar dados da conta, o usuário irá apresentar novas informações, ou trocar as informações já existentes na tela de configuração de conta. Ao confirmar as alterações, o sistema irá realizar as validações das informações, e alterar os dados da conta, buscando o usuário, que está logado no sistema, salvando, novamente, as informações e realizando o redirecionamento para a tela de busca.

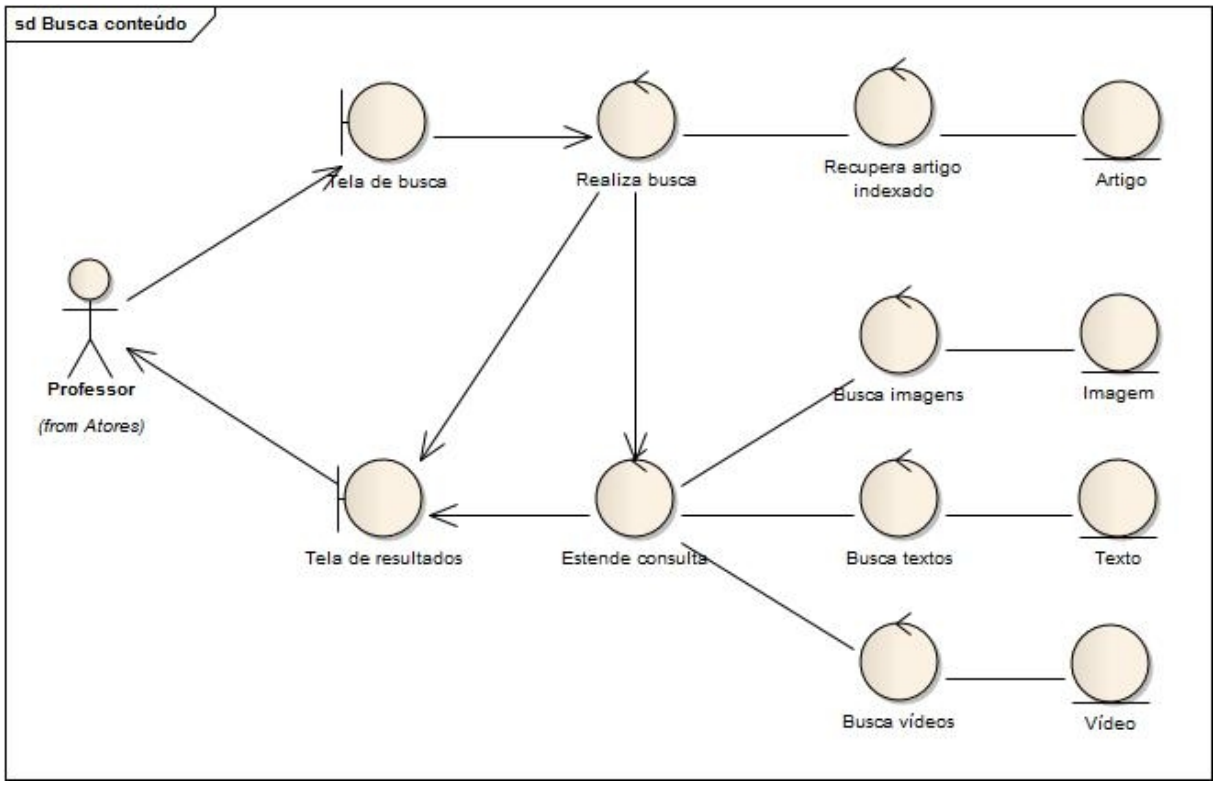

Figura 31 - Robustez: Busca conteúdo (Professor)
Fonte: Autores

Na operação de busca do sistema, o usuário irá apresentar um termo na tela de busca. O sistema há de recuperar o artigo indexado, além de trazer imagens, textos e vídeos referentes ao termo informado, sendo que estas informações serão mostradas na tela de resultados.

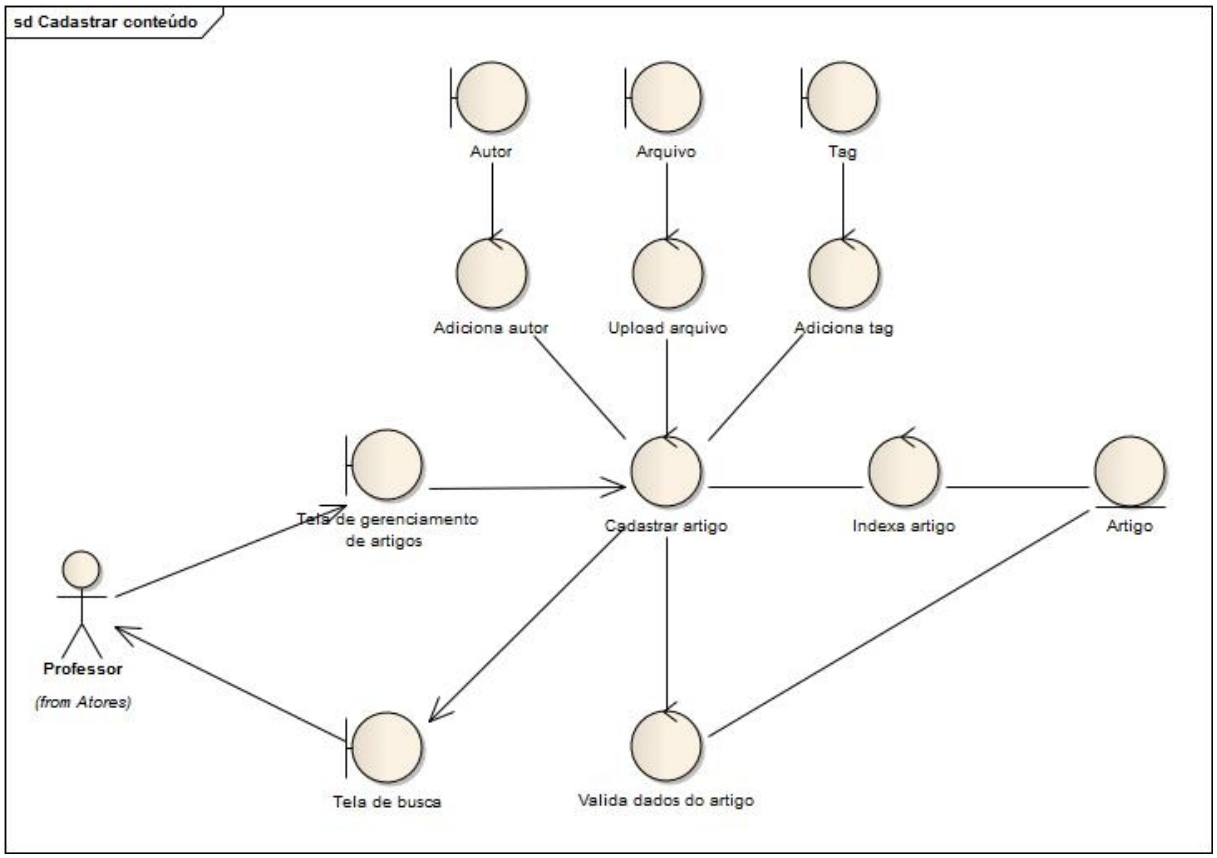

Figura 32 - Robustez: Cadastrar conteúdo (Professor)
Fonte: Autores

Na operação de cadastrar conteúdo, o usuário professor informará os dados do conteúdo, juntamente com o próprio conteúdo, que será anexado no cadastro. O sistema validará as informações, e, caso a validação esteja correta, o sistema salvará o conteúdo indexando-o.

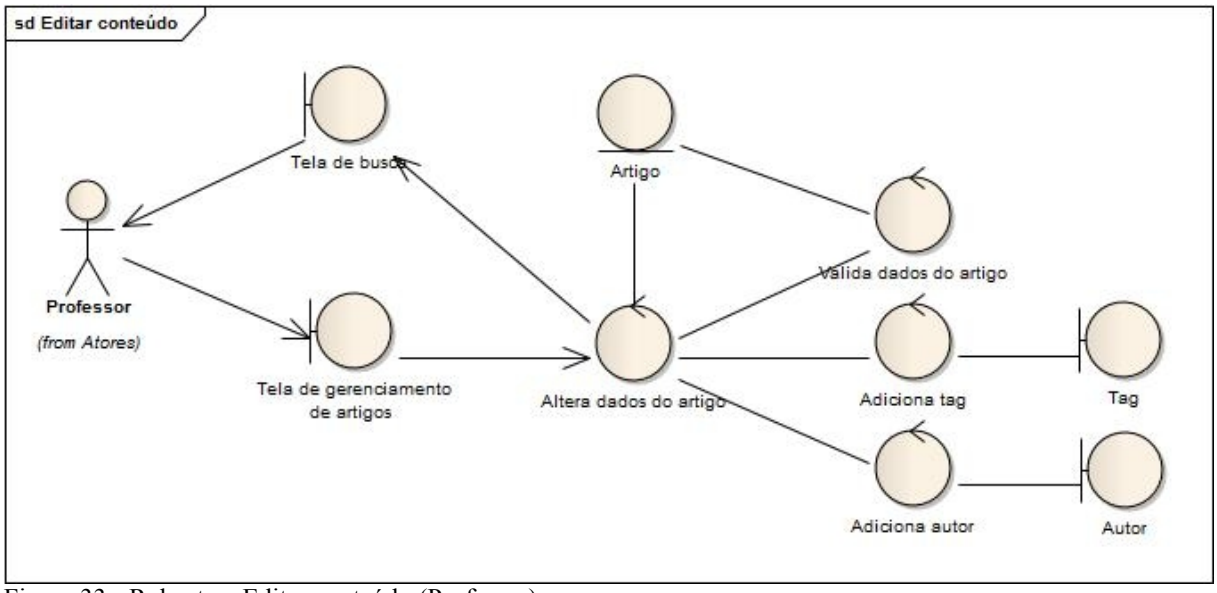

Figura 33 - Robustez: Editar conteúdo (Professor)
Fonte: Autores

Na operação de edição do conteúdo, o usuário apresenta novas informações ou troca as informações já existentes do conteúdo (artigo). Ao confirmar as alterações, o sistema irá realizar as validações das informações e alterar os dados do conteúdo (artigo), buscando o conteúdo que foi selecionado na tela de resultados do sistema, salvando, novamente, as informações, e realizando o redirecionamento para a tela de busca.

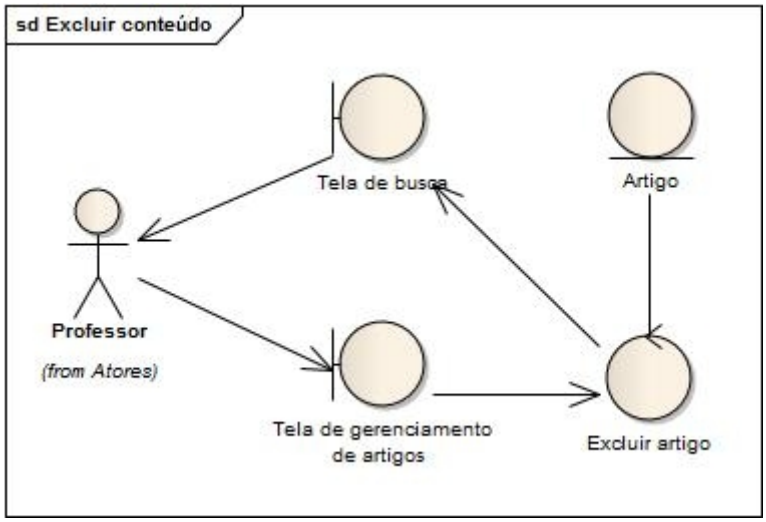

Figura 34 - Robustez: Excluir conteúdo (Professor)
Fonte: Autores

Na operação de exclusão de conteúdo, o usuário irá excluir um conteúdo (artigo). O sistema irá excluir o artigo selecionado, redirecionando para a tela de busca.

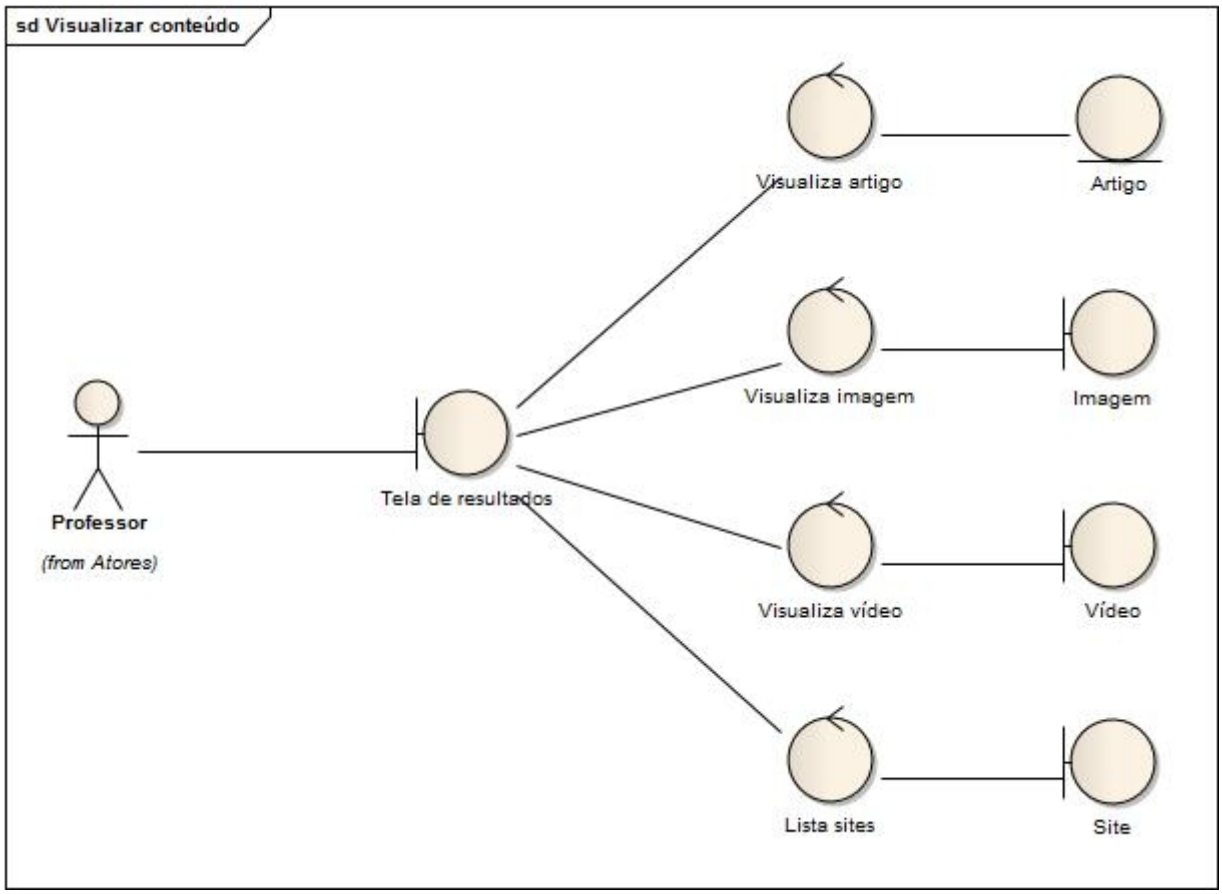

Figura 35 - Robustez: Visualizar conteúdo (Professor)
Fonte: Autores

Após a operação de busca, o usuário poderá visualizar as informações referentes ao termo, informado na busca. A tela de resultados irá mostrar imagens, textos e vídeos, referentes aos artigos encontrados, além de trazer links para os mesmos.

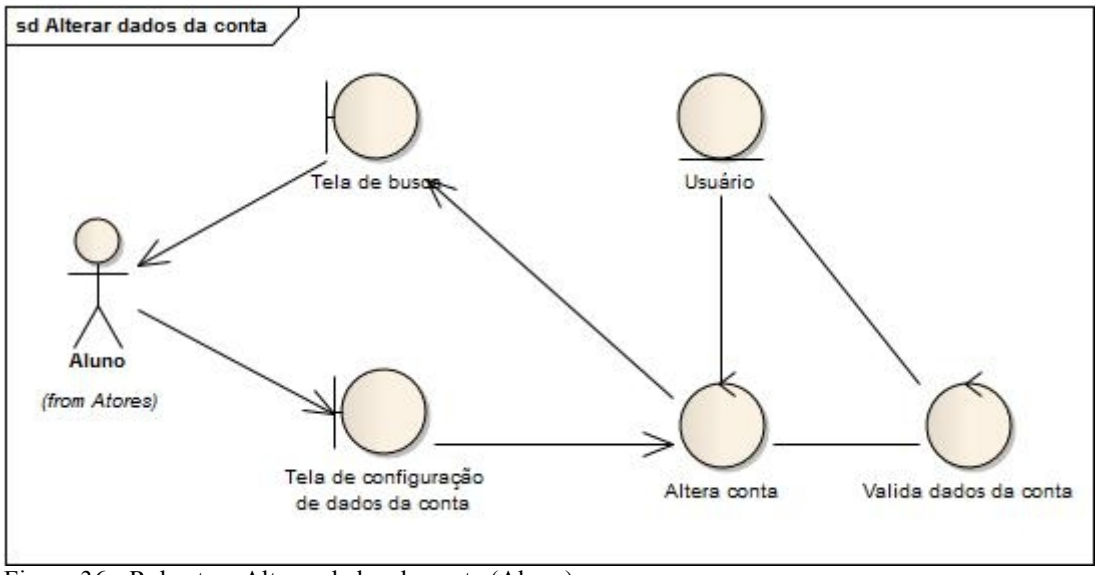

Figura 36 - Robustez: Alterar dados da conta (Aluno)
Fonte: Autores

Na operação de alterar dados da conta, o usuário irá apresentar novas informações, ou trocar as informações já existentes na tela de configuração de conta. Ao confirmar as alterações, o sistema irá realizar as validações das informações e alterar os dados da conta, buscando o usuário que está logado no sistema, salvando, novamente, as informações, e realizando o redirecionamento para a tela de busca.

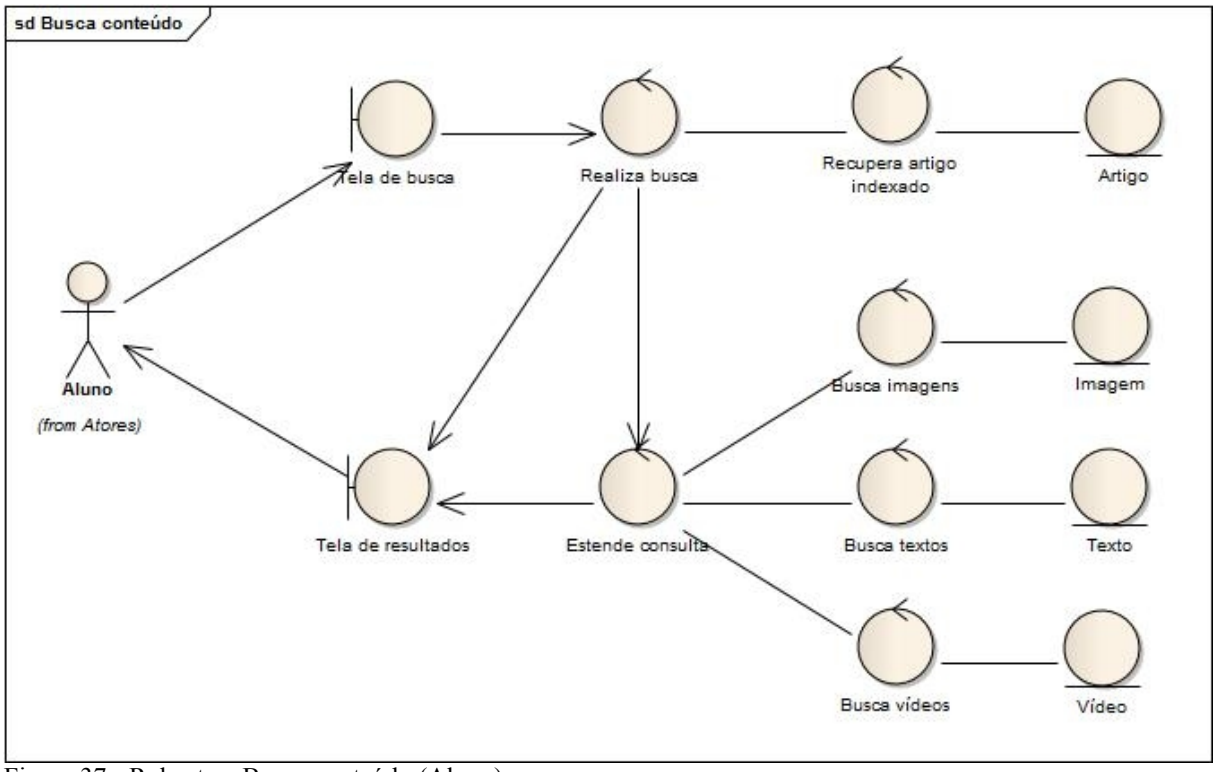

Figura 37 - Robustez: Busca conteúdo (Aluno)
Fonte: Autores

Na operação de busca do sistema, o usuário irá informar um termo na tela de busca. O sistema irá recuperar o artigo indexado, além de trazer imagens, textos e vídeos, referentes ao termo informado, sendo que estas informações serão mostradas na tela de resultados.

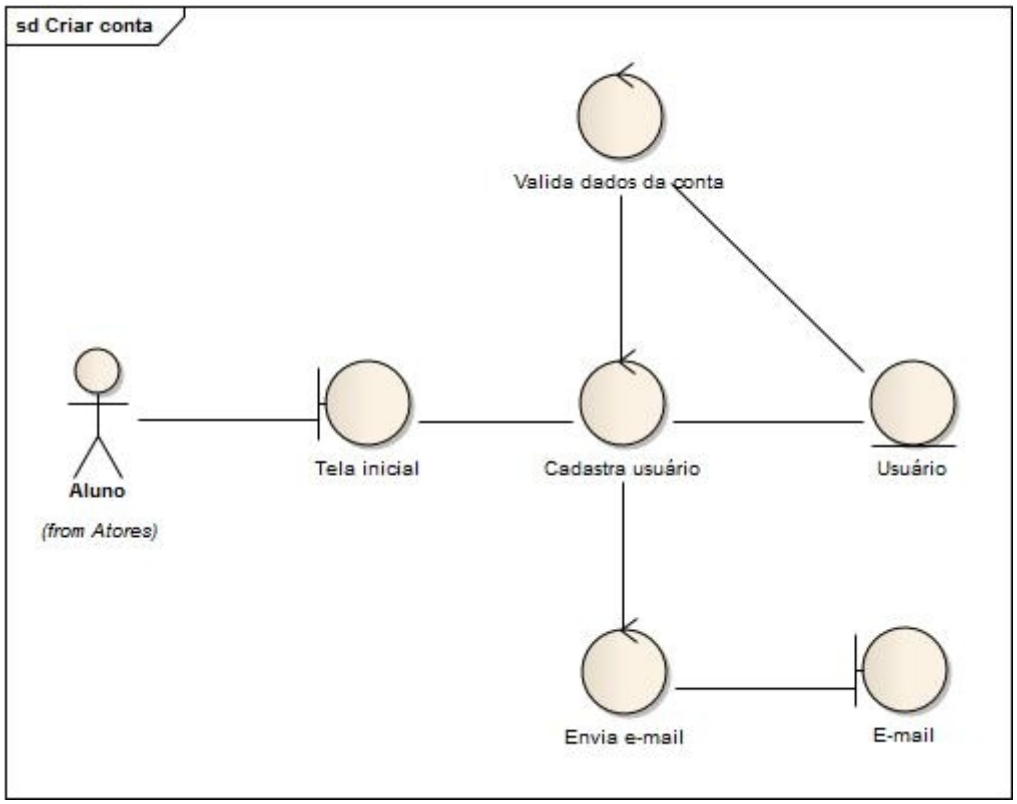

Figura 38 - Robustez: Criar conta (Aluno)
Fonte: Autores

Na operação de cadastro de usuário, o usuário irá informar os dados do cadastro na tela inicial (login). O sistema validará os dados informados, e caso a validação esteja correta, o sistema irá salvar os dados, e enviar um e-mail ao usuário, com os dados preenchidos no cadastro.

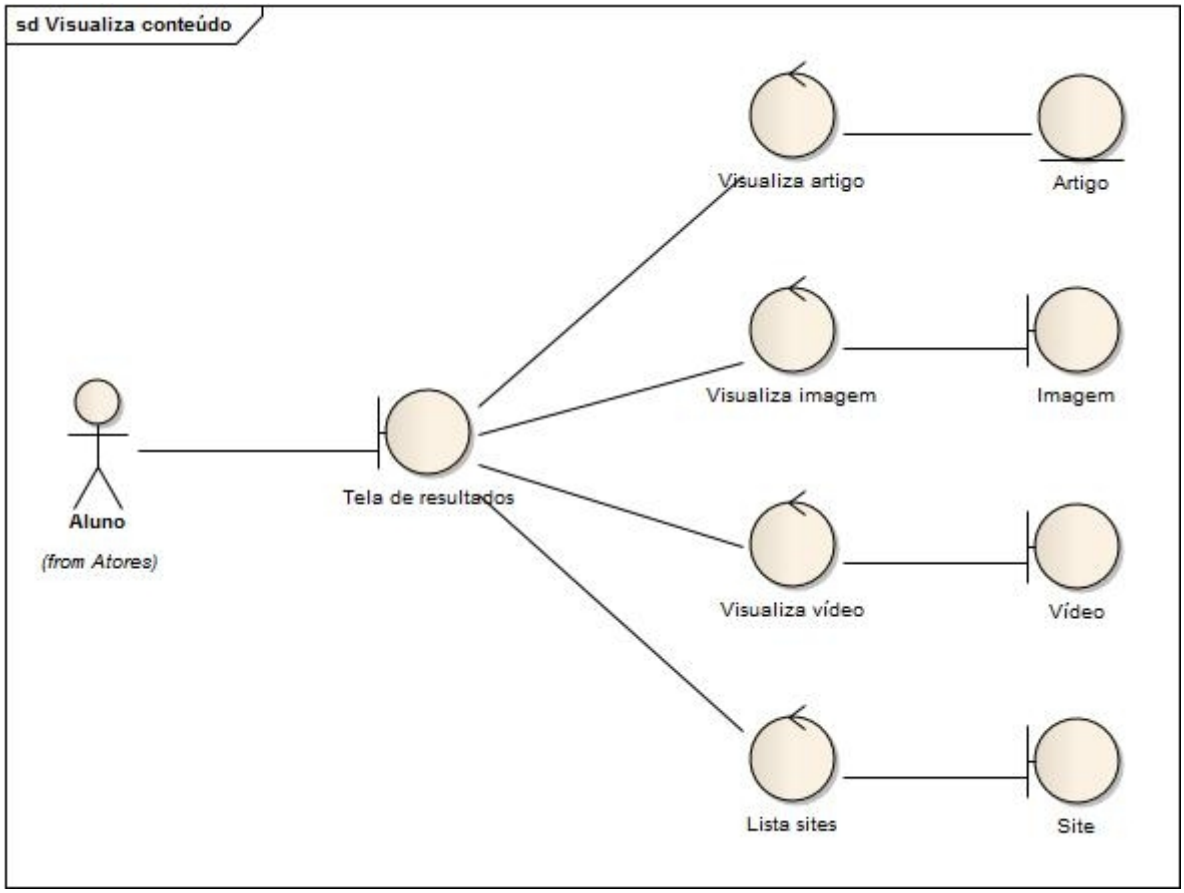

Figura 39 - Robustez: Visualiza conteúdo (Aluno)
Fonte: Autores

Após a operação de busca, o usuário poderá visualizar as informações referentes ao termo informado na busca. A tela de resultados irá mostrar imagens, textos e vídeos, referentes aos artigos encontrados, além de trazer links para os mesmos.

4.2.7 DIAGRAMA DE SEQUÊNCIA

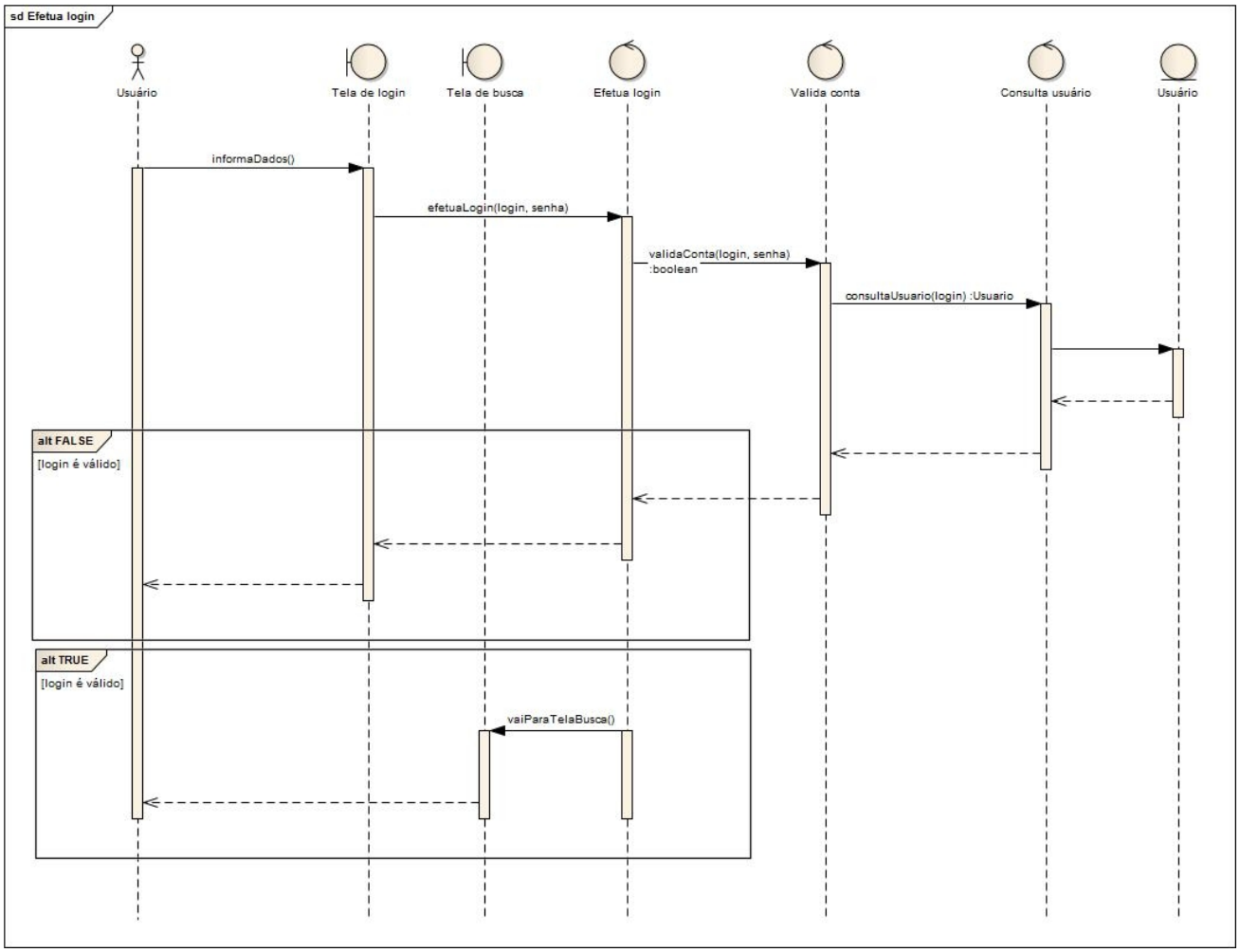

Figura 40 - Sequência: Efetua login
Fonte: Autores

No diagrama de sequência da operação de efetuar o login, pode-se verificar que o usuário deve informar o login e senha para autenticação, sendo que o sistema irá verificar se as informações são compatíveis com as do usuário cadastrado. Se o usuário for autenticado, o sistema redireciona para a tela de busca; caso contrário, o sistema exibe uma mensagem na tela de login.

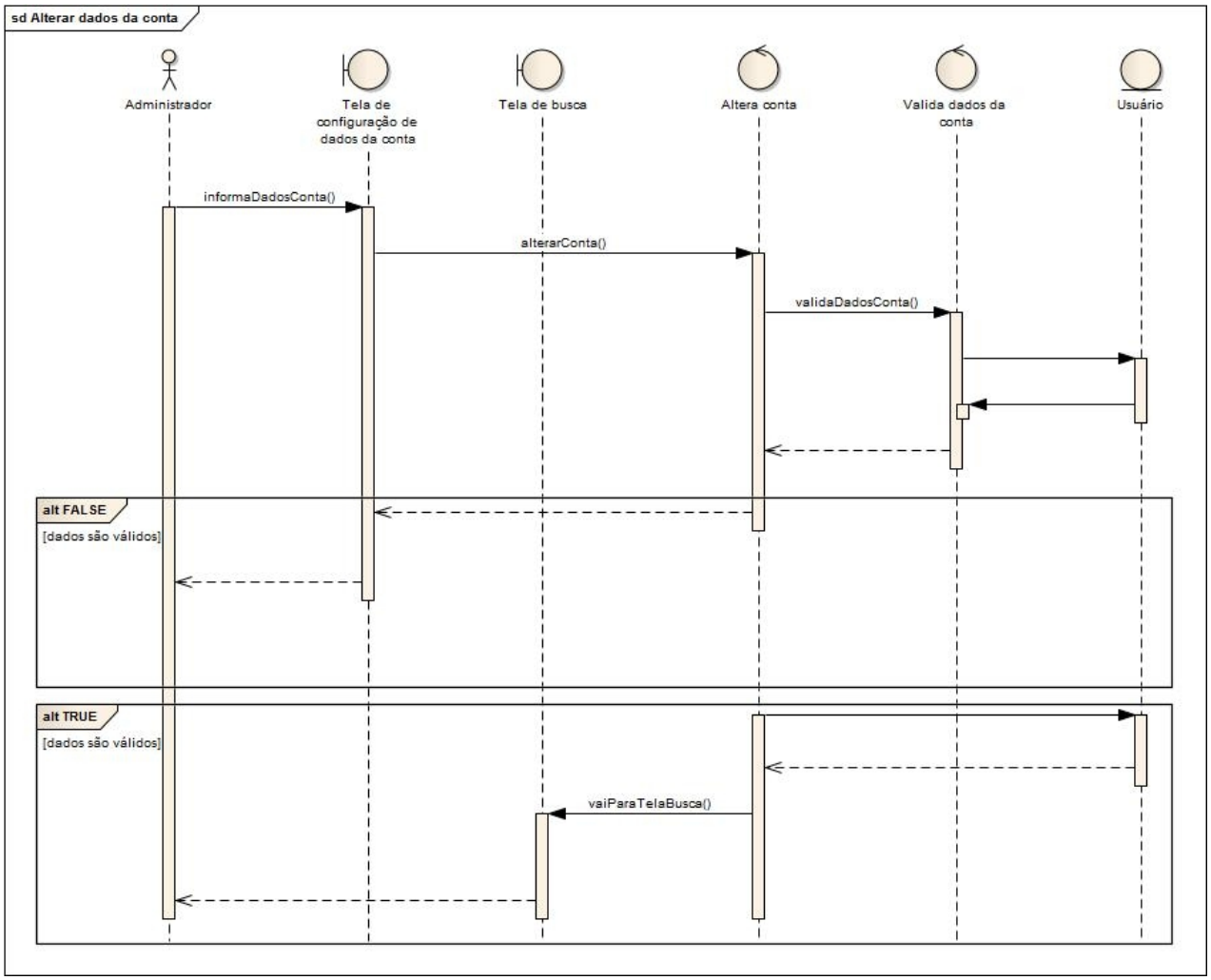

Figura 41 - Sequência: Altera dados da conta (Administrador)
Fonte: Autores

No diagrama de sequência da operação de alteração dos dados da conta, pode-se verificar que o usuário informa os dados referentes à sua conta para alteração dos mesmos. O sistema valida os dados e salva as alterações da conta. Se os dados não forem válidos, o sistema exibe uma mensagem na tela de configuração de dados da conta; caso contrário, o sistema redireciona para a tela de busca.

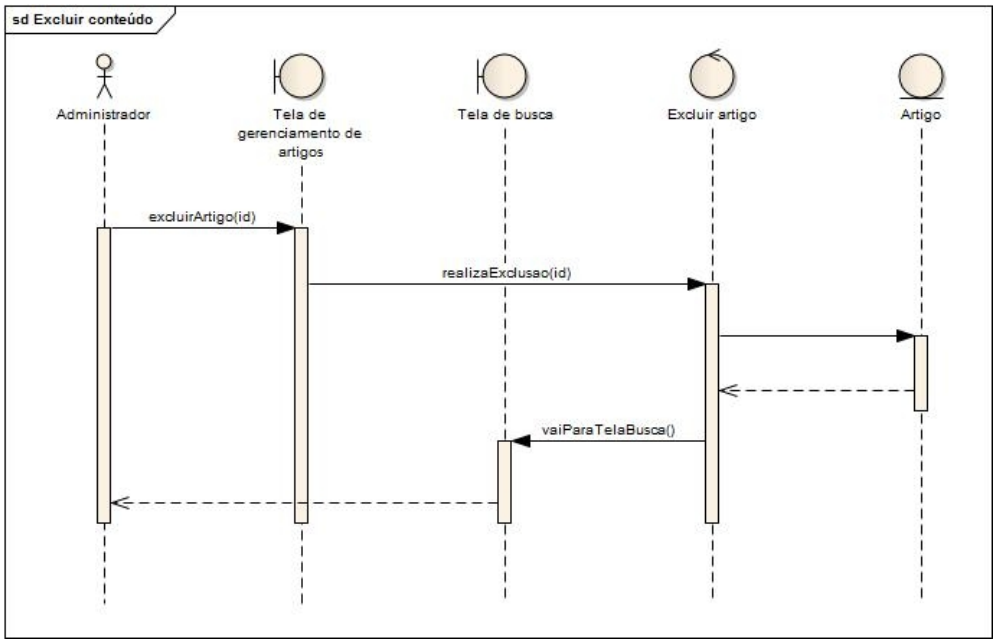

Figura 42 - Sequência: Excluir conteúdo (Administrador)
Fonte: Autores

No diagrama de sequência da operação de excluir conteúdo, pode-se verificar que o usuário exclui um conteúdo selecionado, através do id do mesmo, lembrando que o usuário seleciona o conteúdo na tela de resultados.

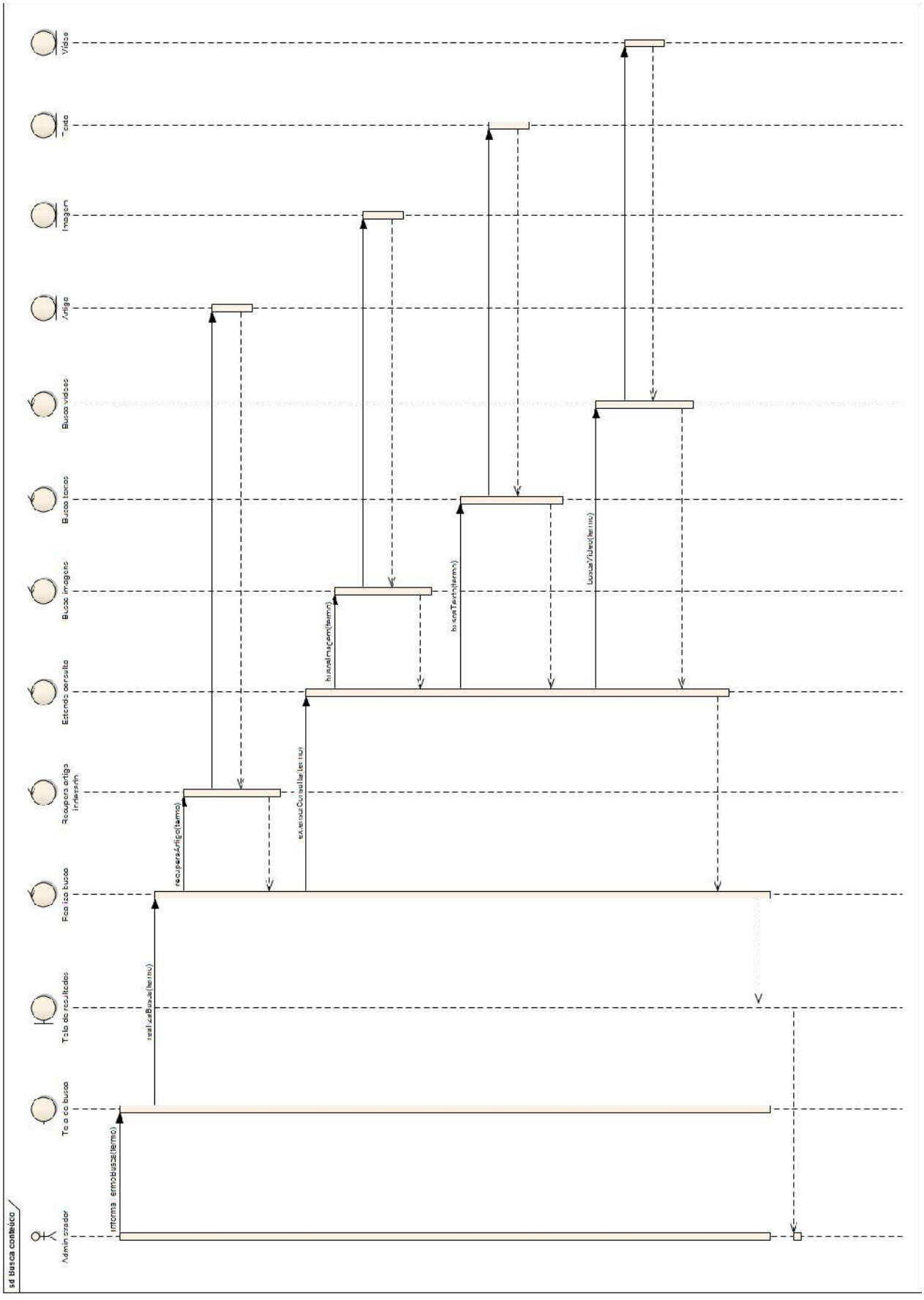

Figura 43 - Sequência: Busca conteúdo
Fonte: Autores

No diagrama de sequência da operação de busca conteúdo, pode-se verificar que o usuário informa um termo na tela de busca. O sistema, então, recupera o conteúdo indexado, através do termo informado, além de trazer imagens, textos e vídeos, relacionados aos conteúdos recuperados. As informações são visualizadas na tela resultado.

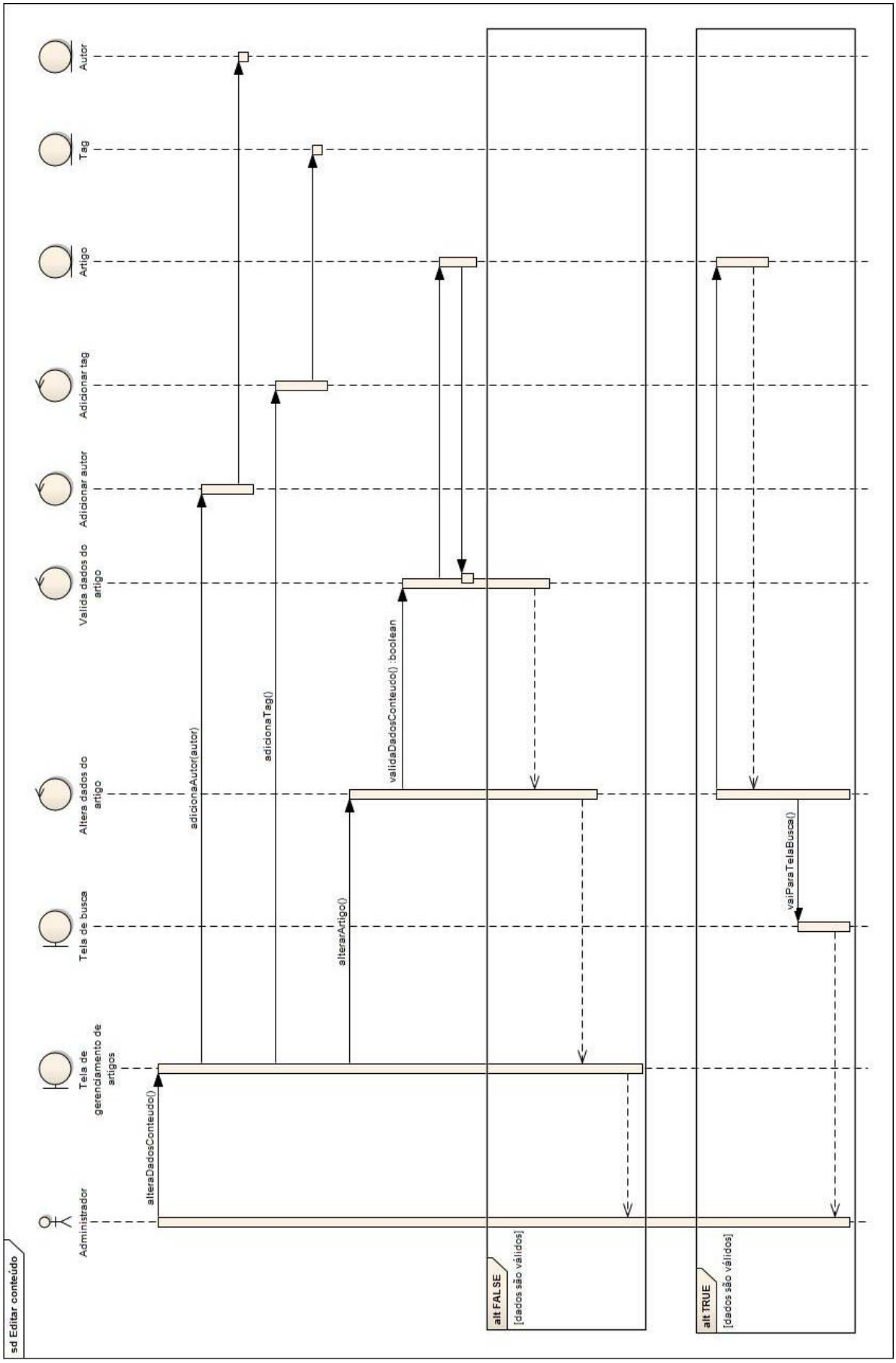

Figura 44 - Sequência: Editar conteúdo (Administrador)
Fonte: Autores

No diagrama de sequência da operação de edição de conteúdo, pode-se verificar que o usuário informa dados do conteúdo para alteração, como adição de *tags* ou autores. O sistema, então, valida essas informações. Se as informações são válidas, o sistema salva as alterações e redireciona para a tela de busca; caso contrário, é exibida uma mensagem para o usuário na tela de gerenciamento do conteúdo.

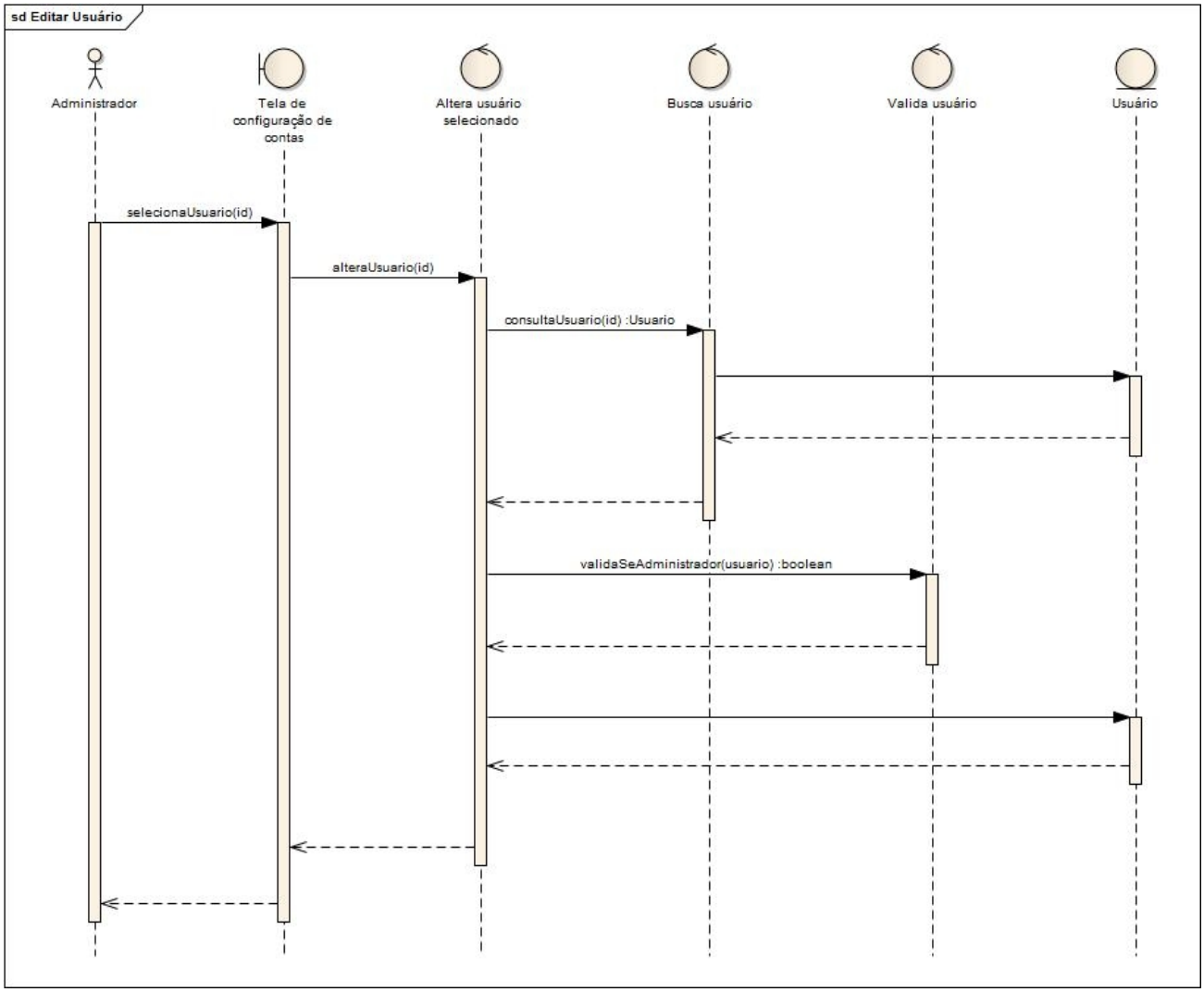

Figura 45 - Sequência: Editar usuário (Administrador)
Fonte: Autores

No diagrama de sequência da operação de edição de usuário, pode-se verificar que o usuário administrador seleciona um usuário na tela de configuração de conta, podendo alterar os dados do mesmo. O sistema, então, valida os dados informados. Se os dados são válidos, o sistema salva a alteração; caso contrário, é exibida uma mensagem de erro na tela de configuração de contas.

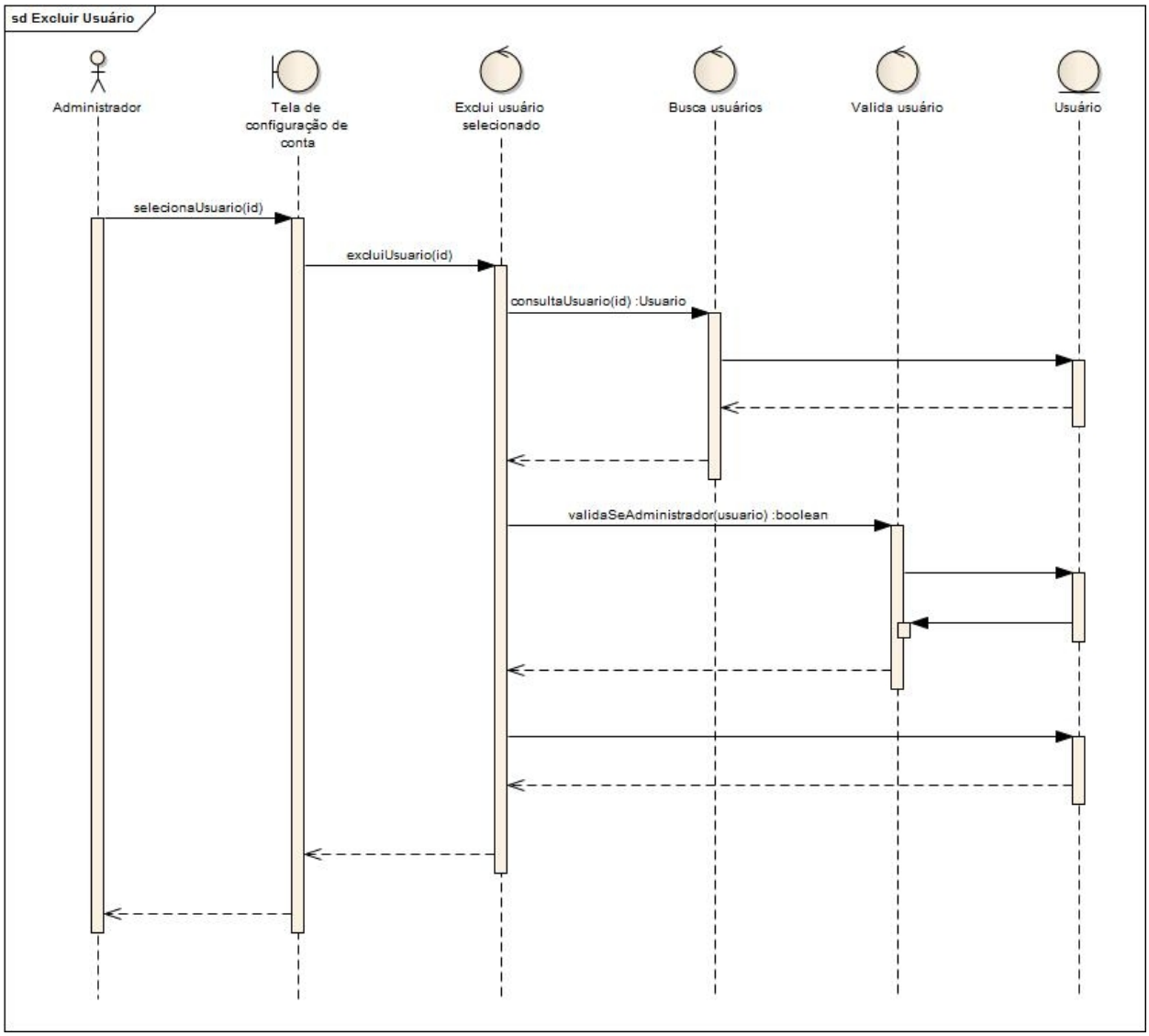

Figura 46 - Sequência: Excluir usuário (Administrador)
Fonte: Autores

No diagrama de sequência da operação de exclusão de usuário, pode-se verificar que o usuário administrador seleciona um usuário na tela de configuração de conta, podendo excluir a conta. O sistema, então, valida a operação, verificando se o usuário é um administrador. Se o usuário selecionado for um administrador, o sistema não permite a exclusão do mesmo; caso contrário, a conta é excluída.

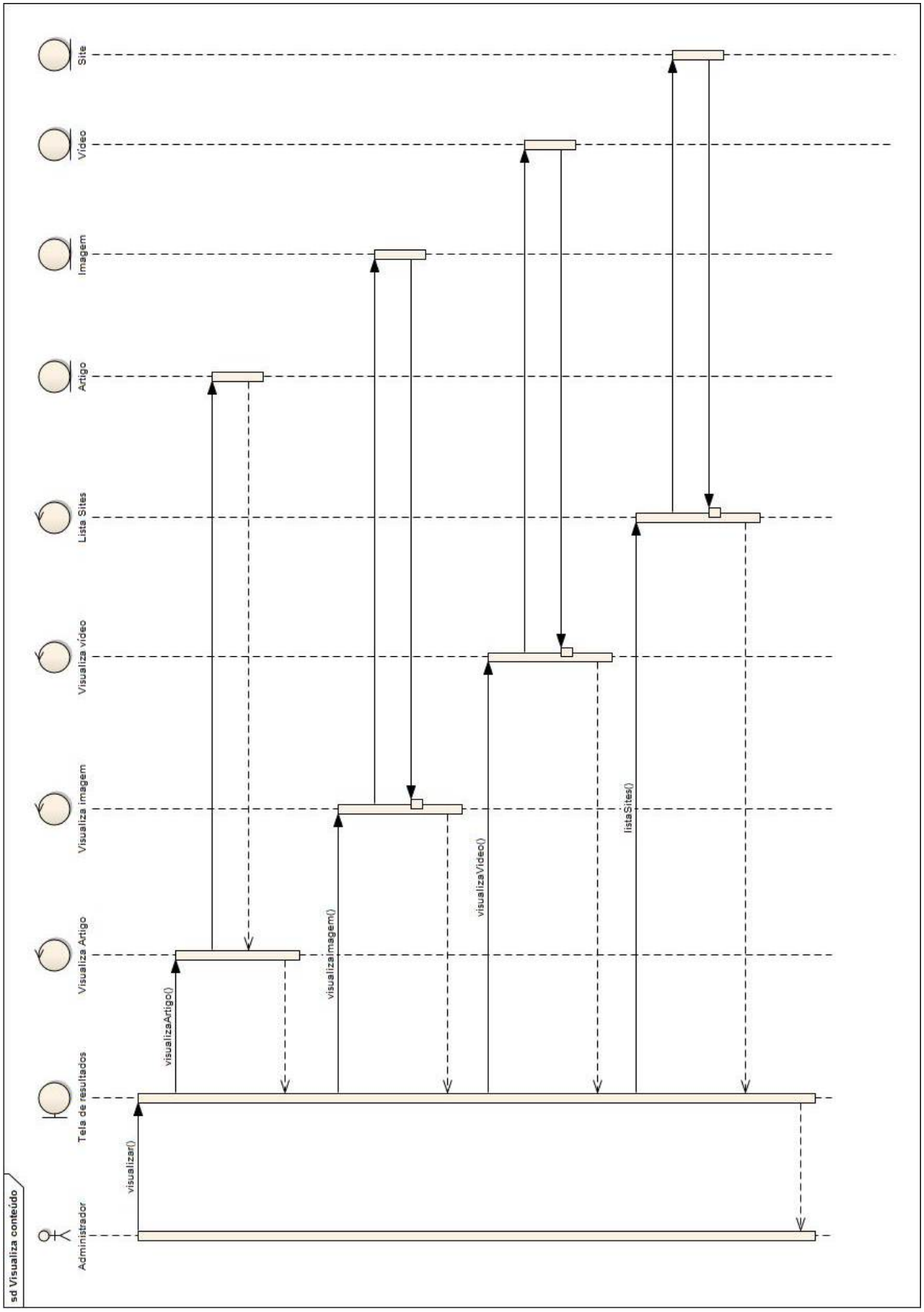

Figura 47 - Sequência: Visualiza conteúdo (Administrador)
Fonte: Autores

No diagrama de sequência da operação de visualizar conteúdo, pode-se verificar que o usuário poderá, simplesmente, visualizar as informações, vindas do resultado da busca de conteúdo.

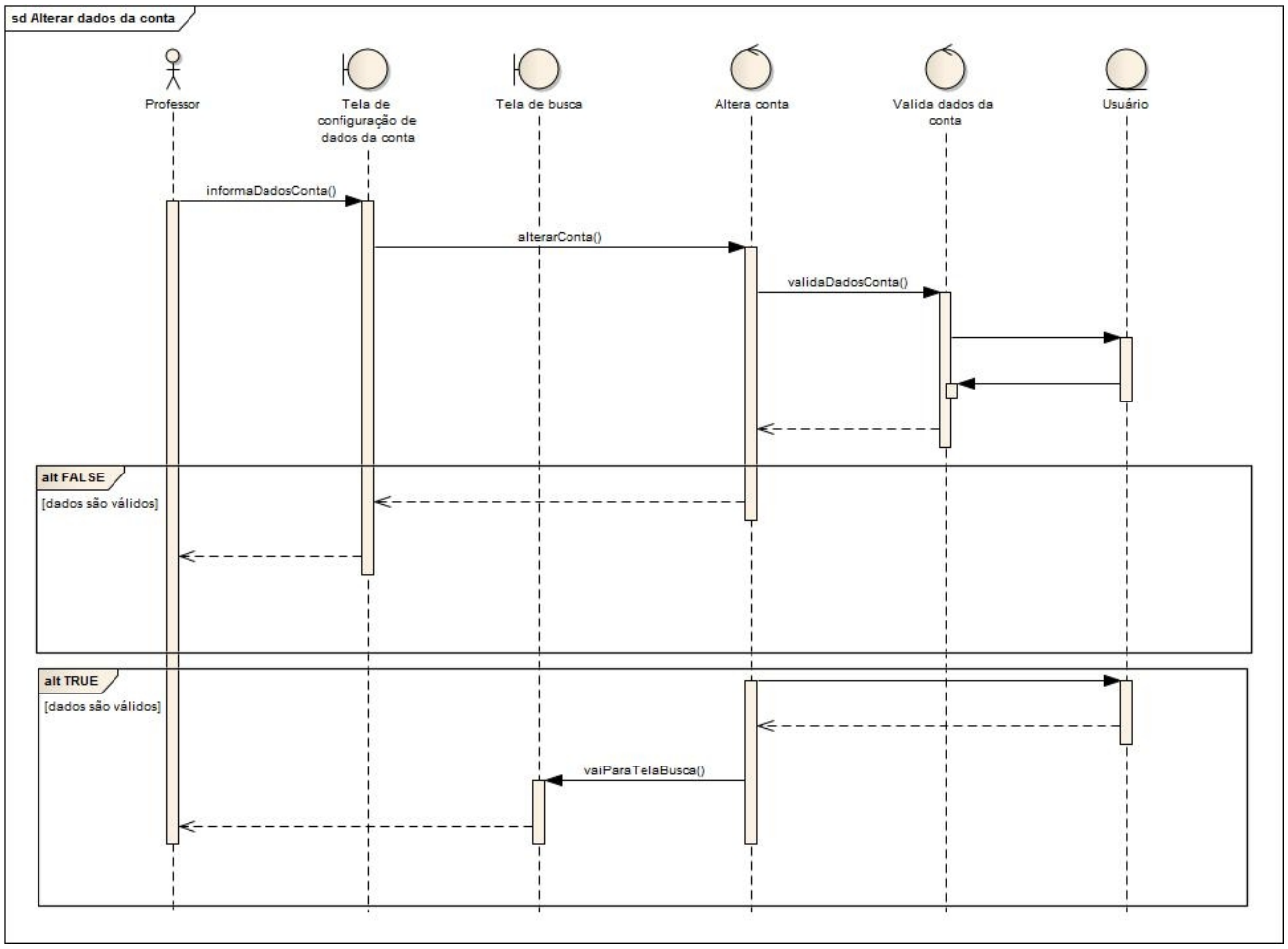

Figura 48 - Sequência: Altera dados da conta (Professor)
Fonte: Autores

No diagrama de sequência da operação de alteração dos dados da conta, pode-se verificar que o usuário informa os dados referentes à sua conta para alteração dos mesmos. O sistema valida os dados e salva as alterações da conta. Se os dados não forem válidos, o sistema exibe uma mensagem na tela de configuração de dados da conta; caso contrário, o sistema redireciona para a tela de busca.

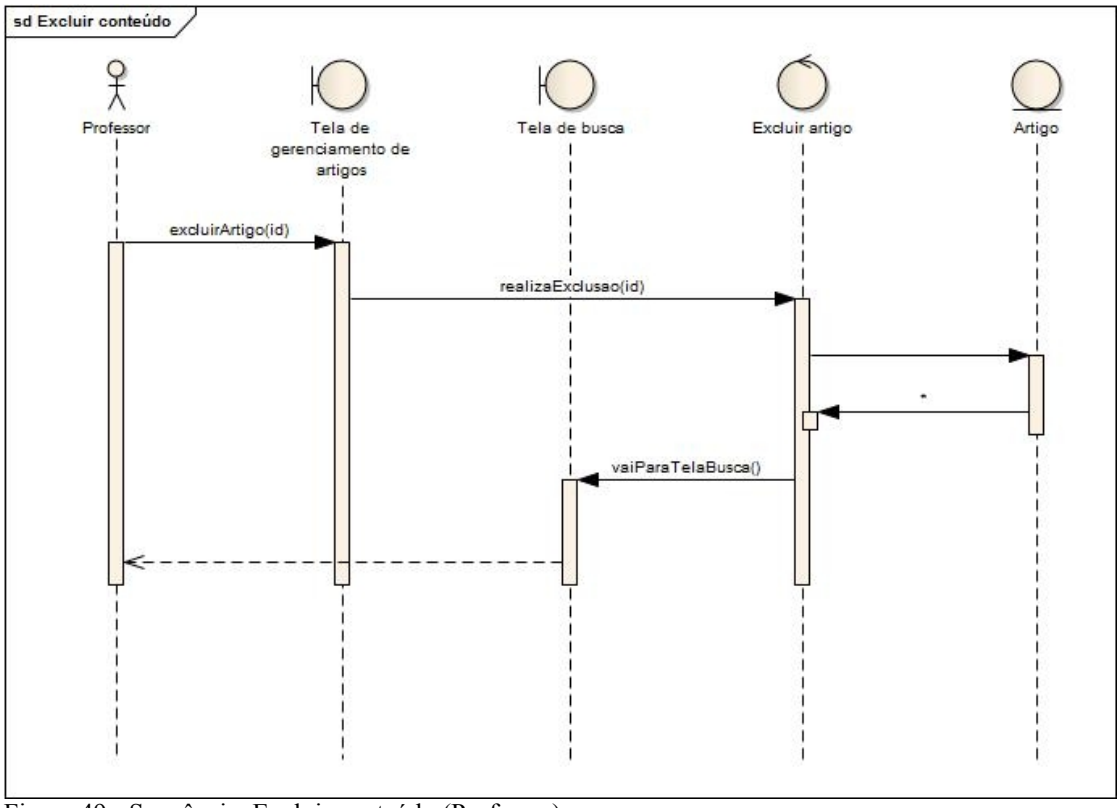

Figura 49 - Sequência: Excluir conteúdo (Professor)
Fonte: Autores

No diagrama de sequência da operação de excluir conteúdo, pode-se verificar que o usuário exclui um conteúdo selecionado através do id do mesmo, lembrando, também, que o usuário seleciona o conteúdo na tela de resultados.

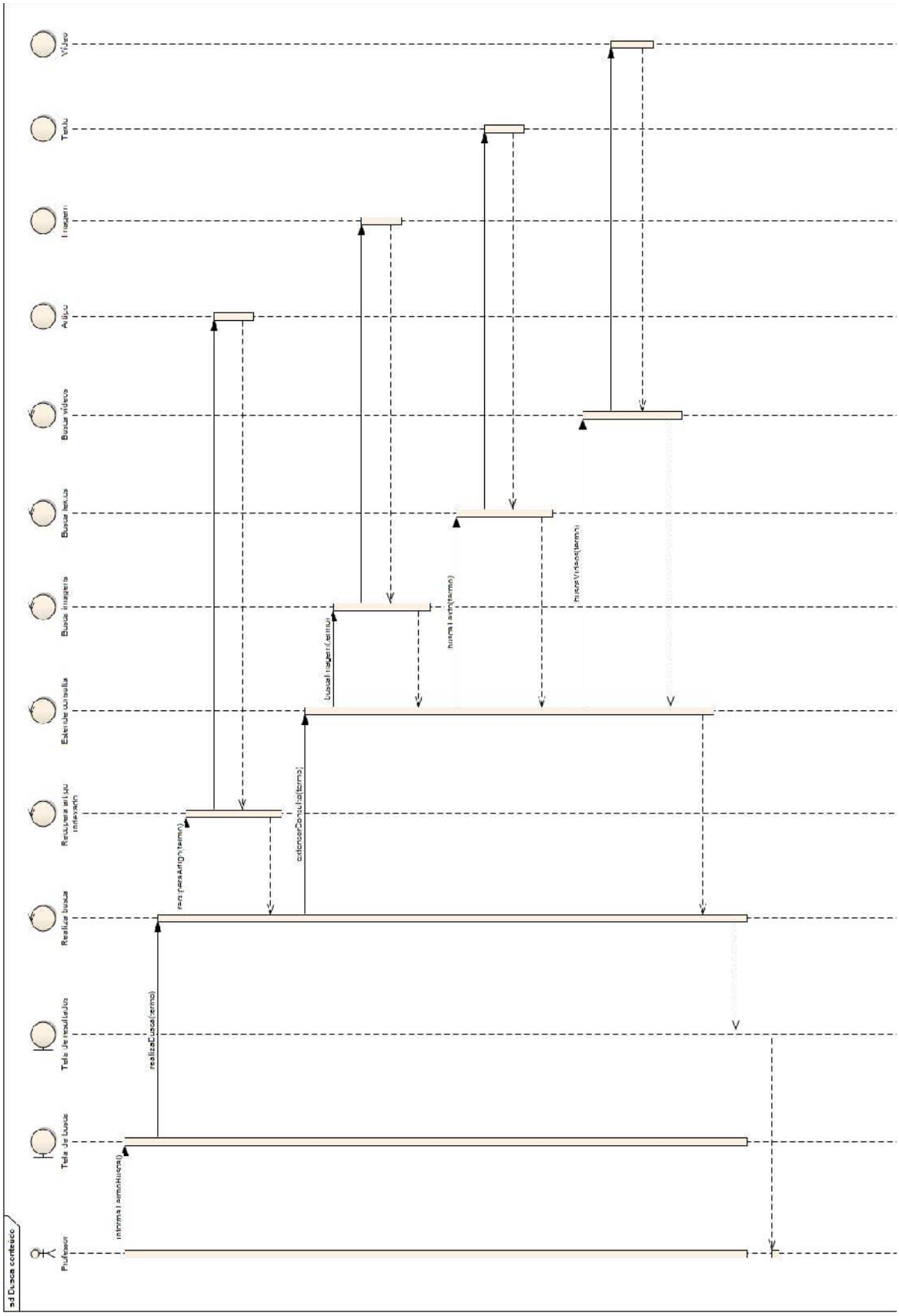

Figura 50 - Sequência: Busca conteúdo (Professor)
Fonte: Autores

No diagrama de sequência da operação de busca conteúdo, pode-se verificar que o usuário informa um termo na tela de busca. O sistema, então, recupera o conteúdo indexado através do termo informado, além de trazer imagens, textos e vídeos, relacionados aos conteúdos recuperados. As informações são visualizadas na tela de resultados.

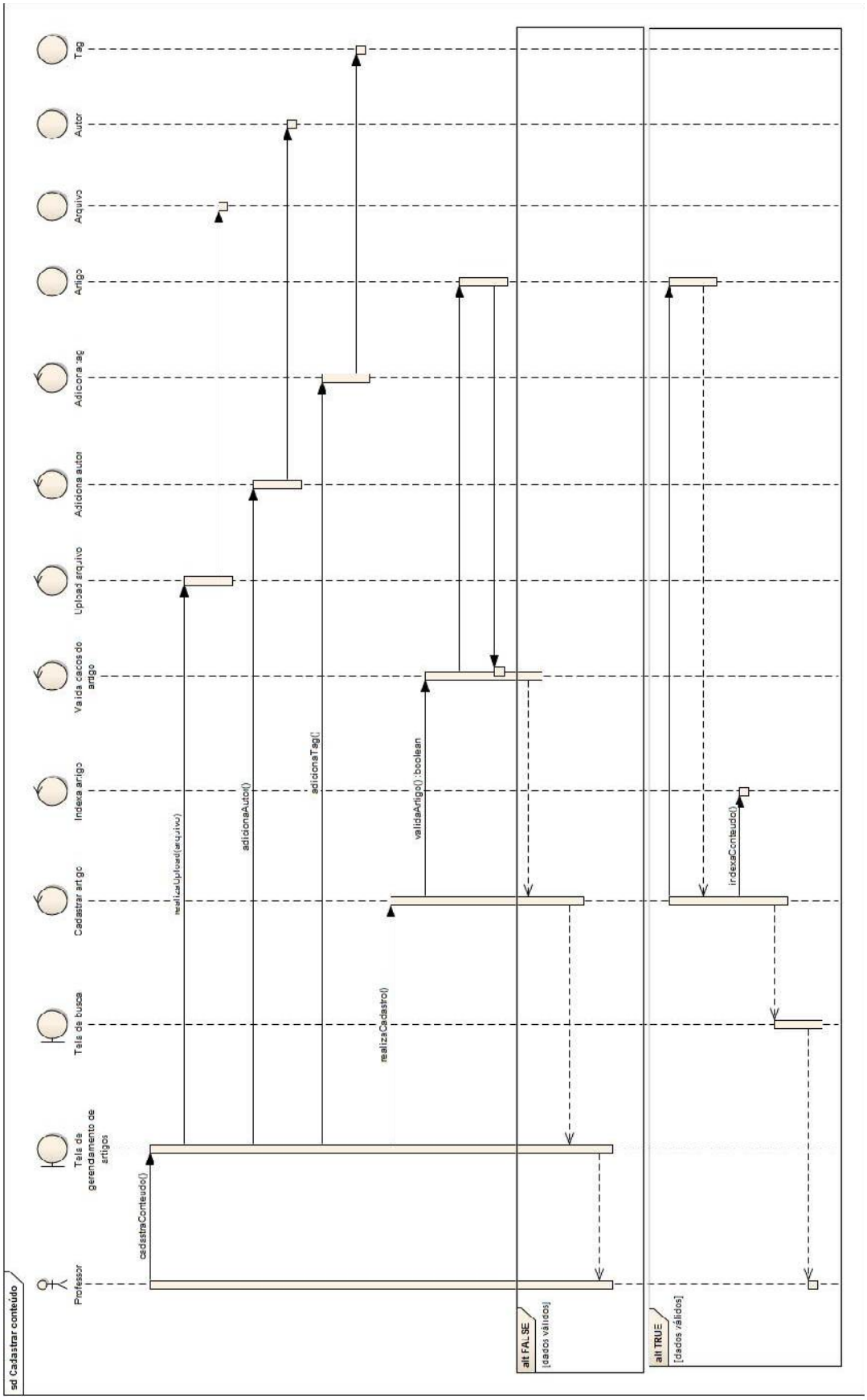

Figura 51 - Sequência: Cadastrar conteúdo (Professor)
Fonte: Autores

No diagrama de sequência da operação de cadastrar conteúdo, pode-se verificar que o usuário informa os dados do conteúdo para o cadastro, como as *tags* e os autores. O sistema, então, valida essas informações. Se as informações são válidas, o sistema salva as alterações e redireciona para a tela de busca; caso contrário, é exibida uma mensagem para o usuário na tela de gerenciamento de artigos.

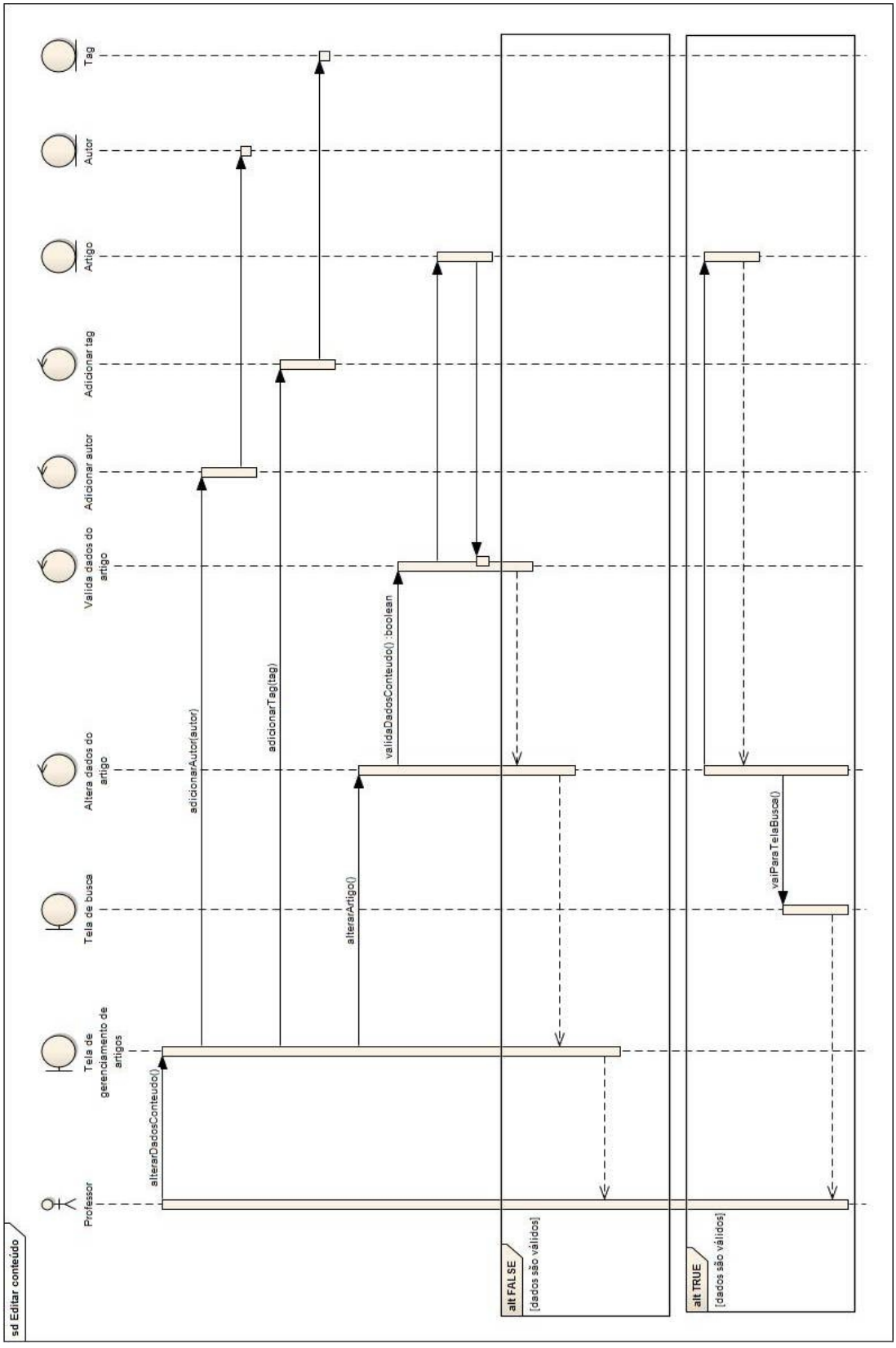

Figura 52 - Sequência: Editar conteúdo (Professor)
Fonte: Autores

No diagrama de sequência da operação de edição de conteúdo, pode-se verificar que o usuário informa dados do conteúdo para alteração, como adição de tags ou autores. O sistema, então, valida essas informações. Se as informações são válidas, o sistema salva as alterações e redireciona para a tela de busca; caso contrário, é exibida uma mensagem para o usuário na tela de gerenciamento do conteúdo.

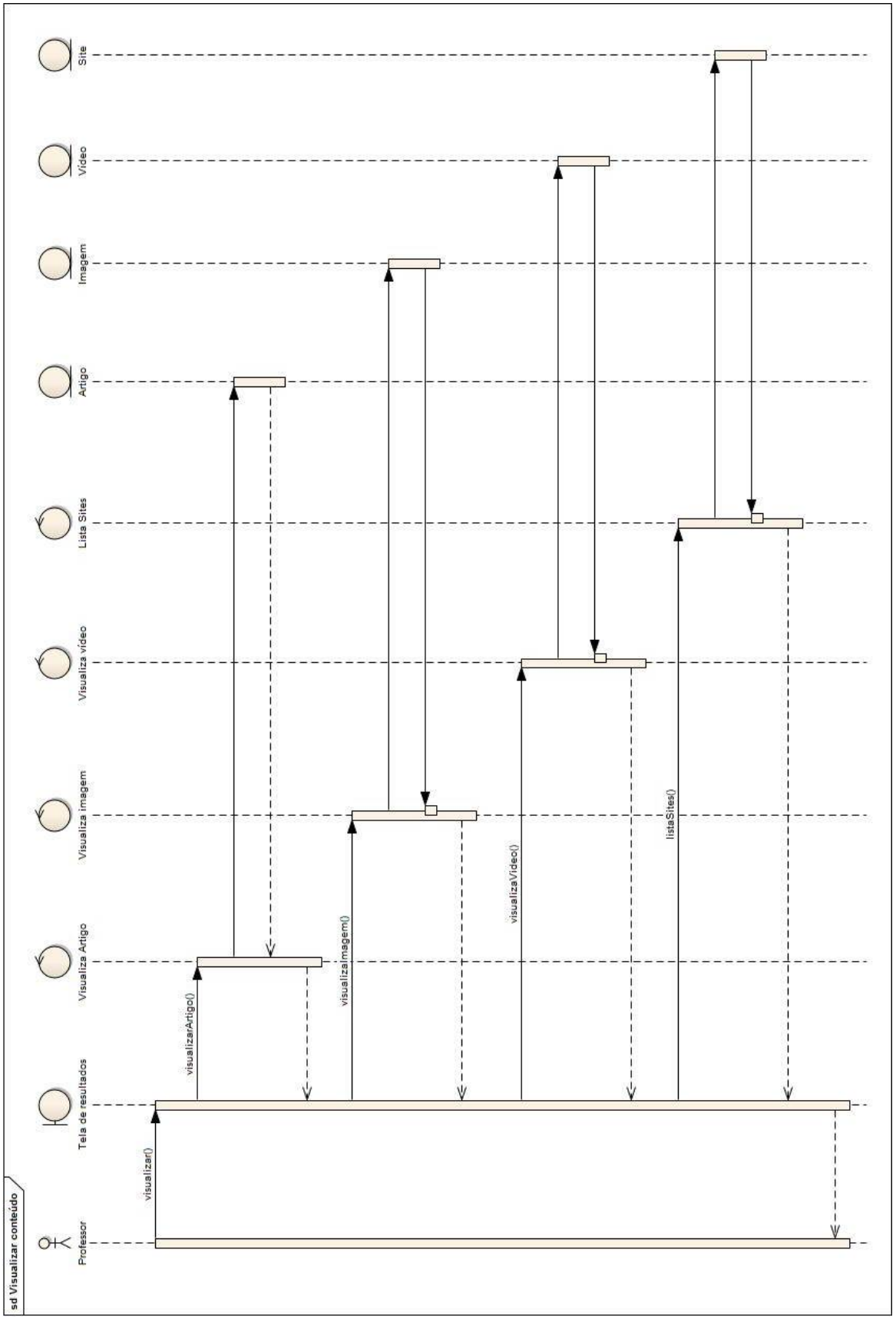

Figura 53 - Sequência: Visualizar conteúdo (Professor)
Fonte: Autores

No diagrama de sequência da operação de visualizar conteúdo, pode-se verificar que o usuário poderá, simplesmente, visualizar as informações vindas do resultado da busca de conteúdo.

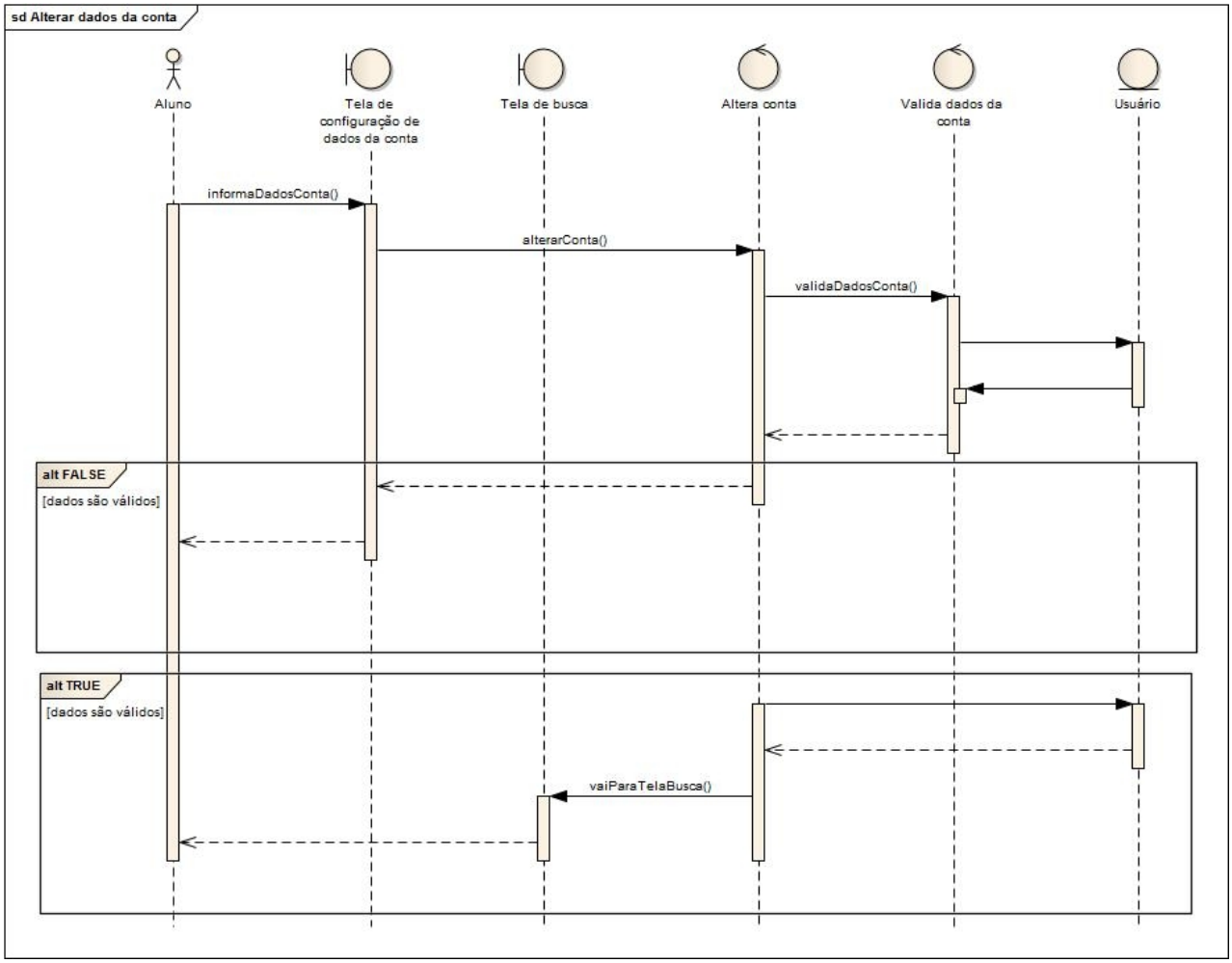

Figura 54 - Sequência: Alterar dados da conta (Aluno)
Fonte: Autores

No diagrama de sequência da operação de alteração dos dados da conta, pode-se verificar que o usuário informa os dados referentes à sua conta para alteração dos mesmos. O sistema valida os dados e salva as alterações da conta. Se os dados não forem válidos, o sistema exibe uma mensagem na tela de configuração de dados da conta; caso contrário, o sistema redireciona para a tela de busca.

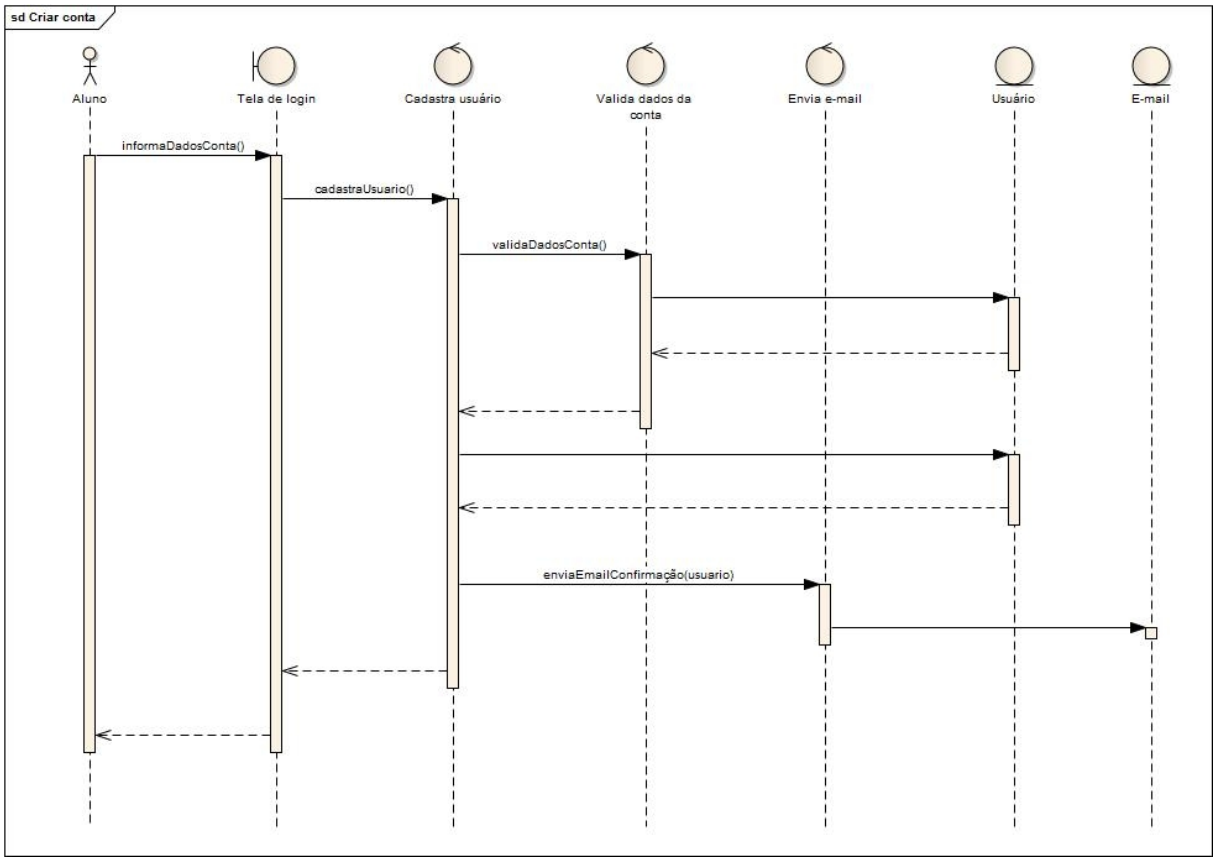

Figura 55 - Sequência: Criar conta (Aluno)
Fonte: Autores

No diagrama de sequência da operação de criação de conta, pode-se verificar que o usuário informa os dados para o cadastro. O sistema valida os dados informados. Se os dados não forem válidos, o sistema exibe uma mensagem na tela de login; caso contrário, o sistema salva as informações e envia um e-mail para o usuário com o login e senha do mesmo.

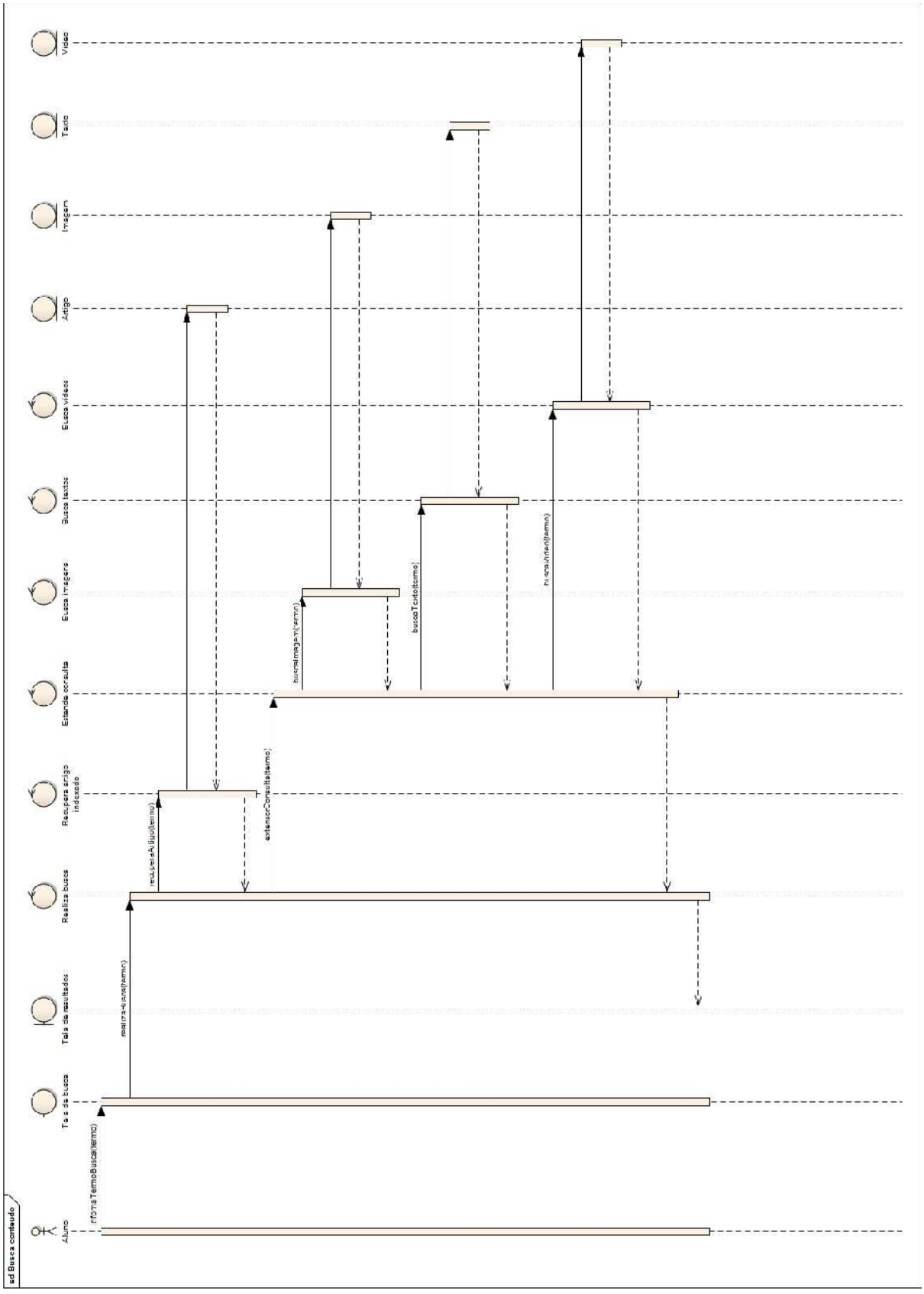

Figura 56 - Sequência: Busca conteúdo (Aluno)
Fonte: Autores

No diagrama de sequência da operação de busca conteúdo, pode-se verificar que o usuário informa um termo na tela de busca. O sistema, então, recupera o conteúdo indexado através do termo informado, além de trazer imagens, textos e vídeos, relacionados aos conteúdos recuperados. As informações são visualizadas na tela de resultados.

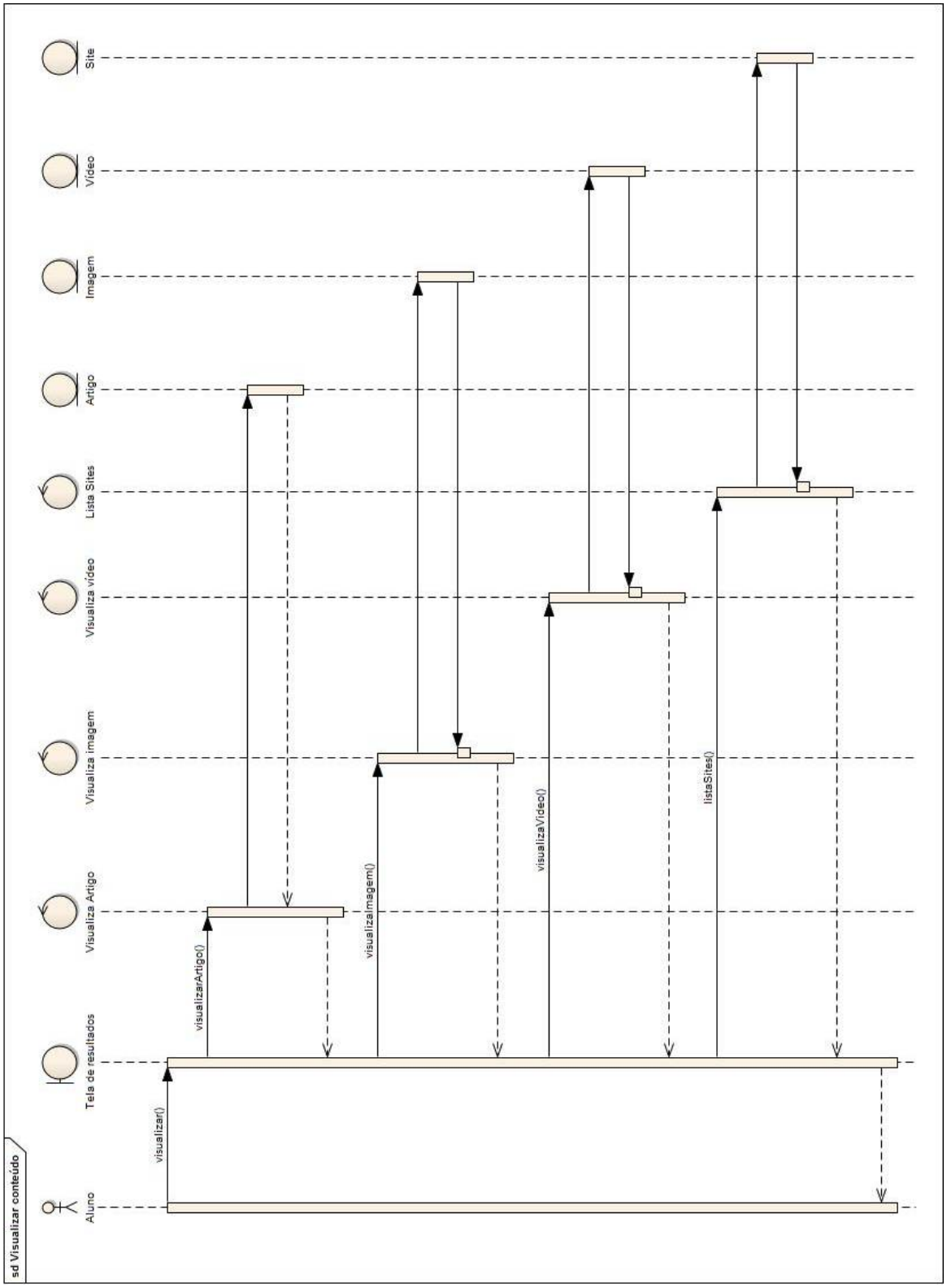

Figura 57 - Sequência: Visualizar conteúdo (Aluno)
Fonte: Autores

No diagrama de sequência da operação de visualizar conteúdo, pode-se verificar que o usuário poderá, simplesmente, visualizar as informações vindas do resultado da busca de conteúdo.

4.2.8 DIAGRAMA DE CLASSE

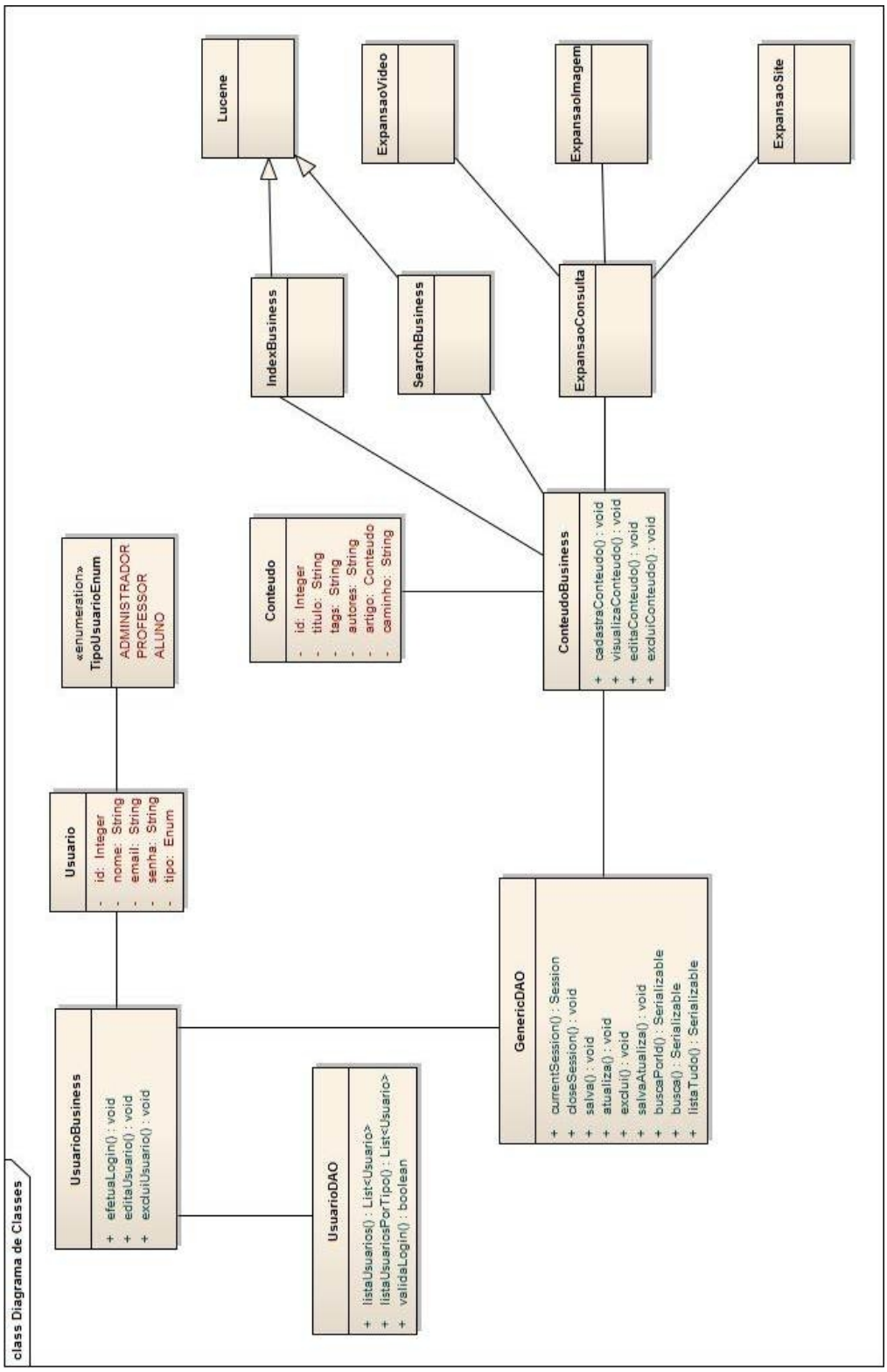

Figura 58 - Diagrama de classe
Fonte: Autores

5 SISTEMA PARA BUSCA E EXTENSÃO DE CONSULTA

O seguinte capítulo descreve, inicialmente, o esquema físico do sistema e cada uma das tecnologias utilizadas para o desenvolvimento do mesmo. Serão apresentadas todas as telas do protótipo desenvolvido, e uma explicação de cada operação, que um usuário poderá efetuar.

Esse capítulo, também, mostra como foi feita a validação do protótipo, descrevendo o método de validação, seu cenário e os resultados obtidos. Por fim, um caso de teste é apresentado, mostrando passo-a-passo o fluxo principal do sistema.

5.1 ESQUEMA DO SISTEMA

Esta seção irá mostrar o esquema físico do sistema, de uma forma um pouco mais detalhada. A Figura 59, ilustra o esquema como um todo, porém, ela é dividida em 3 módulos que serão descritos nos tópicos dessa seção.

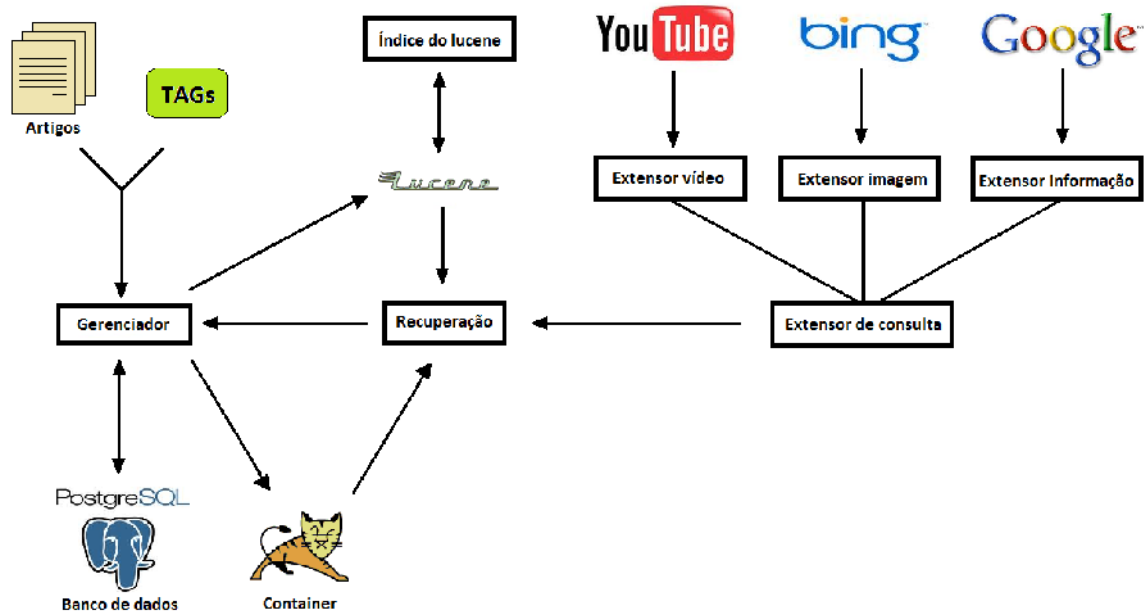

Figura 59 - Esquema físico do sistema
Fonte: Autores

5.1.1 Gerenciamento do sistema

Este módulo é responsável pelo gerenciamento de informações no sistema, permitindo que o usuário adicione o documento (artigo) para a indexação, além das tags referentes ao artigo, esta etapa é a mesma indicada pelo número 1 na Figura 9 – Esquema do sistema proposto. Após o usuário realizar a operação de upload do documento desejado e informar as tags para esse documento, o sistema irá salvar o documento em um diretório, dentro do container web (tomcat), e irá gravar os dados referentes a esse documento, como o nome e o caminho, onde foi salvo, no banco de dados (postgres). Ao salvar as informações no banco, o id desse registro e o próprio texto do documento são recuperados para, posteriormente, serem utilizados na indexação pelo sistema Lucene.

Outra função desse módulo é a própria visualização das informações do documento. Após a busca ser realizada e as informações serem recuperadas, tanto pelo Lucene, quanto pelo extensor de consultas, - o sistema exibe todas elas, em uma página de resultados. Essa é o passo de número 7 na Figura 9.

5.1.2 Indexação e recuperação

A função deste módulo é indexar o documento, utilizando a ferramenta Lucene. Após o documento ser salvo no banco de dados (postgre), o texto, as tags e o id do registro formam o índice, que será utilizado e gravado pelo Lucene. Esse etapa pode ser relacionada com o passo 2 e 6 da Figura 9.

Para a recuperação do conteúdo, o Lucene irá retornar às informações contidas no índice, com base no termo informado pelo usuário e com base nas tags, que estão gravadas no índice. Assim sendo, com os ids retornados poderão ser recuperadas as informações gravadas no banco. Com as tags poderá ser feita a extensão de consulta e o caminho do arquivo irá trazer o documento salvo no diretório do container (tomcat). Esse passo é o mesmo indicado pelo número 3 na Figura 9.

5.1.3 Extensão de consulta

O módulo de extensão de consulta terá como objetivo recuperar arquivos de texto, imagem e vídeo, referentes à busca de artigos.

Quando o usuário informa um termo para a busca, o sistema recupera alguns artigos, relacionados a esse termo, através das tags do documento, que, no caso, também, irão servir para recuperar outros tipos de arquivos da web.

O extensor de consulta por imagem recupera algumas imagens, utilizando a api do Bing, um motor de pesquisa da Microsoft. O extensor de consulta, por vídeo, irá trazer links de alguns vídeos do Youtube, para visualização na página de resultados. E o extensor de consultas, por texto, irá apresentar os primeiros resultados do sistema de busca Google, lembrando que todas as informações recuperadas estão, diretamente, relacionadas às tags dos documentos recuperados.

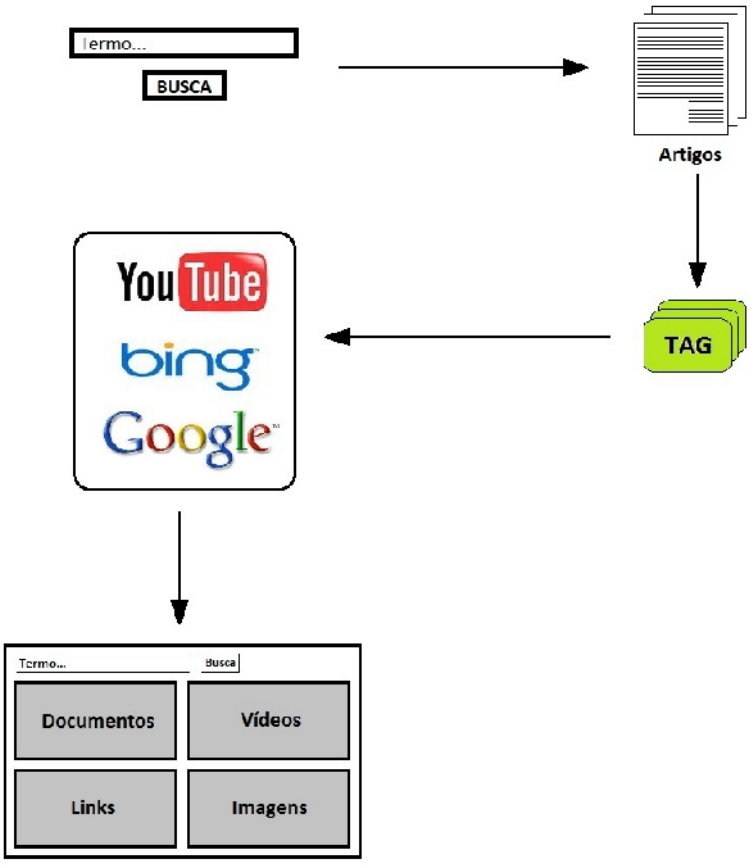

Figura 60 - Extensão de consulta
Fonte: Autores

A figura 60 ilustra como a extensão de consulta ocorre. Após o usuário informar um termo e realizar a consulta, - o sistema irá buscar, no corpo dos artigos indexados, o termo informado pelo usuário. Depois de recuperados os artigos, o sistema utiliza as tags pertencentes aos artigos, para efetuar outra consulta. O resultado são os links do Google, os vídeos do Youtube e as imagens do Bing que são recuperadas, através das tags dos artigos. O termo informado pelo usuário não tem relação direta com o resultado da expansão de consulta.

Essa etapa de expansão de consulta pode ser relacionada pelo passo de número 4 e 5 apresentado na Figura 9.

5.2 FERRAMENTAS UTILIZADAS

Nesta seção do trabalho, serão apresentadas as tecnologias utilizadas no desenvolvimento da solução proposta, bem como os motivos que influenciaram na escolha destas.

5.2.1 Plataforma Java

Java (2011) explica que :

> Java é uma linguagem de programação e uma plataforma de computação lançada pela primeira vez pela Sun Microsystems em 1995. É a tecnologia que capacita muitos programas da mais alta qualidade, como utilitários, jogos e aplicativos corporativos, entre muitos outros, por exemplo. O Java é executado em mais de 850 milhões de computadores pessoais e em bilhões de dispositivos em todo o mundo, inclusive telefones celulares e dispositivos de televisão.

De acordo com Java (2011), Java é necessário por vários fatores, como o fato de que muitos aplicativos e sites funcionam, somente, com o Java instalado. E muitos outros

aplicativos e sites são desenvolvidos dando suporte a essa tecnologia todos os dias. O Java é rápido, seguro e confiável.

Além desses motivos, outro que nos influenciou, na escolha pela linguagem de programação Java, foi o fato de termos experiência com a linguagem, tanto acadêmica como profissional.

5.2.2 JSF

Este framework é o resultado de um projeto apoiado pela Sun, e teve sua primeira versão apresentada em setembro de 2002 [GEARY e HORSTMANN, 2005].

Segundo Geary e Horstmann (2007), o JSF traz o desenvolvimento rápido de interfaces de usuário para o Java server-side, possuindo conjuntos de componentes pré-fabricados de IU (interface de usuário). O JSF traz um modelo de programação orientado a eventos e um modelo de componentes que permite a desenvolvedores independentes fornecerem componentes adicionais.

Dentre os motivos para a escolha desta tecnologia, podemos citar a facilidade para criar interfaces, usando os componentes do framework, além de possuir inúmeras IDEs e plugins para desenvolvimento, sem falar que é a especificação de desenvolvimento para a web, indicada pela Sun Microsystems, desenvolvedora da plataforma Java.

5.2.3 Hibernate

De acordo com Hibernate (2011), Hibernate é um serviço de alto desempenho de consultas e persistência objeto/relacional. Representa a solução mais flexível e poderosa encontrada no mercado, ficando o Hibernate responsável por cuidar do mapeamento de classes Java para tipos de dados SQL.

Segundo Bauer e King (2005), Hibernate é um projeto ambicioso, que visa ser uma solução completa para o problema de gerenciamento de dados persistentes em Java. Este

medeia a interação do aplicativo com um banco de dados relacional, deixando o desenvolvedor livre para se concentrar no problema em questão.

Este framework facilita a consulta de dados, o que reduz, significativamente, o tempo de desenvolvimento. É, por isso, um grande fator para a escolha desta tecnologia.

5.2.4 Apache Lucene

Segundo Apache Software Foundation (2011), o projeto Apache Lucene desenvolve software open-source de pesquisa, incluindo Apache Lucene Core, antigamente chamado de Lucene Java, - fornece indexação de dados e implementações de pesquisa com base na linguagem Java.

De acordo com Gospodnetic e Hatcher (2005), Lucene é uma biblioteca de recuperação de informação de alto desempenho, projetada para ser agregada a sistemas de indexação e pesquisas textuais, em acervos de documentos eletrônicos.

Um dos grandes fatores para a escolha do SRI Lucene, foi ele ter sido desenvolvido com base na linguagem Java, o que oferece uma maior flexibilidade na implementação do protótipo de solução, que é desenvolvido nesta linguagem. Outro grande fator para a escolha do Lucene é que, de acordo com Hatcher e Gospodnetic (2005), esta é a biblioteca de recuperação de informação, a mais popular entre as existentes.

5.2.5 Enterprise Architect

Enterprise Architect, segundo Sparx Systems (2011), é uma plataforma de design, baseado na padrão UML 2.3, a qual disponibiliza uma modelagem de alto desempenho e visualização. Além disso conta com uma completa rastreabilidade através de requisitos de negócios e conta com uma interface intuitiva.

5.3 SISTEMA DESENVOLVIDO

Nesta seção, é apresentado o sistema desenvolvido e suas funcionalidades. O sistema é apresentado tela-a-tela, descrevendo e demonstrando as funcionalidades disponíveis e como utilizar.

Figura 61 - Tela de login do sistema
Fonte: Autores

Inicialmente ao acessar o site do sistema, depara-se com uma tela, onde é possível registrar um usuário e efetuar o login. Para se registrar no sistema, basta informar sua conta de e-mail e uma senha para acessar o sistema. Feito isto, o sistema ira sinalizar com uma mensagem de sucesso de registro e poderá ser feito o login no sistema com a conta criada.

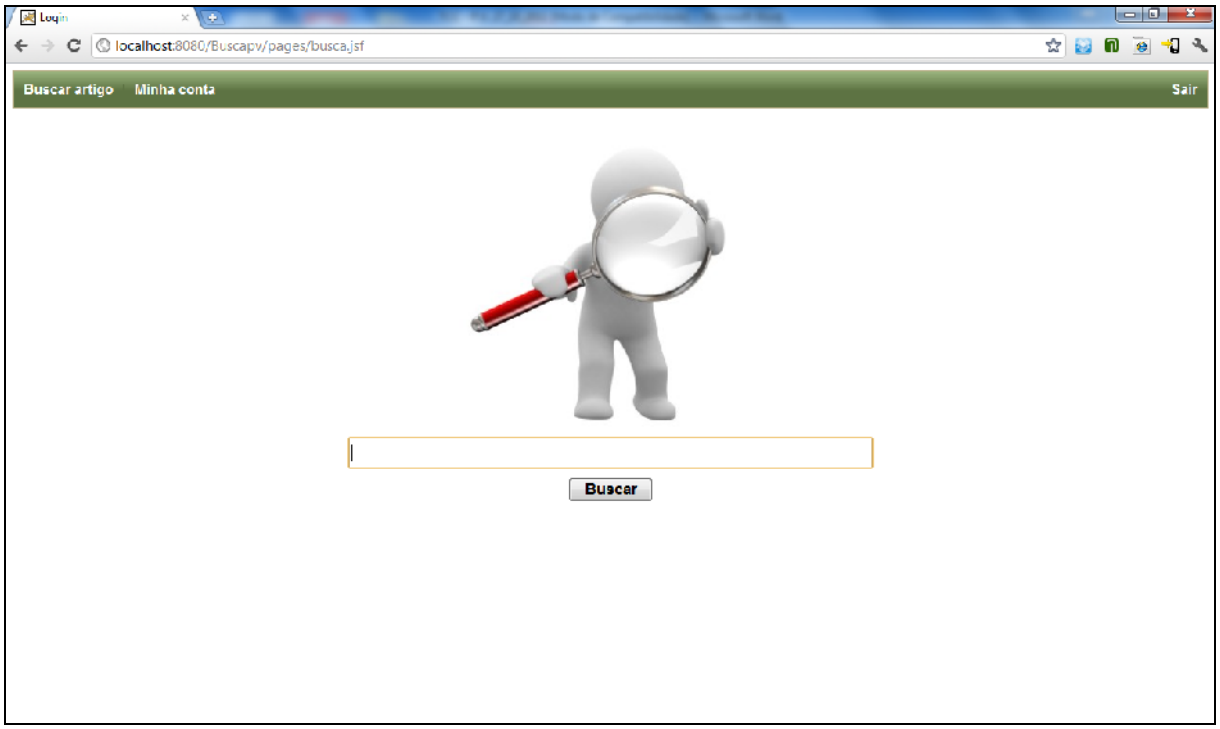

Figura 62 - Tela de busca do sistema
Fonte: Autores

Ao efetuar login, será exibida a tela principal do sistema, onde poderá ser feita a busca dos documentos. Na parte superior da página, é exibido um menu de opções. Para cada perfil de usuário (Administrador, Professor e Aluno), serão exibidas diferentes opções. Todos os perfis terão as opções “Buscar artigo”, “Minha conta” e “Sair”. O perfil Administrador terá uma opção chamada “Gerenciar contas”. E o usuário de perfil Professor terá a opção “Publicar artigo”. Todas as opções de menu citadas serão explicadas a seguir.

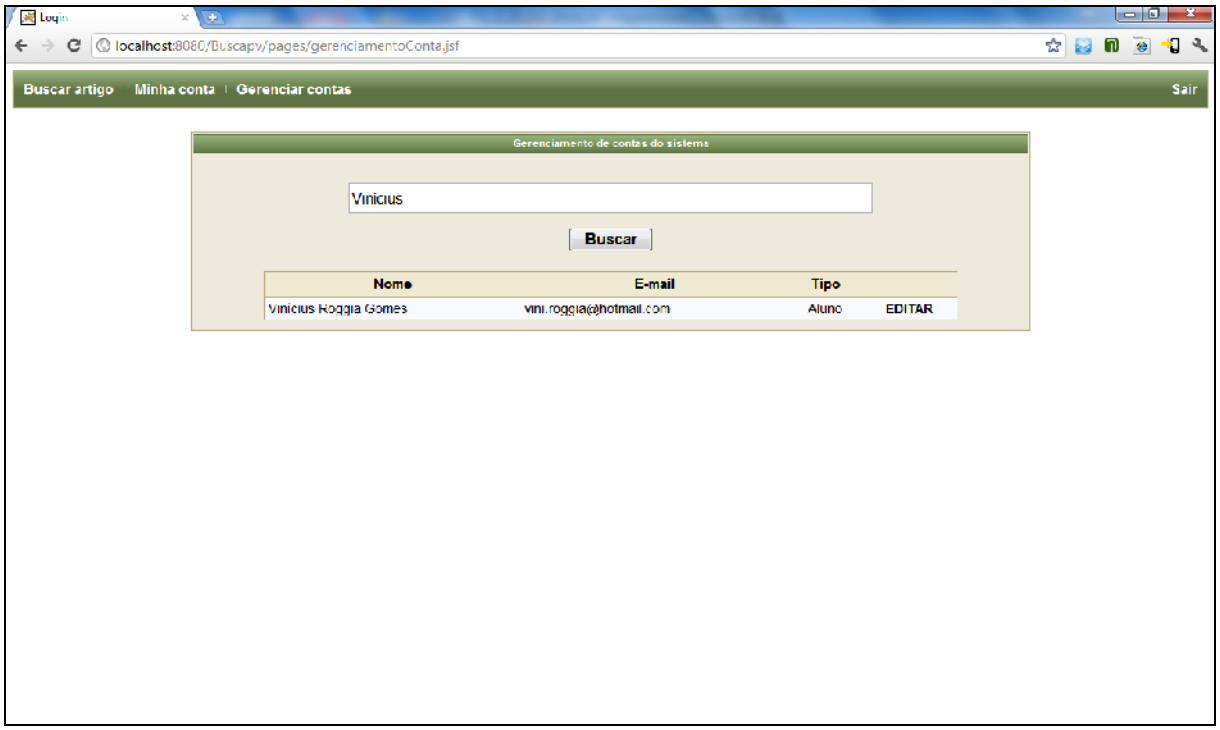

Figura 63 - Tela de gerenciamento de contas do sistema
Fonte: Autores

Esta tela é acessada através do menu “Gerenciar contas”. Apenas o usuário de perfil Administrador tem acesso à mesma. Essa tela tem como objetivo, buscar os usuários pelo nome, exibir os dados dos usuários que atendem à consulta, e permitir editar dados das contas do usuário, tais como nome, data de nascimento. Seu principal propósito é a alteração do tipo de usuário. Importa notar que todo usuário registrado é, inicialmente um usuário do tipo Aluno. Através desta conta e desta opção gerencial é possível passar de um usuário para outro tipo/perfil de usuário. Assim, por exemplo, passar de um usuário recém criado (aluno) para o tipo Professor, com o propósito de permitir que esse usuário publique artigos.

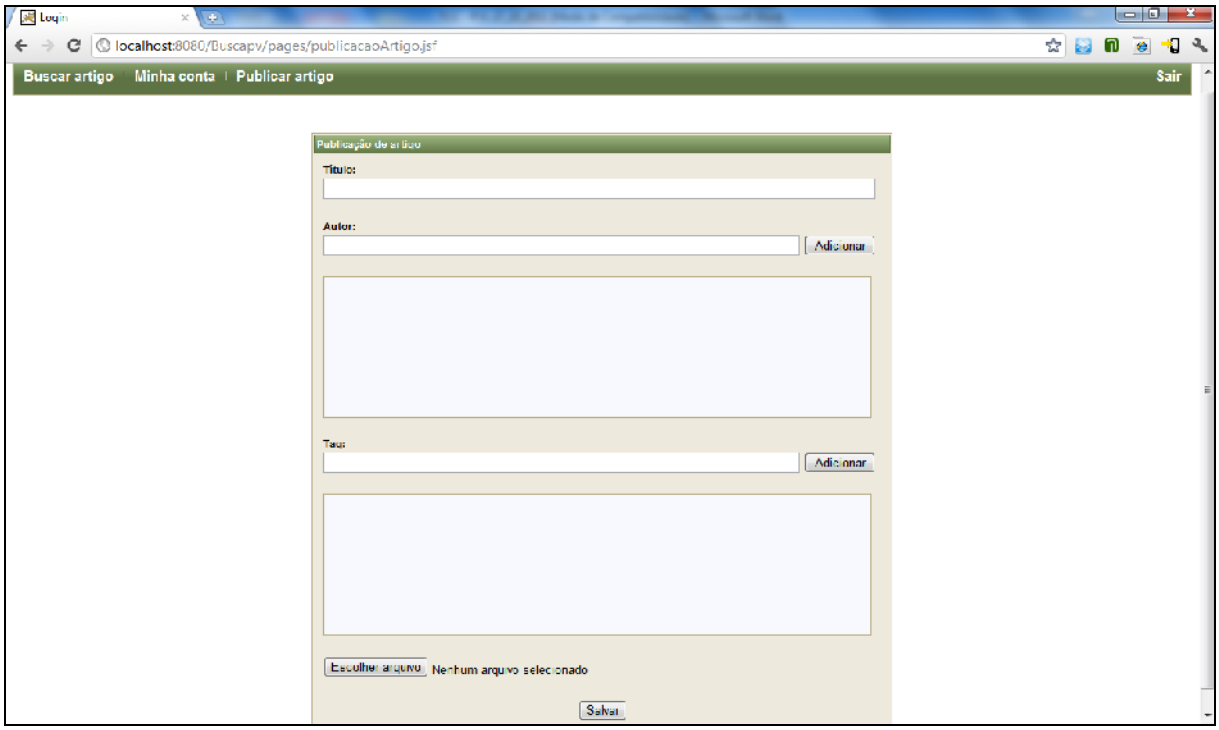

Figura 64 - Tela de publicação de artigo do sistema
Fonte: Autores

Essa é a tela para a publicação de artigos, opção exclusiva do usuário de perfil professor. Pode-se dizer que é uma das telas de maior importância no sistema. Através dessa tela, o usuário irá preencher um formulário, com as informações básicas de um artigo, tais como título, autor(es), tag(s) e, por fim, anexar o documento/artigo, finalizando a operação através do botão salvar. Essa tela irá pegar as informações passadas pelo usuário de perfil professor, salvar num banco de dados, indexar o documento/artigo, e salvar o documento, em um diretório, para depois ser recuperado.

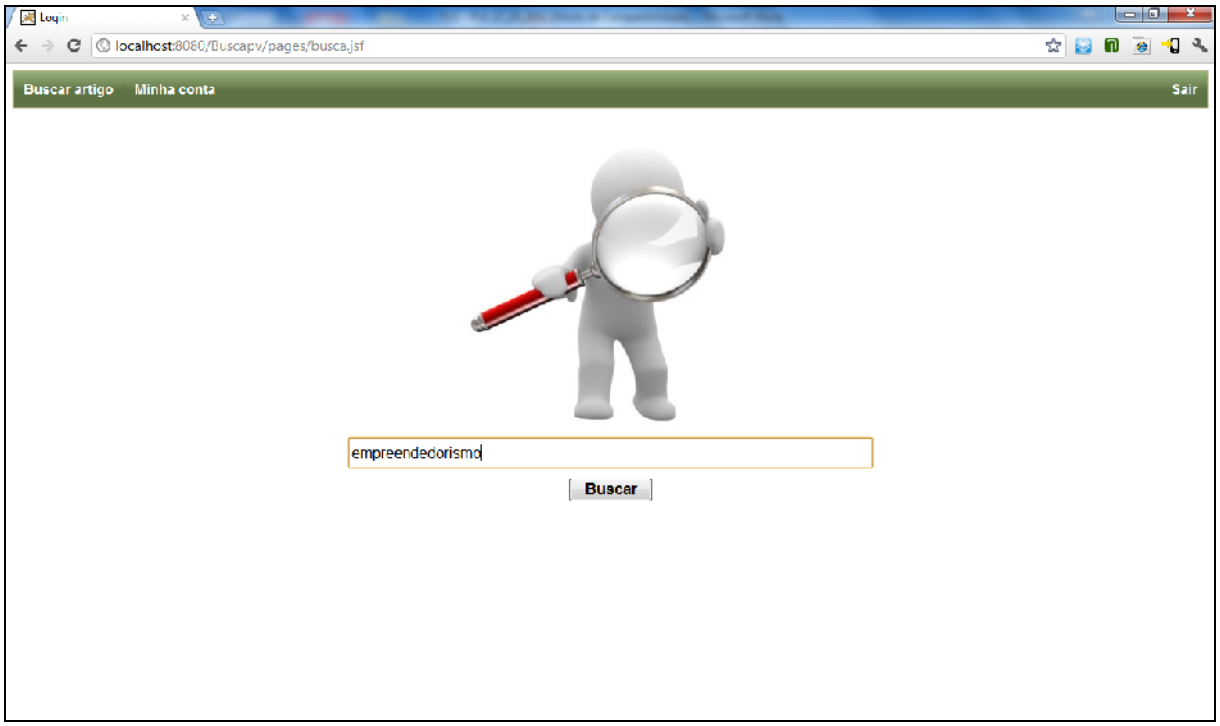

Figura 65 - Tela de busca do sistema com termos
Fonte: Autores

Voltando à tela de “Buscar artigo”, que está disponível para todos os usuários do sistema. Agora que se conhece o procedimento de publicar um artigo, pode-se, através da referida tela, recuperar as informações, através de termos de consulta. Ao inserir um termo, o sistema irá buscar, no índice do SRI, e retornar uma lista de documentos/artigos que tem em seu conteúdo aqueles termos.

Figura 66 - Tela de resultados do sistema
Fonte: Autores

Nessa tela, que é apresentada, após a busca de um documento/artigo por um termo, será apresentada uma lista de documentos/artigos que contém, em seu conteúdo/corpo, os termos informados. Retornando para o usuário o título do artigo, um breve resumo e um link para download do mesmo. Para os usuários de perfil professor, será apresentado, também, um botão "editar", o qual permite que um professor altere os dados do mesmo. Além disso, serão apresentadas informações relacionadas com os documentos/artigos em outros tipos de mídia, como, por exemplo, vídeos (Youtube), imagens (Bing), e uma extensão aos resultados de um dos mais populares buscadores da atualidade, o Google. Esses resultados multimídias, o qual se chama de extensão de consulta, são recuperados, através das tags informadas, ao registrar o artigo na opção "Publicar artigo". Essa tela, além de apresentar diferentes tipos de mídias para uma consulta, permite, ainda, que você expanda um painel para visualizar, apenas, aquele tipo de informação, fazendo, assim, com que aquela janela se expanda e tome conta do restante da tela do navegador/*browser*.

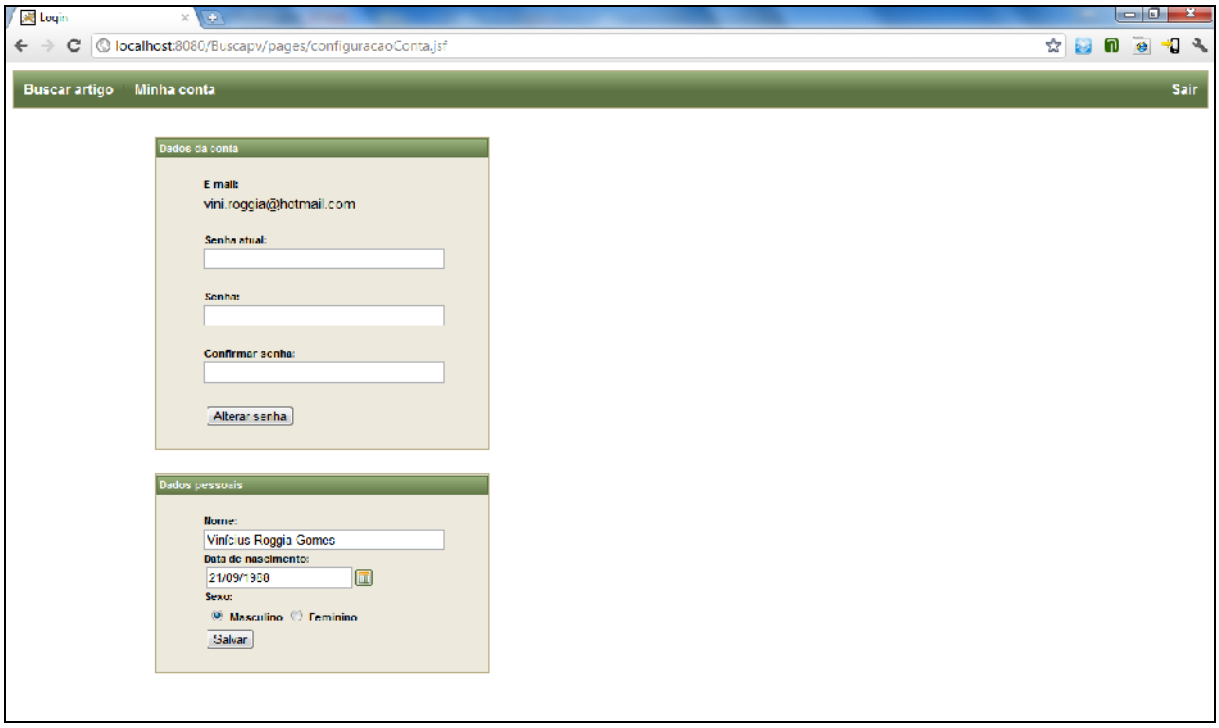

Figura 67 - Tela de gerenciamento de dados da conta do usuário
Fonte: Autores

Por fim, outra opção, que é disponibilizada a todos os perfis de usuário, é o menu "Minha conta". Nesta tela, é possível que o usuário veja o seu login e altere sua senha e dados pessoais, como nome, data de nascimento e sexo.

5.4 VALIDAÇÃO DO SISTEMA

O protótipo de sistema é validado através de entrevistas com o usuário. Assim, através de um questionário, busca-se tanto um resultado qualitativo como quantitativo para a proposta de resolução do problema apresentado. Esse sistema está atrelado a um estudo de caso/caso de teste, apresentado, anteriormente, no qual o cenário se restringe à recuperação de informação de documentos/artigos científicos, que foram, previamente, indexados por um usuário de perfil professor.

5.4.1 Entrevistas com o usuário

Esta validação foi feita através de um questionário, referente ao sistema e à sua proposta de solução. É realizada uma breve apresentação do sistema, seus objetivos e suas funcionalidades. Após isso, o entrevistado é liberado para testar o sistema. Na sequência, ao usuário é apresentado um questionário de 10 questões, que tem como objetivo validar a proposta de solução.

O questionário, apresentado ao entrevistado, apresenta 10 afirmativas, que terão, como resposta, 4 alternativas. Estas são:

1. Não atende.
2. Atende, em partes.
3. Atende.
4. Atende completamente.

As afirmativas encontradas, no questionário, são as seguintes:

1. Efetua o registro de documentos/artigos científicos, e indexa estes para posterior recuperação.
2. Recupera informações de documentos/artigos, previamente, registrados no sistema.
3. O sistema traz resultados multimídias, relacionados à busca do documento/artigo e seus resultados.
4. O sistema traz resultados relevantes aos termos de busca.
5. Tem desempenho satisfatório, quanto ao tempo de busca.
6. O sistema tem interface amigável, ou seja, é fácil manuseá-lo.
7. Apresenta uma forma interessante de exibir resultados.
8. A solução apresentada, neste sistema, facilita encontrar informações pertinentes aos termos de busca.
9. Permite criar uma base de dados de documentos/artigos de fácil manutenção.
10. Em uma pesquisa simples no sistema, geralmente, as informações desejadas são encontradas.

5.4.1.1 Cenário de validação

A amostra é composta por 10 entrevistados, de ambos os sexos, com idades entre 19 e 45 anos. Essa amostra é composta por profissionais das mais diversas áreas, como tecnologia da informação, direito, administração, psicologia e comércio.

O cenário apresentado aos participantes da entrevista foi de uma instituição acadêmica, que tem como objetivo, disponibilizar aos alunos artigos e documentos científicos, de forma que o aluno acesse um site e possa buscar por artigos e documentos científicos através de termos, e, além disso, apresentar material multimídia complementando os resultados da busca.

5.4.1.2 Resultado da validação

O resultado da validação é apresentado, baseado no questionário feito, com uma amostra de 10 pessoas. Os gráficos a seguir exibem o resultado deste questionário.

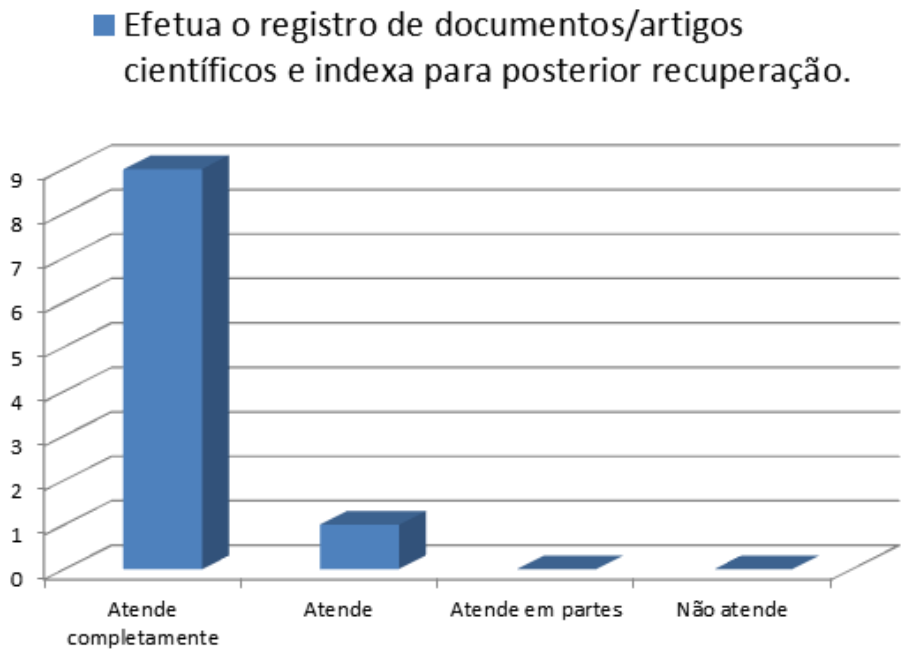

Figura 68 - Questão 1
Fonte: Autores

Com a afirmação encontrada, na Figura 68 – Questão 1, tínha-se como propósito comprovar a eficácia do sistema, quanto ao registro/publicação de artigos e sua indexação. O gráfico comprova que, em 90% dos casos, o sistema atende, completamente, o seu propósito; em 10%, atende, e, em 0% atende em partes ou não atende.

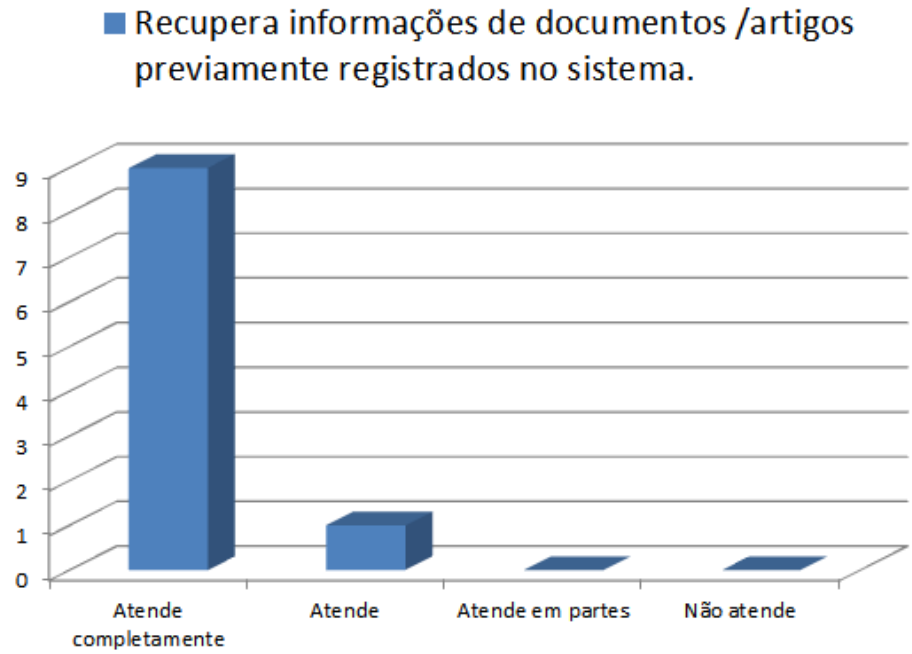

Figura 69 - Questão 2
Fonte: Autores

A Figura 69 – questão 2, comprova que, para 90% dos entrevistados, o sistema atende completamente o requisito de recuperação de informações de documentos; em 10%, atende, e, em 0%, atende em partes ou não atende.

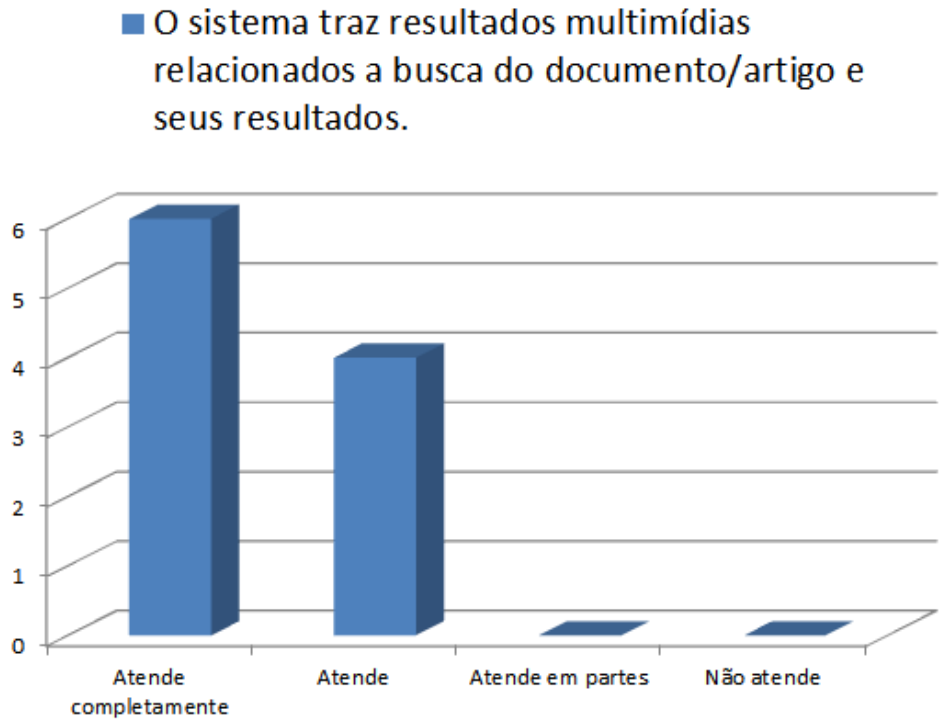

Figura 70 - Questão 3
Fonte: Autores

O gráfico apresentado na Figura 70 – Questão 3, tem como objetivo comprovar que o sistema atende ao objetivo de retornar resultados multimídias, relacionados aos documentos/artigos consultados. Observa-se que, em 60% dos casos, atende completamente; em 40%, atende; e, em nenhum caso, foi constatado que atendia em partes ou não atendia.

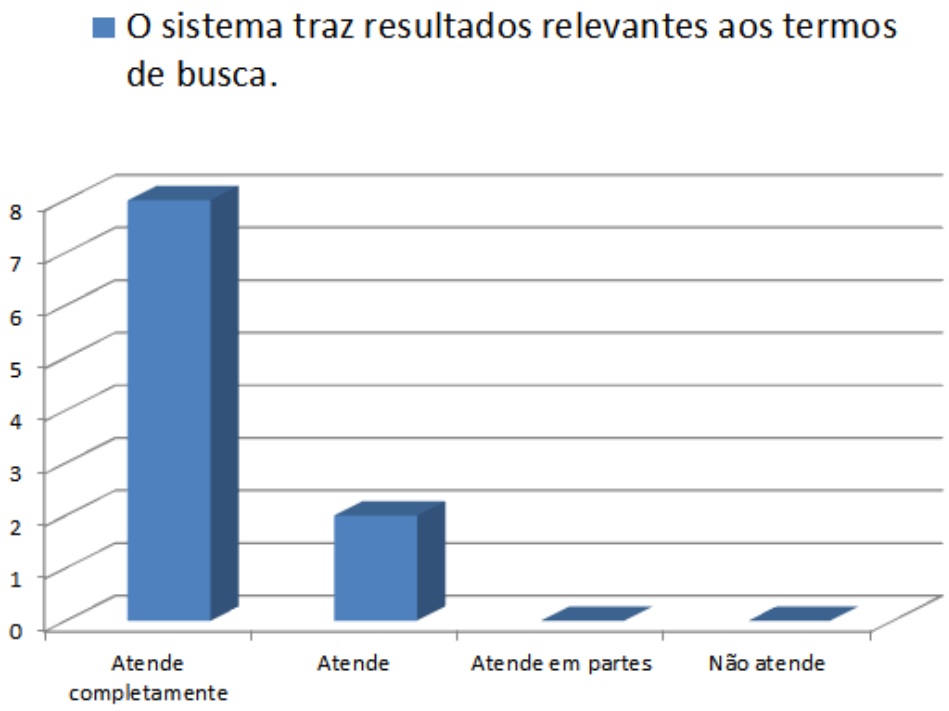

Figura 71 - Questão 4
Fonte: Autores

Um dos fatores mais importantes para recuperação de informação e, principalmente, para um sistema de busca, é se os resultados retornados são relevantes. A Figura 71 – Questão 4, comprova que, para 80% dos entrevistados, este requisito é atendido completamente; em 20%, apenas atende; e, em nenhum caso, foi constatado que atendia em partes ou não atendia.

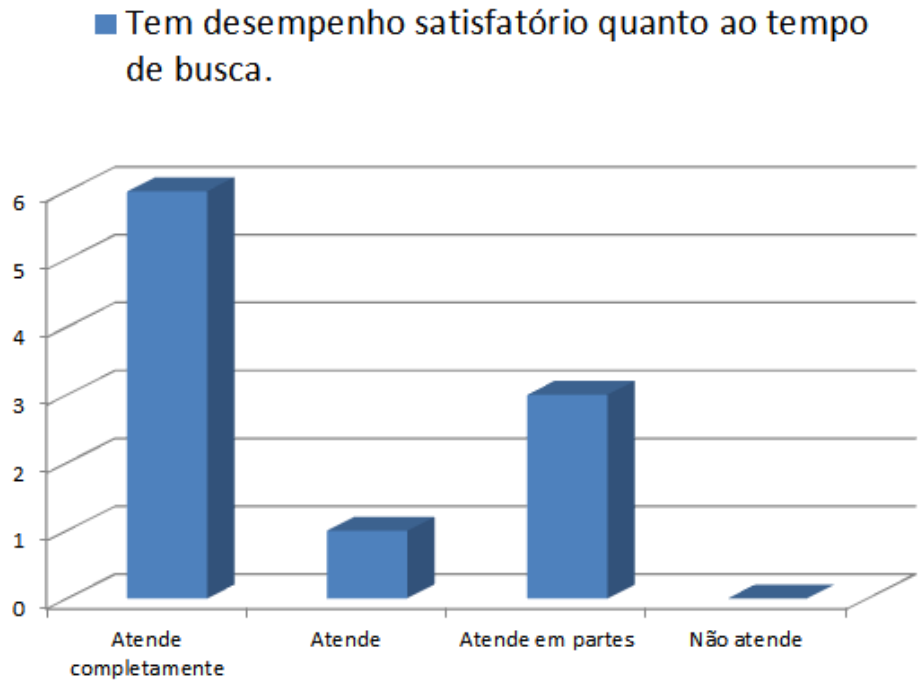

Figura 72 - Questão 5
Fonte: Autores

Outro fator importante para satisfação de usuários, com relação a um sistema de busca é o tempo de resposta. Neste requisito, visualizamos, através da Figura 72 – Questão 5, que, em 60% atende completamente ao requisito; para 10%, atende; para 30%, atende em partes e em nenhum caso não atendeu ao requisito.

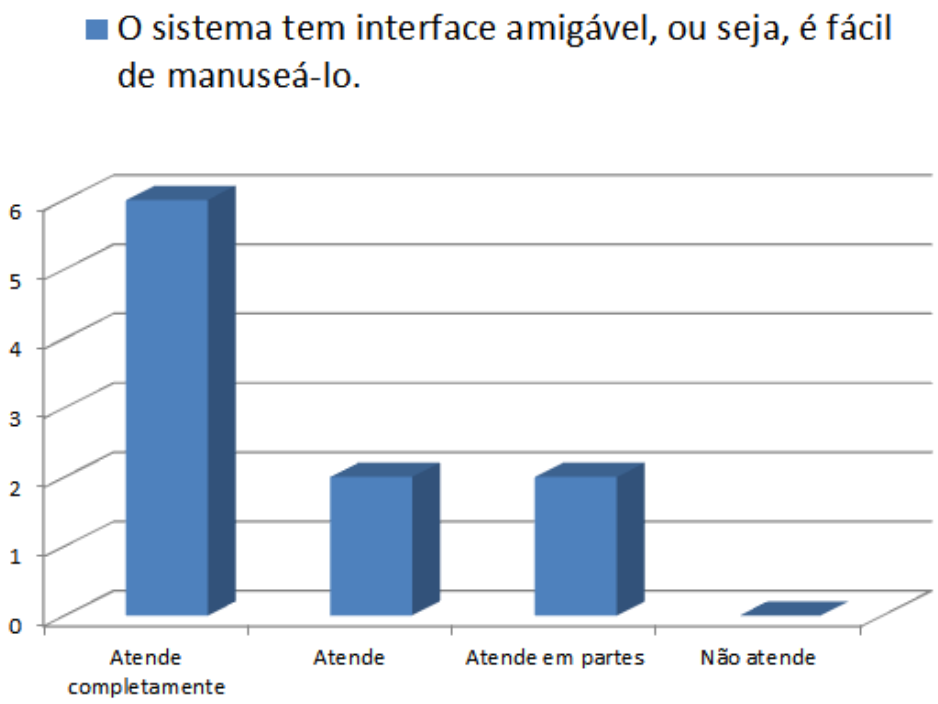

Figura 73 - Questão 6
Fonte: Autores

Uma forma de avaliar qualitativamente um sistema é saber se este tem uma interface amigável, se é fácil de usar, ou seja, autoexplicativo. Neste quesito, a Figura 73 – Questão 6, assinala que, para 60% dos entrevistados atende completamente a este requisito; para 20%, atende; para outros 20%, atende em parte; e, para 0%, não atende.

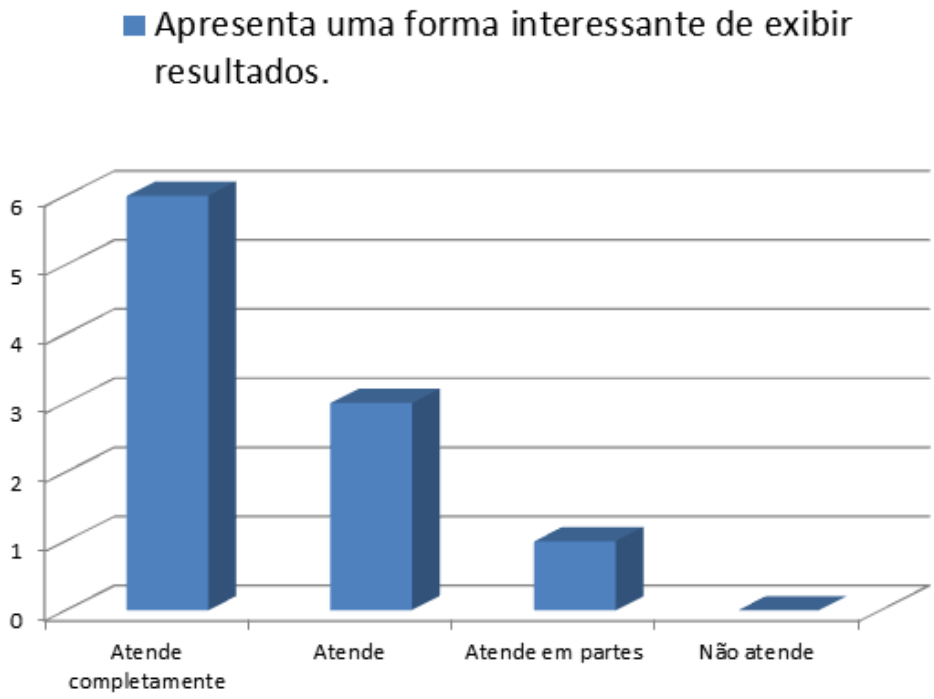

Figura 74 - Questão 7
Fonte: Autores

Esta questão tem como objetivo verificar se a forma de apresentação dos resultados, ou seja, da forma a exibir dados multimídia, relacionados a um documento é interessante. Na Figura 74 – Questão 7, vemos que, em 60%, atende completamente; em 30%, atende; em 10%, atende em partes; e, em 0%, não atende.

Figura 75 - Questão 8
Fonte: Autores

Na questão apresentada na Figura 75 - Questão 8, verifica-se que o sistema facilita a busca de informações. Analisando o gráfico: 80% dos entrevistados, diz que atende completamente a afirmação; para 20%, atende; enquanto para 0%, atende em partes ou não atende.

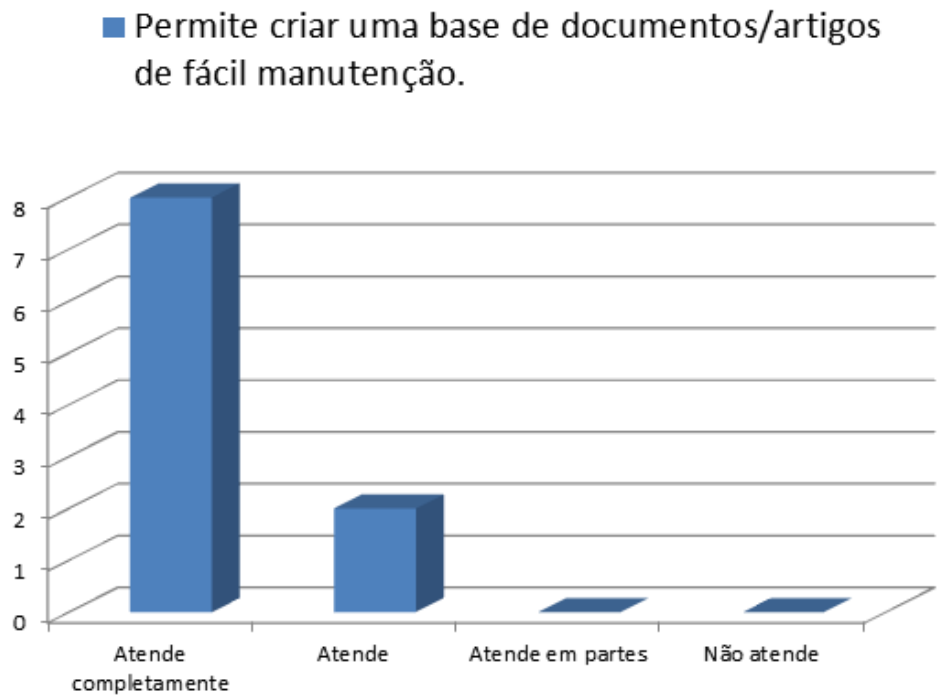

Figura 76 - Questão 9
Fonte: Autores

A Figura 76 – Questão 9, visa verificar se o sistema viabiliza criar e administrar uma base de documentos/artigos. A afirmação é verificada quando o usuário passa de perfil Aluno para perfil Professor. Lembrando que o perfil Professor tem como opção editar e excluir um documento. Para 80% dos entrevistados, a afirmação atende completamente; para 20%, atende; e, para 0%, atende em partes ou não atende.

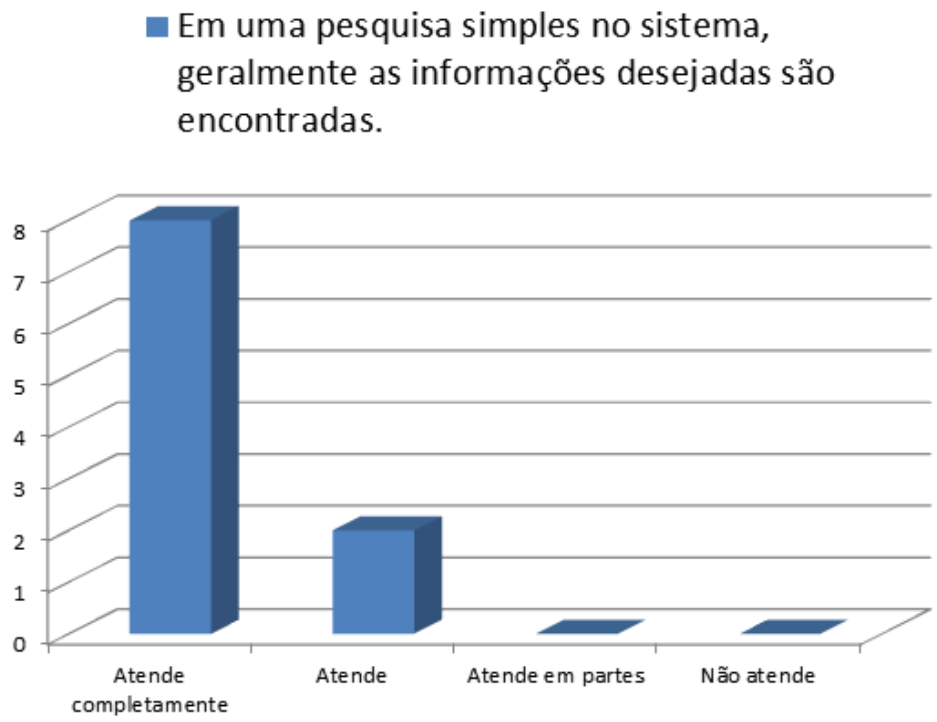

Figura 77 - Questão 10
Fonte: Autores

A última afirmação do questionário apresentada na Figura 77 – Questão 10, visa verificar a principal função do sistema, que é retornar informações desejadas. Para 80%, este requisito é atendido completamente; para 20%, atende; e, para 0%, atende em partes ou não atende.

5.4.2 Caso de teste

Nesta seção, será apresentado um caso de teste, desde a criação de um usuário, até ao resultado da busca. Leva-se em conta um cenário acadêmico, onde os professores poderão publicar, no sistema, artigos de interesse da comunidade acadêmica.

Tanto os usuários Professores, quanto os usuários Alunos poderão realizar uma busca, tendo como resultado artigos que se relacionam com o termo da busca. Junto com os artigos encontrados são “retornados” arquivos multimídia que possuem uma relação relevante com o assunto da pesquisa.

Inicia-se com o cadastro de um usuário. Para isto, basta acessar o sistema e lhe será apresentado, na tela inicial duas opções, Login e Criar conta. Deve-se preencher os campos para criar conta e clicar em cadastrar conforme Figura 78 – Tela de cadastro.

Figura 78 - Tela de cadastro
Fonte: Autores

A seguir, na mesma tela deve-se fazer login com os dados informados para criar a conta. Ao efetuar login no sistema têm-se como principal opção efetuar a busca por um documento. Para este caso de teste, deve-se preencher o campo de busca com o termo "empreendedorismo", conforme Figura 79 – Tela de busca por termo.

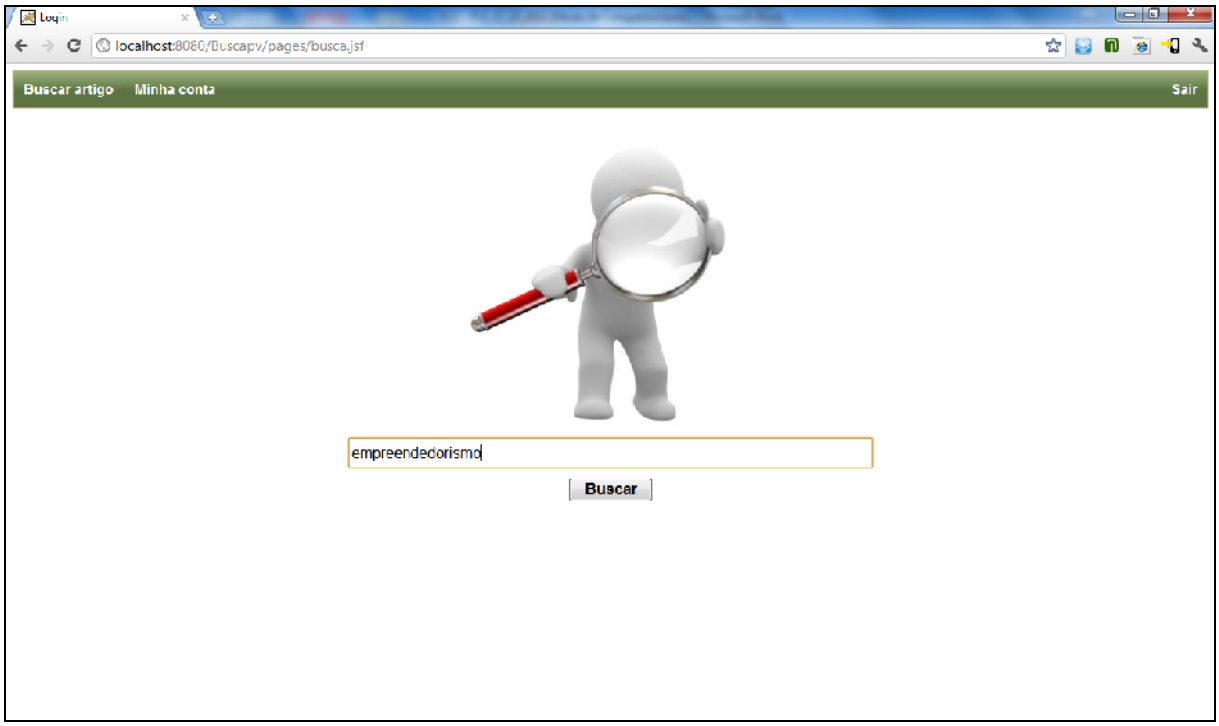

Figura 79 - Tela de busca por termo
Fonte: Autores

Em poucos segundos, será exibido uma tela de resultado, com os documentos/artigos encontrados com aquele termo e a expansão de busca, de acordo com as tags cadastradas para estes documentos. Abaixo, Figura 70 – Tela de resultados por termo, demonstra tela de resultado.

Figura 80 - Tela de resultados por termo
Fonte: Autores

Acima foi demonstrado como um usuário do tipo aluno poderá buscar por artigos. Agora, será demonstrado como um usuário do tipo professor poderá popular essa base, indexando documentos/artigos de interesse. Volta-se à tela de login, e entra-se com um usuário que tem privilégios do tipo professor. Ao efetuar login, será exibida a mesma tela principal do sistema, com uma diferença: um menu na parte de cima que permite publicar artigos. Ao clicar no menu Publicar Artigo, será apresentado ao professor uma tela de cadastro que deverá ser preenchida, com todo cuidado, pois, através dos dados ali informados, poderão ser retornados resultados relevantes ou não à pesquisa. Segue na Figura 81 – Tela de publicação de artigo, a tela de cadastro de artigo.

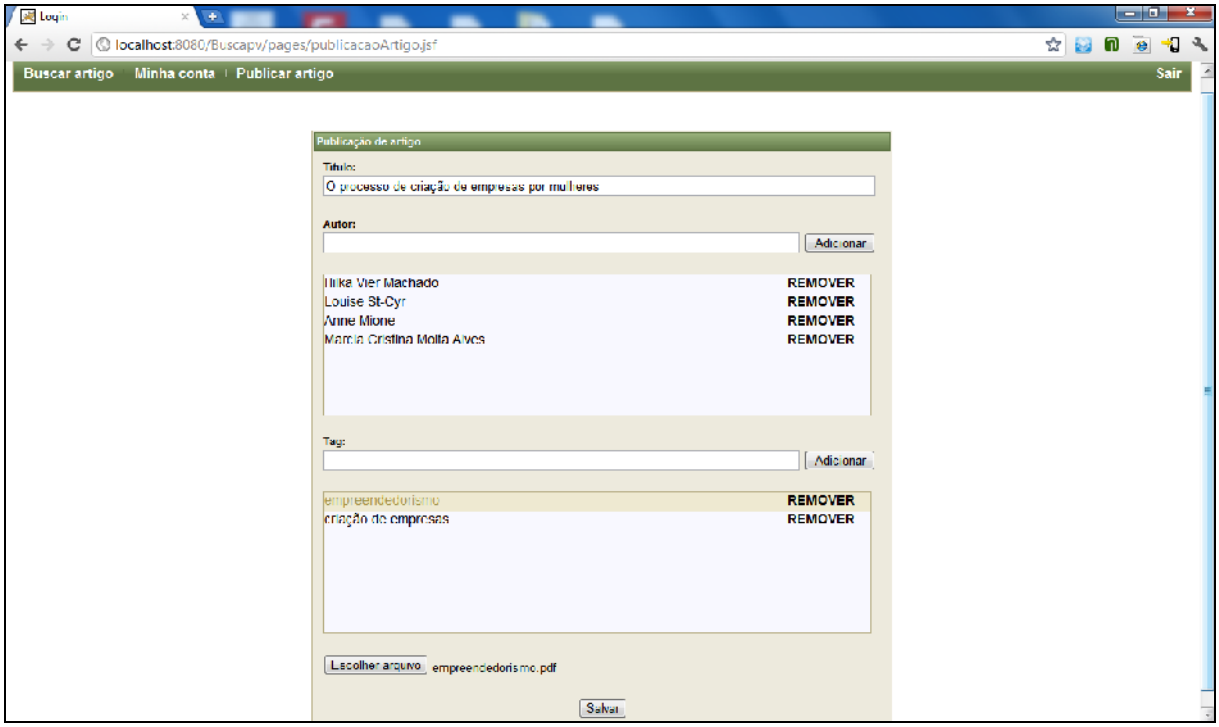

Figura 81 - Tela de publicação de artigo
Fonte: Autores

Nesta seção, foi apresentado um caso de teste, dando uma visão geral do sistema e demonstrando a proposta de solução deste trabalho.

5.5 CONSIDERAÇÕES FINAIS

Este capítulo teve como objetivo apresentar o protótipo de sistema para a solução proposta, exibindo o esquema do sistema e as tecnologias utilizadas. Além disso, foi exibido o sistema tela-a-tela e suas funcionalidades, bem como foi validada a proposta de solução, através de um questionário e seus resultados, além de um caso de teste.

6 CONCLUSÕES E TRABALHOS FUTUROS

Nesse capítulo serão abordados os resultados obtidos com o desenvolvimento do trabalho e opiniões sobre trabalhos futuros, visando melhorar os resultados da busca e melhorar a qualidade do protótipo criado. Os problemas encontrados no desenvolvimento também serão citados, assim como nossa satisfação com os resultados obtidos.

6.1 CONCLUSÃO

Este trabalho apresentou conceitos sobre o tema Recuperação de Informações, tendo como objetivo principal o desenvolvimento de um sistema de RI, que indexa e recupera documentos. Outro objetivo do sistema em questão, - é a realização da expansão de consulta envolvendo tags, relacionadas ao documento indexado.

Para a validação do sistema, foi feito um estudo com base em uma pesquisa realizada por usuários de diferentes perfis, sendo que os mesmos analisaram todas as operações do sistema desenvolvido. As opiniões dos usuários foram analisadas, através de um questionário, levando em consideração a qualidade dos resultados obtidos pela busca, além da facilidade e desempenho do sistema.

O maior problema encontrado foi a possibilidade de um usuário poder adicionar qualquer tipo de tag no documento, sendo que a extensão de consulta não irá trazer bons

resultados se as tags adicionadas não se relacionarem com o tema do documento, fazendo-se assim, necessário que o usuário que irá publicar conteúdo seja treinado.

Com relação ao sistema desenvolvido, pode-se afirmar que, as APIs utilizadas para a extensão de consulta, possuem algumas limitações perante a busca. O caso mais relevante é da API do Youtube, sendo que a base de dados não é muito abrangente, restringindo uma boa parte dos resultados de uma determinada tag. Já as consultas realizadas no Bing Imagens e no Google trazem resultados satisfatórios, com relação às tags relacionadas ao documento.

Analisando os problemas apresentados no início deste trabalho, a proposta de solução, e os resultados obtidos através da pesquisa, pode-se concluir que a proposta é válida, podendo ser usada em outros cenários e não apenas no apresentado no estudo de caso. A proposta se apresenta como uma alternativa aos buscadores atuais, trazendo vários tipos de dados relacionados com os resultados obtidos.

Pode-se concluir também que a técnica de Extensão de Consulta, utilizada no sistema desenvolvido, mostrou-se uma técnica interessante, mesmo não sendo utilizada pelos buscadores mais conhecidos atualmente. A extensão de consulta de certa forma enriquece as informações retornadas de uma busca.

É valido afirmar também que o sistema entra no escopo da Web 2.0, pois permite que os usuários compartilhem informações de uma maneira mais fácil, além de permitir a visualização de mais de um tipo de dado em uma mesma tela de uma forma clara e organizada.

Com base nos resultados da pesquisa, pode-se, então, concluir que o sistema atende, completamente, à maior parte de seus objetivos. O sistema recupera corretamente um artigo que foi publicado, sendo utilizadas palavras que estejam no corpo do artigo como termo para a pesquisa. Também traz vídeos, imagens e outros links relacionados às tags (palavras-chaves) pertencentes aos artigos retornados pela busca. Entretanto, o sistema não possui um desempenho satisfatório com relação ao tempo de busca. Em alguns casos não são retornados vídeos no resultado, devido à limitação do próprio Youtube referente às informações que o mesmo possui em sua base de dados. E por fim, foram encontradas, no sistema, algumas pequenas falhas em relação à interface com o usuário.

6.2 TRABALHOS FUTUROS

Tendo em vista possíveis trabalhos futuros, pode-se citar que a expansão da consulta pode ser melhorada, trazendo resultados mais relacionados aos documentos indexados. Essa melhoria poderá ser feita, através de um algoritmo que verifique, antes de realizar a consulta da expansão, as tags que mais se identificam com o termo da busca e o documento indexado.

O desenvolvimento de um algoritmo para anotação automática, nos documentos, seria de grande ajuda, evitando que o usuário adicione palavras-chaves que não se relacionem com o documento. Outra alternativa seria a possibilidade de o usuário escolher as tags mais relevantes, dentro de uma lista de tags, identificadas pelo sistema, utilizando desse modo, o conceito de anotação semiautomática.

Para contornar o problema das APIs, na extensão de consulta, seria interessante a criação de um *crawler* para recuperar os dados na web. Por fim, poderiam ser utilizados conceitos de busca semântica, como ontologias, para obtermos melhores resultados na busca das informações relacionadas ao documento.

REFERÊNCIAS

AQUINO, Maria Clara. Hipertexto 2.0, folksonomia e memória coletiva: Um estudo das tags na organização da web. **E-Compós**, v. 9, ago. 2007. Disponível em: <http://www.compos.org.br/seer/index.php/e-compos/issue/view/9>. Acesso em: 07 maio 2011.

BAEZA-YATES, Ricardo A.; RIBEIRO-NETO, Berthier. **Modern information retrieval**. New York: ACM Press, 1999.

BARROS, Camila Monteiro de; VIERA, Angel Freddy Godoy. MPEG-7 e a recuperação da informação de objetos multimídia. **Inf. & Soc.:** Est., João Pessoa, v. 20, n. 3, p. 135-144, set./dez. 2010. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/7337>. Acesso em: 15 fev. 2011.

BATISTA, Carlos Eduardo C. F.; SCHWABE, Daniel. LinkedTube: Informações Semânticas em Objetos de Mídia da Internet. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB, 15., 2009, Fortaleza. **Anais eletrônicos**... Disponível em: <http://www2.telemidia.puc-rio.br/telemidia/publicacao_por_tipo.jsp ?tipo=tp_congresso&idioma=pt>. Acesso em: 21 mar. 2011.

BELL, Donald. UML basics: An introduction to the Unified Modeling Language. Disponível em: < http://www.ibm.com/developerworks/rational/library/769.html >. Acesso em: 21 mai. 2011.

BEPPLER, Fabiano Duarte. **Emprego de RBC para recuperação inteligente de informações**. 2002. 100 f. Dissertação (Mestre em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2002.

BETTIO, Raphael Winckler de. **Interrelação das Técnicas Term Extraction e Query Expansion Aplicadas na Recuperação de Documentos Textuais**. 2007. 99 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Universidade Federal de Santa Catarina, Florianópolis, 2007.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web. **Scientific American**, maio 2001.

BITTENCOURT, Guilherme; FREITAS, Frederico Luiz G.; SILVA, Tércio de M. Sampaio. Extração de Informação no Master-Web Baseada em Ontologias. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (RESUMOS), 13., 2002, São Leopoldo.

BLATTMANN, Ursula; SILVA, Fabiano Couto Corrêa. Colaboração e Interação na Web 2.0 e Biblioteca 2.0. **Revista ACB**: Biblioteca em Santa Catarina, Florianópolis, v. 12, n. 2, p. 191/215, jul./dez. 2007.

CARDOSO, O. N. P. **Recuperação de Informação**. 6 p. Departamento de Ciência da Computação – Universidade Federal de Lavras, Lavras, 2002.

CARDOSO, O. N. P. Recuperação de informação. **Journal of Computer Science**, v.2, n.1, 2000.

CECI, Flávio. **Um modelo semiautomático para a construção e manutenção de ontologias a partir de bases de documentos não estruturados**. 2010. 129 f. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina, Florianópolis, 2010.

CERVO, A. L.; BERVIAN, P. A. **Metodologia Científica**. São Paulo: Prentice Hall, 2002. P.65.

CHANG, S. et al. Overview of the MPEG-7 Standard. IEEE transactions on circuits and systems for video technology, v. 11, n. 6, jun. 2001. Disponível em: <http://www.img.lx.it.pt/~fp/cav/Additional_material/MPEG7_overview_1.pdf> Acesso: 05 mai. 2011.

CHELLA, M. T. Sistema para classifi cação e recuperação de conteúdo multimídia baseado no padrão MPEG-7. UNICAMP: São Paulo, 2004. Disponível em: < http://www.nied.unicamp.br/~siros/doc/2232.pdf > Acesso em: 05 mai. 2011.

CHRISTOPHER D. M.; RAGHAVAN, R.; SCHÜTZE, H. Introduction to Information Retrieval, Cambridge University Press. 2008.

COELHO, Alexandre Ramos. **Stemming para a língua portuguesa**: estudo, análise e melhoria do algoritmo RSLP. 2007. 69 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

COTRIN, Gilberto; DEMO, Pedro; PILLETTI, Nelson; OLIVEIRA, Claudino de. **Fundamentos da Filosofia**: ser, saber e fazer. Reformulado. Saraiva, 2002.

COUTINHO, Clara Pereira; BOTTENTUIT JUNIOR, João Batista. Blog e Wiki: os futuros professores e as ferramentas da web 2.0. In: SIMPÓSIO INTERNACIONAL DE INFORMÁTICA EDUCATIVA, 9., 2007, Portugal. Actas… Portugal, 2007. p. 199-204.

DALLACOSTA, A; et al. A utilização da indexação de vídeos com MPEG-7 e sua aplicação na educação. Novas Tecnologias na Educação, Porto Alegre, v. 2, n. 1, mar. 2004. Disponível em: <http://www.cinted.ufrgs.br/ciclo3/af/35-autilizacao.pdf > Acesso em: 05 mai. 2011.

DENNIS, Simon; BRUZA, Peter; MCARTHUR, Robert. Web searching: a process oriented experimental study of three interactive search paradigms. **Journal of the American Society for Information Science and Technology**, v. 53, n. 3, p. 120-133, 2002.

DIAS, Maria Abadia Lacerda. Extração **Automática de Palavras-chave na Língua Portuguesa Aplicada a Dissertações e Teses da Área das Engenharias**. 2004. 127 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, Campinas, 2004.

DIAS, T. D; SANTOS, N. Web Semântica: Conceitos Básicos e Tecnologias Associadas. **Cadernos do IME Série Informática**, 14:p. 25–38, 2003.

DODEBEI, Vera Lúcia Doyle. **Tesauro**: linguagem de representação da memória documentária. Rio de Janeiro: Editora Interciência, 2002.

EIKVIL, Line. **Information Extraction from World Wide Web**: A Survey. Norwegian Computing Center Jul., 1999. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1. 1.41.4905&rank=2>. Acesso em: 9 abr. 2011.

ELLER, Markus Pereira. **Anotações Semânticas de Fontes de Dados Heterogêneas**: Um Estudo de Caso com a Ferramenta Smore. 2008. 89 f. Trabalho de Conclusão de Curso (Graduação em Sistemas da Informação) – Universidade Federal de Santa Catarina, Florianópolis, 2008.

ERIKSSON, Hans-Erik; PENKER, Magnus. **Business Modeling with UML**. Estados Unidos: Wiley & Sons, 2000. 459p.

FACHIN, Odília. **Fundamentos de metodologia**. 3. ed. São Paulo: Saraiva, 2001.

FEDELI, Ricardo Daniel; POLLONI, Enrico Giulio Franco; PERES, Fernando Eduardo. **Orientação a Objeto com Prototipação**. São Paulo: Pioneira Thomson, 2002.

FERNANDES, Ricardo Madeira. **GeoSen_Tags:** um motor de busca geográfico com suporte a Tags. 2010. 107 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Campina Grande, 2010.

FERNEDA, Edberto. **Recuperação de Informação**: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 137 f. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2003.

FLORES, Felipe Nunes. **Avaliando o Impacto da Qualidade de um Algoritmo de Stemming na Recuperação de Informações**. 2009. 48 f. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

FOWLER, Martin. **UML essencial**: um breve guia para a linguagem-padrão de modelagem de objetos. Trad. João Tortello, 3 ed. Porto Alegre: Bookman, 2005.

FRAKES, William B.;FOX, Christopher J. **Strength and similarity of affix removal stemming algorithms**. ACM SIGIR Forum, Volume 37 Issue 1, April 2003.

FRAKES, Willian B.; YATES, Ricardo Baeza-. **Information Retrieval**: Data Structures & Algorithms. New Jersey: Prentice-Hall, 1992.

FURLAN, José David. **Modelagem de objetos através da UML**. São Paulo: Makron Books, 1998.

GALHO, Thaís Silva; MORAES, Silva Maria Wanderley. **Categorização Automática de**

Documentos de Texto Utilizando Lógica Difusa. 2003. 75 f. Monografia (Bacharelado em Ciência da Computação) – Universidade Luterana do Brasil, Gravataí, 2003.

GALLIANO, A. G. **O método científico**: teoria e prática. São Paulo: Harbra, 1986.

GARCIA, E. A. C**. Manual de sistematização e normalização de documentos técnicos**. São Paulo: Atlas, 1998.

GEARY, David.; CAY, Horstmann. Core JavaServer Faces. Alta Books. 2005. p.1-234.

GEARY, David.; CAY, Horstmann. Core JavaServer Faces – Segunda Edição. Alta Books. 2007. p.1-234.

GERALDO, André Pinto. **Aplicando Algoritmos de Mineração de Regras de Associação para Recuperação de Informações Multilíngues**. 2009. 76 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

GONÇALVES,Danilo Brandão; JUNIOR, José Cláudio Vahl. **Web 2.0 – Frameworks de desenvolvimento**.

GONZALEZ, Marco; LIMA, Vera L. S. de. Recuperação de Informação e Processamento da Linguagem Natural. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 23., 2003, Campinas. **Anais**... Campinas: Mini Cursos de Inteligência Artificial, 2003, p. 347 – 395.

GROBMAM, Rafael. Sistema de Busca de Informações Baseado nos Conceitos da Web Semântica. **Anuário da Produção de Iniciação Científica Discente**, V.12, n. 15, p. 311-328, 2009. Disponível em: <http://sare.anhanguera.com/index.php/anuic/ article/view/2438>. Acesso em: 10 abr. 2011.

HATCHER, Erik; GOSPODNETIC, Otis. Lucene in action. Greenwich: Manning Publications, 2005.

Hibernate (2011): Página oficial do framework Hibernate, disponível em <http://www.hibernate.org/>. Última visita em 26/09/2011.

IANNELA, R.; WAUGH, A. **Metadata**: enabling the Internet. 1997

IGARASHI, Wagner. **Construção automática de vocabulários temáticos e cálculo de aderência curricular: uma aplicação aos fundos setoriais**. 2005. 95 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) - Universidade Federal de Santa Catarina, Florianópolis, 2005.

ISOTANI, Seij et al. Web 3.0: Os Rumos da Web Semântica e da Web 2.0 nos Ambientes Educacionais. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 19., 2008, Fortaleza. **Anais eletrônicos**... Disponível em: <http://www.proativa.virtual.ufc.br/~sbie/CD_ROM_COMPLETO/cd.html>. Acesso em: 21 mar. 2011.

Java (2011): Página oficial da plataforma Java, disponível em <http://www.java.com/pt_BR/download/faq/whatis_java.xml>. Última visita em 18/09/2011.

JACKSON, Peter; MOULINIER Isabelle. **Natural language processing for online applications – text retrieval, extraction and categorization**. Philadelphia, PA, USA: John Benjamins Publishing Company, 2002.

JESUS, Rui Manuel Feliciano de. **Recuperação de Informação Multimédia em Memórias Pessoais**. 2009. 220 f. Tese (Doutorado) – Universidade Nova de Lisboa, Lisboa, 2009.

JONES, Meilir Page-. **O que todo programador deveria saber sobre projeto orientado a objeto**. São Paulo: Makron Books, 1997.

JONES, K. Sparck; WALKER, S.; ROBERTSON, S.E. **A probabilistic model of information retrieval**: development and comparative experiments. Cambridge, jan. 2000. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.134.6108&rep=rep1&type=pdf>. Acesso em: 26 abr. 2011.

JOVANOVIC, Jelena.; TORNIAI, Carlo; GASEVIC, Dragan; BATEMAN, Scott; HATALA, Marek. Leveraging the Social Semantic Web in Intelligent Tutoring Systems. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2008. **Proceedings…** Springer: Verlag Berlin Heidelberg, 2008, p. 563 – 572.

KIMMEL, Paul. **UML Demystified**. McGraw-Hill, Única ed., 2005, P.235.

KING, G. & BAUER, C. (2005): Hibernate in Action. Manning Publishing Co., 2005.

LEITE, Maria Angelica de Andrade. **Modelo Fuzzy para Recuperação de Informação Utilizando Múltiplas Ontologias Relacionadas**. 2009. 164 f. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, 2009.

LUCENE PROJECT (2011). Página do projeto Lucene, disponível em <http://lucene.apache.org/>. Última visita em 20/09/2011.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHUTZE, Hinrich. **Introduction to information retrieval**. New York: Cambridge University Press, 2008.

MARINHO, Leandro Balby; GIRALDI Rosario. Mineração na Web. **Sociedade Brasileira de Computação**: Revista Eletrônica de Inicialização Científica. São Luis, v. 3, n. 2, jun. 2003. Disponível em: <http://143.54.31.10/reic/edicoes/2003e2/>. Acesso em: 9 abr. 2011.

MARON, M. E., and J. L. KUHNS. 1960. "On Relevance, Probabilistic Indexing and Information Retrieval." **J. Association for Computing Machinery**, 7(3), 216-44.

MARTINS, Claudia A.; MONARD, Maria Carolina, MATSUBARA, Edson T. Uma Metodologia para Auxiliar na Seleção de Atributos Relevantes usados por Algoritmos de Aprendizado no Processo de Classificação de Textos. In: CONFERENCIA LATINOAMERICANA DE INFORMÁTICA, 30., 2004, Peru. **Anais**... Peru: Sociedad Peruana de Computación, 2004, p. 21 – 32.

MEYER, Bertrand; COLEMAN, Derek; ARNOLD, Patrick; BODOFF, Stephanie; DOLLIN, Chris; GILCHRIST, Helena; HAYES, Fiona; JEREMAES, Paul. **Desenvolvimento Orientado a Objetos**: o método fusion. Rio de Janeiro: Campus, 1996.

MURAKAMI, Tiago Rodrigo Marçal. **Tesauros e a World Wide Web**. 2005. 75 f. Monografia (Bacharel em Biblioteconomia e Documentação) – Universidade de São Paulo, São Paulo, 2005.

MURUGESAN, San. Understanding Web 2.0. **IT Professional**, New Jersey, jul. 2007. v. 9, ed. 4, p. 34 – 41

O'REILLY, T. What is web 2.0: design patterns and business models for the next generation of software. 2005. Disponível em <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html# mememap>. Acesso em: mai. 2011.

PAL, Sankar K.; TALWAR, Varum; MITRA, Pabitra. Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. **IEEE Transactions on Neural Networks**, v. 13, n. 5, p. 1163 – 1177, Set. 2002. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/ summary?doi=10.1.1.11.6928&rank=1>. Acesso em: 9 abr. 2011.

PITA, Marcelo; PAIXÃO, Goedson Teixeira. "Arquitetura de Busca Semântica para Governo Eletrônico". In: **II Workshop de Computação Aplicada em Governo Eletrônico & Congresso da Sociedade Brasileira de Computação**, 2010, Belo Horizonte.

RAMALHO, Franklin; ROBIN, Jacques. **Avaliação empírica da expansão de consultas baseada em um thesaurus**: aplicação em um engenho de busca na web. RITA 10 (2004), no. 2, 9-28.

RAMIRO, Thiago Bortolo; MENEZES, Crediné Silva de; CURY, Davidson; NEVADO, Rosane Aragon de. Uma Ferramenta Web para Gerência de Anotações em Documentos. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 16., 2005, Juiz de Fora. **Anais**... Disponível em: <http://www.br-ie.org/pub/index.php/sbie/issue/view/25>. Acesso em: 06 maio 2011.

REEVE, Lawrence; HAN, Hyoil. Survey of Semantic Annotation Platforms. In: ACM Symposiun on Applied Computing, 2005, Santa Fé.

RIZZI, Claudia Brandelero; WIVES, Leandro Krug; OLIVEIRA, José Palazzo Moreira de; ENGEL, Paulo Martins. Fazendo uso da categorização de textos em atividades empresariais. In: INTERNATIONAL SYMPOSIUM ON KNOWLEDGE MANAGEMENT/DOCUMENT MANAGEMENT, 3., 2000, Curitiba. **Proceedings**... Curitiba: PUC-PR, 2000. p. 125-147.

ROSENBERG, Doug; STEPHENS, Matt; COPE, Mark Collins-. **Agile Development with ICONIX Process**: People, Process, and Pragmatism. New York: Apress , 2005.

ROSS, Ronald G.; Principles of the Business Rule Approach. Boston: Addison-Wesley, 2003.

ROSSINI, Tiago; MEDEIROS, Rodrigo; GUIMARÃES, Gabriel; SILVA, Gilbert; SILVA, George. Utilizando ICONIX no desenvolvimento de aplicações delphi. In: CONGRESSO DE PESQUISA E INOVAÇÃO DA REDE NORTE NORDESTE DE EDUCAÇÃO TECNOLÓGICA, 2., 2007, João Pessoa. **Anais eletrônicos**... Disponível em: <http://www.redenet.edu.br/publicacoes/publicacoes.php?tipo=1&area2=Inform%E1tica#>. Acesso em: 20 maio 2011.

SALTON, G.; McGill, M.J. "**Introduction to Modern Information Retrieval**". McGraw-Hill, New York, NY, 1983.

SANTARÉM, José Eduardo Segundo; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Representação iterativa e folksonomia assistida para repositórios digitais. **Liinc em Revista,** Rio de Janeiro, v. 7, n. 1, p. 283-300, mar., 2011. Disponível em: < http://revista.ibict.br/liinc/index.php/liinc/article/view/414/294 >. Acesso em: 28 ago. 2011

SEIBEL JÚNIOR, Hilário. **Recuperação de informações relevantes em documentos digitais baseada na resolução de anáforas**. 2007. 91 f. Dissertação (Mestrado em Informática) – Universidade Federal do Espírito Santo, Vitória, 2007.

SILVA, Alberto; VIDEIRA, Carlos; UML Metodologias e Ferramentas CASE, Centro Atlântico, 2001

SILVA, Tércio de Morais Sampaio. **Extração de Informação para Busca Semântica na Web Baseada em Ontologias**. 2003. 79 f. Dissertação (Curso de Pós-Graduação em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, 2003.

SOMMERVILLE, Ian. **Engenharia de Software**. 8ª ed. São Paulo: Pearson Addison-Wesley, 2007.

SOUZA, Aleksandro Barboza et al. Recuperação Semântica de Objetos de Aprendizagem: Uma Abordagem Baseada em Tesauros de Propósito Genérico. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 19., 2008, São Paulo. **Anais eletrônicos...** Disponível em: <http://www.br-ie.org/pub/index.php/sbie/ issue/view/28>. Acesso em: 10 abr. 2011.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. Ci. Inf., Brasília, v.33 n.1, p. 132-141, jan./abril. 2004.

SOUZA, Renato Rocha. Sistemas de Recuperação de Informações e mecanismos de Busca na web: panorama atual e tendências. Perspect. Ciênc. Inf., Belo Horizonte, v.11 n.2, p. 161-173, mai./ago. 2006

SPARCK-JONES, K.; WILLET, P. (editores). Readings in Information Retrieval. California: Morgan Kaufmann Publishers, Inc., 1997.

SPARX SYSTEMS (2011): Site oficial da ferramenta, disponível em <http://www.sparxsystems.com/products/index.html>. Último acesso em 27/09/2011.

VAN RIJSBERGEN, C. J. 1979. **Information Retrieval**. London: Butterworths.

VICENTE, P. J. V. El estándar MPEG-7. Revista de Ingeniería Informática del CIIRM, Murcia (Espanha), n.3, p. 1-5, 2005. Disponível em: < http://www.cii-murcia.es/informas/jul05/articulos/El_estandar_MPEG-7.pdf > Acesso em: 03 set. 2011.

VIDOTTI, Silvana Aparecida Borsetti; SEGUNDO, José Eduardo Santarém. Representação iterativa e folksonomia assistida para repositórios digitais. **Liinc em Revista**, Rio de Janeiro, v. 7, n. 1, p. 283-300, mar. 2011.

XU, J.; CROFT, W. B. Query expansion using local and global document analysis. In: SIGIR '96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM Press, 1996. p. 4–11. ISBN 0-89791-792-8

W3C Semantic Web Activity. Disponível em: <http://www.w3.org/2001/sw/>. Acesso em: 07 mai. 2011.

WIEGERS, Karl E. More about Software Requirements. Redmont, Washington: Microsoft Press, 2006.

WIVES, Leandro Krug; LOH, Stanley. Recuperação de Informações usando a Expansão Semântica e a Lógica Difusa. In: CONGRESSO INTERNACIONAL EM INGENIERIA INFORMATICA, Abr, 1998. **Proceedings**... Buenos Aires: Universidad de Buenos Aires, 1998.

WIVES, Leandro Krug. Tecnologias de Descoberta de Conhecimento em Textos Aplicadas à Inteligência Competitiva. 2002. 116 f. Dissertação (Pós Graduação em Computação)-Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

YAN, Rong; NATSEV, Apostol; CAMPBELL, Murray. An efficient manual image annotation approach based on tagging and browsing. In: MS '07: Workshop on Multimedia Information Retrieval on The Many Faces of Multimedia Semantics, 2007, Augsburg. Disponível em: <http://www.wjh.harvard.edu/~mahesh/image%20search%20and%20 annotation/p13-yan.pdf>. Acesso em: 05 maio 2011.

ZADEH, Lotfi A. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man and Cybernetics, v. SMC-3, n.1, January 1973.

ZHANG, Yun-tao; GONG, Ling; WANG, Yong-cheng. An improved TF-IDF approach for text classification. Journal of Zhejiang University Science, China, 2005.

ANEXO 1

QUESTIONÁRIO DE VALIDAÇÃO DE SISTEMA PARA RECUPERAÇÃO DE INFORMAÇÃO

Por favor, leia cada afirmação cuidadosamente e selecione qual opção se adequa melhor a sua opinião a respeito do sistema apresentado neste trabalho. Por favor indique qual grau de satisfação para cada afirmação circulando um número de 1 à 4, onde 1 significa "Não atende" e 4 significa "Atende completamente". Não há respostas certas ou erradas, por isso sinta-se tranquilo para responder cada pergunta da forma mais honesta possível. Certifique-se de responder todas as questões.

	Qual sua opnião a respeito das seguintes afirmações sobre o sistema?	**Não atende**	**Atende em partes**	**Atende**	**Atende completamente**
1.	Efetua o registro de documentos/artigos científicos e indexa estes para posterior recuperação.	1	2	3	4
2.	Recupera informações de documentos/artigos previamente registrados no sistema.	1	2	3	4
3.	O sistema traz resultados multimídias relacionados a busca do documento/artigo e seus resultados.	1	2	3	4
4.	O sistema traz resultados relevantes aos termos de busca.	1	2	3	4
5.	Tem desempenho satisfatório quanto ao tempo de busca.	1	2	3	4
6.	O sistema tem interface amigável, ou seja, é fácil de manuseá-lo.	1	2	3	4
7.	Apresenta uma forma interessante de exibir resultados.	1	2	3	4
8.	A solução apresentada neste sistema facilita encontrar informações pertinentes aos termos de busca.	1	2	3	4
9.	Permite criar uma base de dados de documentos/artigos de fácil manutenção.	1	2	3	4
10	Em uma pesquisa simples no sistema, geralmente as informações desejadas são encontradas.	1	2	3	4