

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

UTILIZAÇÃO DE INTELIGÊNCIA COLETIVA PARA A SEPARAÇÃO DOS
METADADOS DE REFERÊNCIAS BIBLIOGRÁFICAS

GUSTAVO BORGES FOLLADOR
MARLON AUGUSTO MIRETZKI

CURITIBA

2018

GUSTAVO BORGES FOLLADOR

MARLON AUGUSTO MIRETZKI

UTILIZAÇÃO DE INTELIGÊNCIA COLETIVA PARA A SEPARAÇÃO DOS
METADADOS DE REFERÊNCIAS BIBLIOGRÁFICAS

Trabalho apresentado para disciplina de Trabalho de Conclusão de Curso do Bacharelado em Sistemas de Informação do Departamento Acadêmico de Informática - DAINF - da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Alexandre Reis Graeml

CURITIBA

2018



TERMO DE APROVAÇÃO

“Utilização de Inteligência Coletiva para Separação dos Metadados de Referências Bibliográficas”

por

“Marlon Augusto Miretzki e Gustavo Borges Follador”

Este Trabalho de Conclusão de Curso foi apresentado como requisito parcial à obtenção do grau de Bacharel em Sistemas de Informação na Universidade Tecnológica Federal do Paraná - UTFPR - Câmpus Curitiba. O(a)(s) aluno(a)(s) foi(ram) arguido(a)(s) pelos membros da Banca de Avaliação abaixo assinados. Após deliberação a Banca de Avaliação considerou o trabalho _____.

<hr/> <Prof. Alexandre Reis Graemi> (Presidente - UTFPR/Curitiba)	<hr/> <Prof. Marcelo Mikosz Gonçalves> (Avaliador(a) 1 - <Instituição>)
<hr/> <Prof. Luiz Augusto Pelisson> (Avaliador 2(a) - Instituição)	<hr/> <Profa. Leyza Baldo Dorini> (Professora Responsável pelo TCC – UTFPR/Curitiba)
<hr/> <Prof. Leonelo Dell Anhol Almeida> (Coordenador do curso de Bacharelado em Sistemas de Informação – UTFPR/Curitiba)	

“A Folha de Aprovação assinada encontra-se na Coordenação do Curso.”

AGRADECIMENTOS

Gostaríamos de deixar registrado o nosso agradecimento ao Prof. Alexandre Reis Graeml, pela enorme paciência e pela sabedoria com que nos orientou ao longo dos anos no desenvolvimento desse trabalho.

RESUMO

FOLLADOR, Gustavo; MIRETZKI, Marlon. **Utilização de inteligência coletiva para a separação dos metadados de referências bibliográficas**. 2018. 63 f. Trabalho de Conclusão de Curso (Curso Superior de Tecnologia em Sistemas de informação), Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

Neste trabalho será abordado a problemática de separação de metadados de referências bibliográficas, e a utilização de inteligência coletiva como forma de solução, esta solução trata-se da distribuição de pequenas tarefas de classificação, o trabalho de pegar uma referência e separa-la em metadados (autor, título da publicação, ano, etc), quando realizado por poucas pessoas, é um trabalho demorado, mas ao distribuí-lo a uma grande multidão, o esforço é muito pequeno, para isso foi criado uma ferramenta que irá facilitar ainda mais o trabalho, fazendo um pré-processamento dos dados deixando o usuário apenas com uma pequena parte da separação e a verificação dos dados já separados.

Palavras-chave: inteligência coletiva. contribuição colaborativa. metadados. referência bibliográfica.

ABSTRACT

*FOLLADOR, Gustavo; MIRETZKI, Marlon. **The use of collective intelligence for the separation of metadata in bibliographical references.** 2018 63ff. Trabalho de Conclusão de Curso (Curso Superior de Tecnologia em Sistemas de informação), Departamento Acadêmico de Informática, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.*

This work will be approaching the problematic in separating metadata of bibliographical references, and the use of collective intelligence as a form of solution, this solution is about the distribution of small tasks of classification, separating a reference into metadata (author, title, year, etc), when done by few people, is a time consuming work, but if split among many, the effort is very small, for this was created a tool that will further facilitate the work, by doing a pre-processing of the data, leaving the user with only a small part of the separation and checking the data already separated.

Keywords: *collective intelligence. crowdsourcing. metadata. bibliographic reference.*

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de casos de uso.	22
Figura 2 – Diagrama de sequência das requisições.	23
Figura 3 – Diagrama de classes	24
Figura 4 – Artefato em execução.	26
Figura 7 – Sucesso no carregamento do artefato.	36
Figura 8 – Aplicação auxiliar	36
Figura 9 – Artefato e as funcionalidade.	37
Figura 10a e 10b – cenários ao executar.....	38
Figura 11 – erro de estimativa do tipo.....	39
Figura 12 – correção do erro de estimativa do tipo.....	40
Figuras 13a e 13b – erro de estimativa de metadado.....	40
Figura 14 – exemplo dos dados de saída do artefato.	41

LISTA DE TABELAS

Tabela 1 – abordagens computacionais (Lipinski et al., 2013)	14
Tabela 2 – Levantamento inicial da ordem dos metadados por estilo e documento.	30

LISTA DE ABREVIATURAS

ABNT - Associação Brasileira de Normas Técnicas

APA - American Psychological Association

ISO - International Organization for Standardization

SVM - Support Vector Machine

HMM - Hidden Markov Models

CRF - Conditional Random Fields

NBR - Norma Brasileira

TCC - Trabalho de Conclusão de Curso

URL - Uniform Resource Locator

DOI - Digital object identifier

RF - Requisitos Funcionais

HTML - Hyper Text Markup Language

CSS - Cascading Style Sheets

SUMÁRIO

1 INTRODUÇÃO.....	4
1.1 Contextualização do tema.....	4
1.2 Objetivo geral.....	4
1.3 Objetivos específicos.....	5
2 REFERENCIAL TEÓRICO	6
2.1 Inteligência coletiva.....	6
2.1.1 WEB 2.0	10
2.1.2 <i>Crowdsourcing</i>	11
2.1.3 Motivação para colaborar em projetos do tipo crowdourcing	13
2.2 O problema da extração de metadados	13
2.2.1 Abordagens computacionais	13
2.2.2 Problemas de agregação/coleção	15
3 Metodologia.....	16
3.1 Ferramenta inicial	18
3.2 Estrutura dos dados.....	18
3.3 Arquitetura	20
3.3.1 Requisitos.....	20
3.3.2 Diagramas	22
3.4 Tecnologias adotadas	24
3.5 Entrada e Saída.....	25
3.6 Interface com o colaborador.....	26
3.7 Motivação	28
3.8 Pré-processamento	29
3.9 Pós-processamento	33
4 ARTEFATO	34
4.1 Dependências	34
4.2 Instalação	35
4.3 Auxiliar	36
4.4 Execução.....	37
5 CONSIDERAÇÕES FINAIS.....	43
REFERÊNCIAS	45
APÊNDICE A – Grafos por tipo de documento.....	47

APÊNDICE B – Código fonte e arquivos necessários ao artefato.....	50
APÊNDICE C – Descrição das funções publicas do artefato.....	51
ANEXO A – Resumo das normas referenciais	52
ANEXO B – Abreviaturas de meses – ABNT NBR 6023.....	64

1 INTRODUÇÃO

1.1 Contextualização do tema

Existe um termo em inglês, “*wicked problems*”, que se refere a problemas para os quais ainda não existe uma solução computacional boa o suficiente (INTRONE *et al.*, 2013).

Para muitos destes problemas pode-se buscar uma solução alternativa utilizando recursos humanos, ou continuar a utilizar as soluções computacionais que existem até o momento, apesar dos desafios que isso impõe (INTRONE *et al.*, 2013). Para os que se utilizam de esforços humanos, a *web* é um bom lugar para se procurar, pois existe uma grande quantidade de usuários que podem desejar solucionar o problema em questão, ou que possam ser motivados, ou comprados, para trabalhar na solução do problema (ACAR *et al.*, 2011).

O problema de detecção de metadados em referencial bibliográfico cai nesta categoria, porque os autores de textos científicos, ao apresentarem sua lista de referências, não o fazem de forma rigorosamente padronizada. Além de existirem vários padrões (APA, ABNT, Chicago etc.), há variações que decorrem da ampla gama de tipos distintos de referências, que vão desde livros e periódicos científicos até *websites*, catálogos e manuais.

Inteligência coletiva é um dos termos utilizados quando se discute a possibilidade de se obter um resultado que possa beneficiar a todos, a partir de interações de diversas pessoas, direta ou indiretamente (BONABEAU, 2009). Buscase, por meio dessas interações, ainda que de forma inconsciente em alguns casos, potencializar os resultados que poderiam ser obtidos por um único indivíduo, ao tentar solucionar um problema. Afinal, o resultado gerado pela cooperação entre dois ou mais membros de um grupo raramente é apenas igual à soma do potencial individual de cada um deles (BONABEAU, 2009).

1.2 Objetivo geral

O objetivo desse trabalho é estudar a literatura referente a inteligência coletiva, e desenvolver um artefato computacional que possibilite a coordenação de esforços de um grande número de pessoas.

1.3 Objetivos específicos

Para atingir o objetivo geral exposto acima será necessário o cumprimento de cinco objetivos específicos, todos eles necessários individualmente e suficientes em conjunto:

- Compreender o que é inteligência coletiva a partir da perspectiva dos diversos autores que tratam do assunto na literatura;
- Identificar os problemas que podem ocorrer ao se utilizar estratégias puramente computacionais para resolver o problema de extração de referências e separação de metadados;
- Identificar perspectivas de inteligência coletiva que mais se ajustem à coordenação de pequenos esforços individuais de muitas pessoas para a obtenção de resultados grandes, que possam ser adotadas para subsidiar o presente trabalho;
- Verificar os requisitos para a criação de um artefato que facilite e transforme o processo de separação de metadados em algo simples e ágil;
- Desenvolver o artefato computacional;

2 REFERENCIAL TEÓRICO

Nesse tópico são apresentadas considerações encontradas na literatura sobre os fundamentos da inteligência coletiva e as mudanças que levaram a sua concepção atual, assim como considerações sobre a problema apresentado.

2.1 Inteligência coletiva

Lévy (1994, p. 28) conceitua a inteligência coletiva como:

uma inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências.

Verifica-se que a base fundamental da inteligência coletiva é o reconhecimento e enriquecimento mútuo das pessoas, um laço social construtivo e cooperativo em que cada indivíduo, mesmo não sabendo tudo, pode colaborar com aquilo que sabe (ANDRADE *et al.*, 2011).

Já para Noubel (2004, p. 8), a inteligência coletiva:

é o que molda as organizações sociais [...] onde indivíduos se reúnem para compartilhar e colaborar, para encontrar uma vantagem individual e coletiva que é maior do que se cada participante tivesse permanecido sozinho.

Não sendo somente uma característica da espécie humana, a inteligência coletiva está presente também em muitas comunidades animais. Porém, na complexidade das sociedades humanas, a inteligência coletiva sofreu mudanças ao longo do tempo, e diferentes formas de inteligência coletiva coexistem hoje, se diferenciando pela maneira como distribuem, relacionam e coordenam o conhecimento (NOUBEL, 2004).

A inteligência coletiva, na sua forma original, foi introduzida pela própria evolução, e está presente em uma matilha de lobos caçando uma presa ou em um time esportivo (NOUBEL, 2004). É a inteligência formada por um pequeno grupo de indivíduos com proximidade espacial. Mas a quantidade de participantes e proximidade entre eles são limitadores desse tipo de inteligência coletiva. O número elevado de participantes ou dispersão espacial gera mais ruídos que resultados efetivos, afetando diretamente a eficiência do grupo (NOUBEL, 2004).

Pode-se também tratar de uma nova dimensão da comunicação, que possibilita compartilhar e direcionar o conhecimento e que requer uma nova linguagem

ou processo de comunicação em que o tratamento da informação possa ser distribuído e coordenado por toda parte (ANDRADE *et al.*, 2011).

No começo da história humana, caracterizada pelo advento da escrita, os limites impostos pela inteligência coletiva original, quantidade de integrantes e distância entre eles, foram superados, uma vez que o conhecimento a ser compartilhado passou a poder ser escrito, não estando mais fixo em seu emissor (NOUBEL, 2004).

O desempenho dos seres humanos em uma variedade de diferentes tipos de tarefas cognitivas foi encapsulado como um fator estatístico comum chamado *g* ou fator de inteligência geral. Essa inteligência, junto com o fator social, gera uma metodologia psicométrica para quantificar um fator denominado "inteligência coletiva" (*c*), que reflete o quão bem os grupos atuam em um conjunto similar de tarefas de resolução de problemas de grupo (WOOLLEY *et al.*, 2010). Os principais contribuintes para *c* parecem ser os fatores *g* dos membros do grupo, juntamente com uma propensão para a sensibilidade social (em essência, o quão bem os indivíduos trabalham com os outros). O que a inteligência realmente é não está claro e é debatido, mas existe uma associação reprodutível de *g* com resultados de desempenho, como renda e desempenho acadêmico. (WOOLLEY *et al.*, 2010).

Essas novas características deram origem à inteligência piramidal que, segundo Noubel (2004), está presente em todas as estruturas humanas, impérios, governos, até organizações religiosas. Baseiam-se em quatro princípios principais:

- 1) divisão da informação e do trabalho pré-definido para tornar as pessoas permutáveis.
- 2) existência de uma autoridade que se beneficie da assimetria na comunicação e seja responsável pela divisão do trabalho, do território e distribuição de riqueza.
- 3) existência de uma relação de necessidade monetária, que caracteriza hierarquias de inteligência por canais de lealdade.
- 4) existência de padrões e normas que possibilitem a circulação e interoperatividade dentro da comunidade (NOUBEL, 2004).

Esse tipo de inteligência é também limitado, principalmente pela incapacidade de se adaptar a mudanças, tanto pela divisão pré-definida do trabalho quanto pela hierarquia, o que acarreta um grande fluxo de informação para utilização pelas

minorias gerenciadoras, que não conseguem interpretá-las adequadamente, acarretando conflitos (NOUBEL, 2004). Outro ponto limitador é a concorrência, que pode inibir a participação, uma vez que, em uma sociedade igualitária, não há níveis na pirâmide. Ou seja, todos os indivíduos estando em um mesmo nível, necessidades monetárias são iguais ou inexistentes, fato que ocorre também com o papel de autoridade, rompendo a assimetria na comunicação. Por fim, os padrões e regras, na maioria das vezes seguem lógica de concorrência e são utilizados como forma de monopolizar o conhecimento, em vez de maximizar a permeabilidade e a interoperabilidade com o ambiente externo (NOUBEL, 2004).

Nas últimas décadas, comunidades de insetos vem sendo observadas por teóricos de inteligência coletiva (NOUBEL, 2004). Nessas comunidades, criaturas relativamente "estúpidas" conseguem criar uma simbiose inteligente, reativa e adaptável ao ambiente (NOUBEL, 2004). Diferentemente das inteligências coletivas descritas anteriormente, indivíduos dessas comunidades não têm conhecimento do que é a entidade emergente e, na falta de hierarquia de comando, são as variáveis externas, como temperatura, tempo, comida, ameaças, etc. Que definem como se comportam essas comunidades (NOUBEL, 2004). Foram necessários milhões de anos de evolução para introduzir, no código genético de cada indivíduo, meios para que um grande número de indivíduos trabalhasse em uníssono para criar sociedades estáveis e robustas (NOUBEL, 2004).

Esse tipo de inteligência coletiva, que é chamada por Noubel (2004) de "*swarm*" (inteligência de enxame), pode ser vislumbrado nas sociedades humanas no domínio econômico, onde um simples pagamento se assemelha à comunicação entre insetos. Com várias relações de transações um para um, um sistema complexo, com capacidade de resposta e propriedades adaptativas emergentes (NOUBEL, 2004).

O sucesso da inteligência de enxame é fundamentado no anonimato e na uniformidade dos indivíduos, que podem ser sacrificados em nome do equilíbrio global do sistema, o que é fato irrelevante para insetos, mas teoricamente não se aplica a comunidades humanas, pois agimos de forma independente. Porém, no sistema econômico atual, a abordagem liberal postula que o sistema deve atingir equilíbrio independente das perdas (NOUBEL, 2004). Baseado na tese de equilíbrio natural, Noubel (2004) caracteriza a inteligência de enxame como uma forma transitória de inteligência coletiva, que inclui características dos modelos anteriores e assume um novo nível de complexidade.

Já Leimeister (2010) afirma que a inteligência coletiva se transformou, ao longo do tempo, por um processo evolucionário, traçado em torno de grupos de indivíduos, não sendo somente uma característica única de sociedades humanas. Está presente em qualquer outra espécie que possua algum nível de coletividade.

Se separarmos os termos para achar uma definição, coletividade pode ser entendida como um grupo de indivíduos que não precisam ter as mesmas atitudes e pontos de vista. Já a inteligência pode ser definida como a habilidade de aprender, entender e se adaptar a um ambiente, utilizando o próprio conhecimento. Uma definição amplamente aceita, considera inteligência como a habilidade de agir intencionalmente, pensar sensatamente e lidar com o ambiente de forma eficaz (LEIMEISTER, 2010). Essas características da coletividade proporcionam diferentes perspectivas e abordagens, o que acarreta melhores explicações e soluções para um certo problema. Por outro lado, a definição de inteligência proporciona aos indivíduos lidarem com situações desafiadoras e instáveis (LEIMEISTER, 2010). Utilizando a combinação dessas definições, o Centro de Inteligência Coletiva do MIT define inteligência coletiva como um grupo de indivíduos que faz coisas coletivamente as quais parecem ser inteligentes (LEMEISTER, 2010).

A inteligência coletiva original, classificada por Noubel (2004), transcende e inclui o indivíduo, pela entidade emergente e pelo relacionamento harmonioso que promove sua evolução e dá significado de coletividade. Como a inteligência hierárquica e a inteligência de enxame são formas passageiras, mas necessárias para a evolução, não incluindo ou transcendendo o indivíduo, Noubel (2004) acredita que um novo modelo para a inteligência coletiva em escala global, inteligência coletiva global, é um modelo que estende o conceito da inteligência coletiva na sua forma original.

Porém, para que o modelo de inteligência coletiva original suporte, transcenda e inclua milhões de pessoas, segundo Noubel (2004), novas características devem ser adicionadas, para lidar com o novo nível de complexidade dessa comunidade. Assim como na inteligência piramidal, padrões e normas são indispensáveis. Para coesão, permeabilidade e interoperabilidade de grandes comunidades, um espaço artificial simbólico que a comunidade compartilha, o cyber-espço, se faz necessário, dado que as relações de proximidade espacial são anuladas. A utilização de um sistema de informação é fundamental tanto para o acesso ao cyber-espço, quanto para lidar com o nível de complexidade das variáveis que compõem essa nova

comunidade, que a mente humana não é capaz de otimizar e organizar. Por fim, é necessário desenvolvimento pessoal, pois desenvolvimento em larga escala não pode ocorrer sem uma transformação do indivíduo, que envolve questões existenciais e motivacionais (NOUBEL, 2004).

Sobre o ciberespaço, Lévy (1994) afirma que este desterritorializa os saberes e funciona como suporte ao desenvolvimento da inteligência coletiva, um ambiente virtual em que possa existir o “espaço do saber”, onde as relações humanas são baseadas na valorização dos sujeitos e de suas habilidades. O ciberespaço está em constante construção, para maximizar a sinergia entre os sujeitos, utilizando as diferentes tecnologias possíveis e efetivando o espaço do saber.

Em Malone *et al.* (2010), diferentemente de Lévy (1994), a inteligência coletiva é apresentada como uma maneira puramente prática, em que um grupo de indivíduos reunidos parece fazer algo inteligente. Assim, adotando uma forma também prática de definir os requisitos da inteligência coletiva, Malone *et al.* (2010) propõem as seguintes perguntas: O que e como está sendo realizado? E quem e porque está realizando a tarefa?

A utilização da inteligência coletiva junto à computação é uma área emergente, mas ela possui um grande benefício de poder se utilizar de pessoas que possuem o mesmo interesse, como comunidade ou até mesmo pessoas que possuem um problema em comum, mesmo que não se conheçam. Elas podem juntar forças e, por meio da inteligência coletiva, chegar a uma solução para um problema que não conseguiriam resolver sozinhas (LYKOURENTZOU, 2011).

2.1.1 WEB 2.0

A agência de publicidade O'Reilly começou a utilizar o termo web 2.0 em meados de 2004 para definir a maneira como a Internet havia se definido e como passou a ser utilizada por empresas que sobreviveram à explosão da bolha "pontocom", em outubro de 2001 (O'REILLY, 2005).

Dentre as mudanças do modelo anterior (web 1.0) para esse novo modelo, a mais significativa era o foco no desenvolvimento de aplicativos, que se utilizam dos efeitos de rede e inteligência coletiva para se aperfeiçoarem, à medida que vão sendo utilizados (O'REILLY, 2005).

É uma plataforma de rede onde os usuários contribuem para o desenvolvimento de ferramentas, conteúdo e comunidades na Internet (SHANG *et al.*, 2011). A mudança do foco para a Internet como plataforma permitiu que softwares se integrassem e utilizassem as vantagens da conectividade, como a reutilização de dados provenientes de fontes múltiplas (ANDRADE *et al.*, 2011). Outro ponto importante da nova abordagem era o nível alto de produção e edição de conteúdo dos *sites* por parte dos usuários, que passaram de simples expectadores para produtores de conhecimento e opiniões (O'REILLY, 2005). Houve uma mudança das atividades estáticas da Internet para um novo paradigma dinâmico e uma forma interativa de construção do conhecimento, em que os usuários podem extrair informações e contribuir com elas simultaneamente (SHANG *et al.*, 2011).

O alto grau de volatilidade do ambiente também passou a ser algo a se levar em consideração, ou seja, nesse novo ambiente, os softwares devem ser mantidos em constante desenvolvimento, eterno "*beta test*" (O'REILLY, 2005).

2.1.2 Crowdsourcing

O termo em inglês *crowdsourcing*, formado pela junção das palavras multidão "*crowd*" e terceirização "*sourcing*", foi utilizado pela primeira vez em 2006 (HOWE, 2006) e vem sendo utilizado desde então na literatura para definir o modelo que deriva da inteligência coletiva, que se utiliza da Internet, para alcançar objetivos em geral empresariais, que não seriam alcançados por indivíduos em separado.

Howe (2006) define *crowdsourcing* como o ato de uma organização selecionar uma função executada por empregados e terceirizá-la para uma rede indefinida (geralmente grande) de pessoas, na forma de um convite aberto (*open call*). Essa tarefa pode assumir a forma de produção colaborativa, mas também é frequentemente praticada por indivíduos em separado, sendo pré-requisito fundamental o convite aberto e a existência de uma rede de possíveis colaboradores (HOWE, 2006).

Ainda para Howe (2008), indivíduos podem se organizar pela Internet para partilhar conhecimento, recursos e opiniões, dispensando as estruturas organizacionais comuns, para trabalhar como uma comunidade, em busca de interesses semelhantes. É importante salientar que, em vista de limitações de recursos e conhecimento da organização, o indivíduo mais qualificado para uma tarefa nem sempre é o que está disponível para executá-la (HOWE, 2008), principalmente

quando se adota o modelo tradicional de envolvimento de recursos humanos que correspondem aos funcionários de empresa.

Brabham (2008) por sua vez define *crowdsourcing* como, simplesmente, um modelo de produção e um sistema distribuído *on-line* de resolução de problemas, que vem mudando a maneira como muitas organizações solucionam problemas complexos.

Levando em consideração trabalhos desenvolvidos utilizando essa temática e estudando soluções que empregam esse conceito, pode-se entender *crowdsourcing* como o ato de terceirizar tarefas que originalmente precisariam ser realizadas pela organização, internamente a ela, ou por meio de relações de negócios, para um grande número indefinido e heterogêneo de potenciais atores, por meio de convites abertos na Internet, de forma livre de direitos autorais, que pode se basear em incentivos monetários ou não-monetários (HAMMON e HIPPIER, 2012).

Em sua essência, o *crowdsourcing* tem dois tipos de agentes: o *crowdsourcer* e *crowdsourcee*. O primeiro, é responsável pela chamada de abertura do projeto, frequentemente chamado de organização, podendo ser um indivíduo isolado, um grupo de indivíduos, ou uma organização (HOWE, 2006). Quando o objetivo do projeto é comercial, o termo “*company*” é comumente utilizado (HOWE, 2006). Em contrapartida, *crowdsourcee* é o termo utilizado para se referir ao conjunto de atores, responsável por encontrar a solução para o problema de rede indefinida e desconhecida, ou seja, a multidão que é (HOWE, 2006).

A principal característica do *crowdsourcing* é a utilização de multidões heterogêneas e de tamanho indefinido, que são responsáveis por alcançar um novo nível na criação de valor que as organizações, por si sós, não seriam capazes de alcançar (HAMMON e HIPPIER, 2012). Para Hammon e Hippner (2012), esta nova configuração de *sourcing* possibilita ultrapassar as restrições de qualidade e quantidade que seriam impostas pelos recursos limitados a que organizações tradicionais têm acesso, a partir do enorme potencial de conhecimento da multidão acessível pela rede.

Exemplos de projetos de *crowdsourcing* são a Wikipedia, onde pessoas contribuem com conhecimento de sua área (NOV *et al.*, 2010), em uma comunidade *online*, o reCAPTCHA da Google, que utiliza reconhecimento de letras a partir do escaneamento de livros (NOV *et al.*, 2010), o *Galaxy Zoo*, um projeto pelo qual os contribuidores classificam imagens de galáxias ou o *Citizen Weather Observer*

Program (CWOP) em que os voluntários monitoraram o meio ambiente e contribuem com dados meteorológicos (NOV *et al.*, 2010).

2.1.3 Motivação para colaborar em projetos do tipo *crowdsourcing*

De forma geral a literatura expõe dois tipos de motivações diferentes, a saber, intrínsecas e extrínsecas (ACAR *et al.*, 2011). A motivação intrínseca move o indivíduo pela realização da própria atividade, a necessidade do resultado que será obtido ou a simples motivação de ter a atividade realizada. Desta forma, realizar a atividade em um ambiente de *crowdsourcing*, ou simplesmente compartilhar o resultado não gera esforço significativo e, por ajudar uma comunidade no processo, o indivíduo acaba optando por contribuir (ACAR *et al.*, 2011). Já a motivação extrínseca é alavancada por meios externos, podendo ser este o reconhecimento de outras pessoas ou até mesmo alguma recompensa monetária (ACAR *et al.*, 2011).

Ambas as motivações são importantes, dadas as formas como afetam as pessoas, gerando satisfação pessoal e intelectual (VALLERAND, 1997). Isto pode ser amplamente explorado. Um indivíduo que precisa realizar uma atividade que pode ser feita por *crowdsourcing* pode fornecer estímulos às pessoas para que contribuam com seu projeto, não precisando, necessariamente, contratar mão de obra para realizá-la, mas, ao invés disso, motivando uma grande multidão para realizar pequenas partes, a fim de atingir seu objetivo (ACAR *et al.*, 2011).

2.2 O problema da extração de metadados

O problema de extração de metadados possui duas frentes, sendo a primeira o fato de existirem vários padrões para referência (APA, ABNT, Chicago etc.) além de variações dentro desses padrões, e a segunda é que, ao escrever sua lista de referências, muitos autores não o fazem de forma rigorosamente padronizada, gerando assim complexidade na abordagem de extração.

2.2.1 Abordagens computacionais

Existem métodos e ferramentas específicas para extração de metadados de artigos. O grande problema encontrado, muitas vezes, é que se deve prover a estes

programas o arquivo em um formato específico (PENG, MCCALLUM 2006), como por exemplo em LaTeX, onde as informações são extraídas pelas tags “<Metadata>”, “<Title>” e outros *tags* específicos.

Métodos fundamentais utilizados para a extração de metadados são análise estilística, aprendizado de máquina (*machine learning*) e utilização de bases de conhecimento (LIPINSKI *et al.*, 2013).

Ferramentas de extração de metadados usando a análise estilística extraem títulos usando heurísticas como: tamanhos e estilo de fonte e informações de posição de elementos examinados (LIPINSKI *et al.*, 2013).

O aprendizado de máquina em extração de metadados se utiliza de algumas técnicas: *support vector machines* (SVM), *hidden Markov models* (HMM), e *conditional random fields* (CRF). Estas abordagens se baseiam em formatação prévia e processamento de linguagem natural (LIPINSKI *et al.*, 2013).

A abordagem da base de conhecimento faz uso de banco de dados, como o *Google Scholar*, por exemplo, listas de nomes comuns, para atuar como uma referência às entidades extraídas (LIPINSKI, 2013).

Abaixo está uma análise feita por Lipinski *et al.* (2013) de softwares para extração de metadados e a técnica de extração adotada.

Nome da ferramenta	Link	Abordagem usada
Docear's PDF Inspector	http://docear.org	Análise estilística
GROBID	https://github.com/kermitt2/grobid	CRF
Mendeley Desktop	http://www.mendeley.com/	SVM
ParsCit	http://aye.comp.nus.edu.sg/parsCit/	CRF
PDFMeat	http://code.google.com/p/pdfmeat/	Base de conhecimento: Google Scholar
PDFSSA4MET	http://code.google.com/p/pdfssa4met/	Estrutura/análise de sintaxe do XML
SciPlore Xtract	http://sciplore.org/	Estrutura/análise de estilo do XML

Tabela 1 – abordagens computacionais (Lipinski *et al.*, 2013)

2.2.2 Problemas de agregação/coleção

As referências contidas em artigos acadêmicos possuem uma pequena quantidade de informação e conhecimento. No caso de vários artigos contidos em um repositório, tem-se uma grande quantidade de informação isolada com difícil acesso. Mas ao se agregar essas informações de forma ordenada é possível extrair uma grande quantidade de conhecimento (CHEN e PLOTT, 2002).

3 Metodologia

Para o desenvolvimento do trabalho proposto for necessária uma revisão da literatura quanto ao tópico de inteligência coletiva para compreender o que a motiva e como obter a adesão de uma “multidão” para a realização de uma tarefa, de forma voluntária, que, de outra forma, exigiria a contratação de muitos funcionários. Além, disso, prevê-se o desenvolvimento de um artefato com a utilização da tecnologia Web para um problema que não possui uma solução computacional ótima.

Para a realização do trabalho, foram definidos blocos de atividades que precisam ser realizadas, que estão descritos a seguir:

- Estudo e compreensão do conceito de inteligência coletiva: A compreensão do conceito e aplicações da inteligência coletiva for a base para todo o trabalho, pois é imprescindível uma profunda compreensão do tema para identificar o ponto motivador da “multidão” e a melhor maneira possível de aplicar os esforços de cada indivíduo.
- Análise e compreensão da base de periódicos de onde seriam obtidas as referências e para onde seriam devolvidas depois de tratadas e extraídos os metadados. Também analisar alguns periódicos para entender o que deve ser feito, e como deve ser feito.
- Planejamento e análise do projeto: Nesta fase foi definida a estrutura do projeto, cronograma, escopo do projeto, assim como levantamento de requisitos, modelagem, UML e demais atividades de engenharia de software.
- Desenvolvimento: esta fase se referiu ao desenvolvimento do sistema.
- Avaliação do resultado.

Inicialmente o sistema foi desenvolvido para ser aplicado em conjunto com a base de periódicos considerando os elementos da inteligência coletiva. Isto é, ao acessar qualquer recurso da base de periódicos, o usuário que nesse contexto passa a ser um colaborador, é apresentado ao problema da separação dos metadados de referência, e cabe a esse, resolver parte do problema. Para isso, é composto de duas aplicações que atuam em ambientes disjuntos mas funcionariam em conjunto para a solução do problema de forma colaborativa. O primeiro, aqui chamado de “servidor de controle”, opera no ambiente da base de periódicos e é responsável tanto pela entrega do problema a ser resolvido, para os colaboradores, quanto pela agregação e

tratamento das respostas decorrentes. Porém devido a atrasos no desenvolvimento desse trabalho, foi desconsiderado a implementação desse controlador. Limitando-se ao desenvolvimento de um artefato que, de forma objetiva, é entregue ao colaborador para que esse, alcance a resolução do problema, utilizando, ou não, os conceitos da inteligência coletiva, ou seja, o artefato nesse contexto passou a ser apenas a apresentação do problema para os colaboradores, desconsiderando as operações sobre o servidor de controle. Mesmo que essas operações e suas implicações sejam abordadas superficialmente nesse tópico, elas não foram efetivamente implementadas.

Assim o artefato objeto foi desenvolvido para a base, mas passou a ser considerado como um processador de informações independente, a ser aplicado aos colaboradores. Isto é, independente da fonte da entrada ou da saída, o artefato recebe referências em forma de texto e, após a sua execução, produz uma resposta ordenada, contendo todas as informações relevantes a respeito das referências. Mais adiante, são fornecidas mais informações referentes à entrada e à saída e às interações com o servidor de controle.

Essa independência do artefato permite que ele funcione local ou remotamente, se apoiado por um servidor, o qual, por sua vez, pode diferir do servidor do *website* que disparou esse artefato.

Dado que o artefato foi orientado para a base, uma aplicação existente no *site* da própria base serviu de guia para o desenvolvimento. Essa ferramenta, foi fundamental para a identificação da classe de dados utilizada pela base. Esses dados se fazem necessários para a comunicação com a base.

Vale destacar, que esse trabalho considera que os recursos da base de periódicos são acessados por meio de um *website*, então o artefato se utiliza da mesma tecnologia.

Também, como uma forma de maximizar o seu uso, o artefato, foi desenvolvido para permitir a edição dos dados de configuração, isso é, os tipos de referências, os campos das referências e os separadores de metadados das referências. Esse recurso se faz necessário, para adequação de possíveis alterações que a base tenha feito, e que não tenham sido contempladas nesse artefato, ou possíveis alterações futuras. Além disso, esse recurso possibilita que o artefato seja configurado para funcionar, para o mesmo problema, em situações em que haja outros quadros de dados, como exemplo, uma outra base de periódicos.

Essa seção ainda contém informações sobre a arquitetura utilizada, sobre a plataforma de implementação e a disponibilidade, sobre considerações acerca da interface com o usuário e sobre o pré e pós-processamento, que descrevem a implementação de técnicas para aliviar o esforço requerido para a execução da tarefa.

Vale destacar, que nesse contexto, quando referido o termo “base” é para citar a base de dados de periódicos a que se destina esse artefato. Já para os seguintes subtópicos se faz necessário esclarecer que o termo, “tipo de referência” é utilizado para se referir à composição de elementos utilizados pela base, para se descrever um tipo de documento, independente dos estilos. Simultaneamente, o termo “campo da referência” é empregado para se referir aos elementos usados pela base para descrever um tipo de documento, também independente do estilo, e o termo “resposta” é utilizado para se referir aos metadados separados e classificados de dada referência.

3.1 Ferramenta inicial

Esta ferramenta permite que os usuários insiram uma determinada referência que está contida em um documento, juntamente com a indicação de qual tipo de documento essa referência descreve. Após essa indicação, são apresentados um formulário onde os campos textuais são os elementos descritivos relevantes para o cadastro daquela referência. O papel do usuário, a partir desse ponto, é transcrever pela digitação ou cópia, os metadados nos seus referidos campos.

Uma cópia automática dos metadados nos campos do formulário é visível, principalmente nos campos numéricos ou nos campos com abreviações bem definidas, mas o grau de eficiência dessa funcionalidade é impossível de se determinar.

3.2 Estrutura dos dados

Na aplicação inicial, há um quadro de dados bem definido com relação às informações importantes para o problema. São visíveis, nesse quadro, os tipos de documento que a base considera relevantes para a classificação, e quais as informações que devem ser cadastradas para cada tipo de documento.

Esse quadro de dados foi considerado como a estrutura preliminar do artefato. Dispostos abaixo, foram organizados pelo tipo do documento e as informações necessárias para sua referência.

- Livro: Autor(es), Título do Documento, Ano, Edição, Cidade, Editora, URL, DOI.
- Capítulo: Autor(es), Título do Documento, Nome Livro, Editor, Ano, Cidade, Editora, URL, DOI.
- Anais: Autor(es), Título do Documento, Nome do Evento, Editor, Edição, Ano, Páginas, Cidade, Editora, URL, DOI.
- Matéria de Jornal: Autor(es), Título do Documento, Nome do Jornal, Ano, Páginas, Cidade, Editora, URL, DOI, Mês:
- Matéria de Revista: Autor(es), Título do Documento, Nome da Revista, Ano, Volume, Número, Páginas, URL, DOI.
- Caso: Autor(es), Título do Documento, Nome do Livro, Editor, Ano, Número, Páginas, Cidade, Editora, URL, DOI.
- Relatório: Autor(es), Título do Documento, Ano, Páginas, Cidade, Instituição, URL, DOI.
- Tese: Autor(es), Título do Documento, Ano, Páginas, Curso, Instituição, URL, DOI.
- Dissertação: Autor(es), Título do Documento, Ano, Páginas, Curso, Instituição, URL, DOI.
- Apostila: Autor(es), Título do Documento, Ano, Páginas, Curso, Instituição, URL, DOI.
- Monografia: Autor(es), Título do Documento, Ano, Páginas, Curso, Instituição, URL, DOI.
- Web: Autor(es), Título do Documento, Periódico, Ano, Páginas, Cidade, URL:
- Working Paper: Autor(es), Título do Documento, Ano, Volume, Número, Páginas, Instituição, URL, DOI.
- Genérico: Autor(es), Título do Documento, Ano, Páginas, Cidade, Editora, URL, DOI.

Porém, ao longo do desenvolvimento do artefato, alguns dos grupos de documentos se mostraram difíceis de diferenciar principalmente pelas similaridades nos seus metadados. Logo poderiam ser agrupados para minimizar a amplitude das alternativas da resolução, evitando conflitos na etapa do pré-processamento. Assim, monografias, teses e dissertações foram agrupadas em trabalhos acadêmicos. Adicionalmente, as informações necessárias para identificar as referências foram modificadas em alguns dos tipos de referência, conforme o apêndice A. Esse agrupamento deve ser tratado posteriormente, ao nível do servidor, para adequar a resposta obtida à segmentação dos dados da base.

Outra alternativa aplicada para facilitar a generalização da solução foi a adoção do termo “conjunto do documento” em referências de documentos que fazem parte de uma agregação, para se referir ao conjunto da obra em que está contido, como exemplos, se refere à revista, no caso de artigos de revista, ao livro, no caso de capítulos, ou ao evento, no caso de anais. Assim, o conjunto da obra em que o documento está contido passa a ser referenciado como título do conjunto e diferenciado pontualmente, simplesmente pelo tipo do documento indicado na referência. Em exemplo, o título do conjunto no caso de anais seria o título do evento, no caso de capítulos seriam o título do livro, etc. Essa distinção também se fez necessária no nível do servidor de controle para adequação à base de dados inicial.

Ressaltadas as ressalvas anteriores, o quadro de dados foi mantido semelhante ao quadro observado na ferramenta inicial. Os dados podem ser observados no apêndice A, na forma de grafos, onde cada grafo representa um tipo de documento, cada nó representa os tipos de metadados descritivos para a referência desse documento e as arestas representam os separadores entre os nós. Esses conceitos são apresentados no subtópico 3.8.

3.3 Arquitetura

3.3.1 Requisitos

Os requisitos foram definidos com relação a natureza do problema, a ferramenta inicial, a plataforma de implementação e as considerações adicionais contidas nesse tópico. Essa definição dos requisitos se fez necessária para a distinção

entre tarefas delegadas ao cliente de tarefas delegadas ao servidor, bem como a definição do pré e pós-processamento.

- Requisitos Funcionais

RF01: o artefato deve permitir que o usuário selecione o tipo de referência.

RF02: o artefato deve permitir que o usuário requisiite nova referência, sempre que não conseguir resolver a referência atual.

RF03: o artefato deve permitir a alteração da referência.

RF04: o artefato deve permitir que o usuário informe algum erro.

RF05: o artefato deve estipular a que o tipo de documento a referência corresponde.

RF06: o artefato deve estipular os elementos textuais relevantes para o tipo de referência do documento.

RF07: o artefato deve permitir que o usuário informe os metadados correspondentes a cada elemento textual.

RF08: o artefato deve estipular quais metadados correspondem aos elementos descritivos da referência, e ressaltá-los.

RF09: o artefato deve ressaltar de alguma forma a visualização dos metadados.

- Requisitos não funcionais

RN01: o artefato deve exibir sempre uma referência válida.

RN02: o artefato deve ter suporte multilíngue.

RN03: o artefato deve ser exibido sempre em destaque.

RN04: o artefato deve bloquear a navegação normal da página até que seja concluído.

RN05: o artefato deve ser de fácil integração com qualquer página web.

RN06: o artefato deve funcionar em qualquer navegador.

RN07: o artefato deve funcionar localmente.

RN08: o artefato deve minimizar o esforço do usuário.

RN09: o artefato deve funcionar com entradas locais ou externas.

RN10: o artefato deve ser executado no ambiente local do colaborador.

RN11: o artefato deve executar todas operações de processamento no ambiente local.

3.3.2 Diagramas

O diagrama de casos de uso abaixo foi utilizado para definir quais interações os colaboradores poderiam executar intermitentemente para chegar em uma resposta.



Figura 1 – Diagrama de casos de uso.

O diagrama de sequência abaixo demonstra as requisições entre o cliente e o servidor em ordem cronológica da ativação e finalização do artefato.

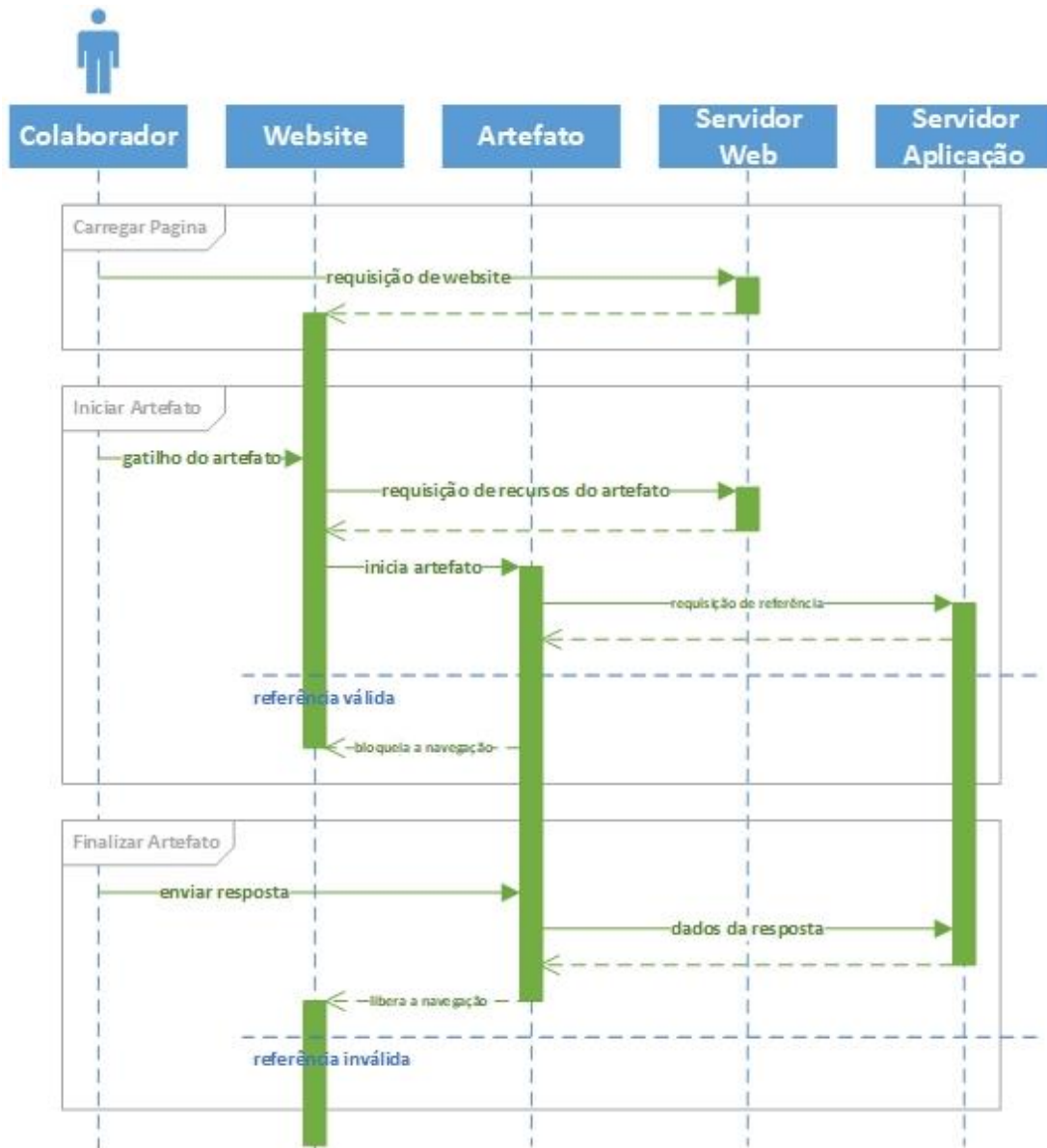


Figura 2 – Diagrama de seqüência das requisições.

O diagrama abaixo modela as classes que compõem o artefato, mesmo que na maior parte utilizando o paradigma de programação estruturada, o artefato é considerado uma biblioteca funcional, a ser inserida em outros códigos. Assim, a sua organização em classe ajuda a manter a organização e expor a suas dependências, no caso, a biblioteca de funções JQuery e as classes de dados estáticos.

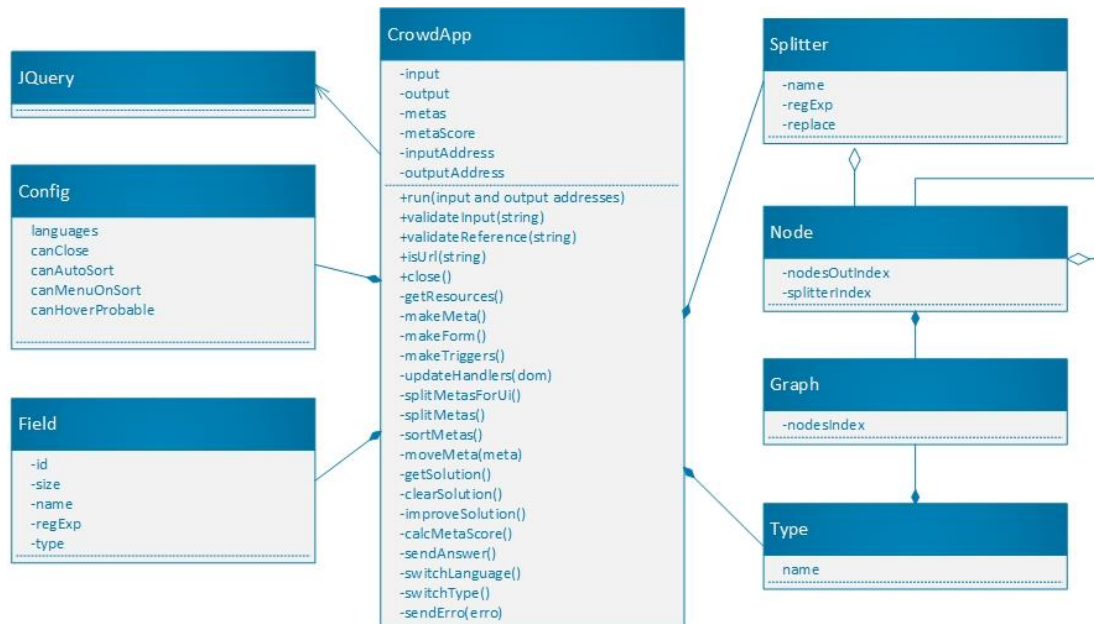


Figura 3 – Diagrama de classes.

3.4 Tecnologias adotadas

O padrão nos navegadores atuais, as tecnologias de HTML (*Hyper Text Markup Language*), *javascript* e CSS (*Cascading Style Sheets*) foram utilizadas para o desenvolvimento do artefato. Adicionalmente, a biblioteca funcional de códigos em *javascript*, *JQUERY*, foi empregada para facilitar a implementação do artefato e simplificar a compatibilidade com os diferentes navegadores. Essa escolha não só permitiu a compatibilidade com o *website* da base, mais também com qualquer outro *website* comum.

A tecnologia *javascript*, disponível nos navegadores, permite que o processamento ocorra do lado do cliente, concordando com uma das premissas do artefato. Outro ponto positivo do *javascript* é o suporte à comunicação com o servidor, que é imprescindível para a execução do artefato.

Já para o armazenamento do quadro de dados necessários para a execução do artefato, foi empregue o simples padrão JSON (*Java Script Object Notation*) utilizado para a troca de dados. Isso permite que os dados fiquem armazenados em outro arquivo, de forma compacta, independente da base e do código do artefato, e, por ser um padrão simples, a sua visualização e edição são facilitadas. Esse padrão, também é utilizado para o armazenamento e envio da resposta do artefato, o que possibilita que qualquer tecnologia seja adotada pela base, devido a sua alta compatibilidade com outras linguagens.

O problema encontrado na plataforma utilizada do padrão do *World Wide Web*, é que no ambiente dos navegadores, os documentos HTML possuem a possibilidade de alteração de qualquer dado recebido por parte do cliente, tanto por usuários experientes, pela edição de código, ou por usuários comuns, pelo uso, por exemplo, de aplicações bloqueadoras de código. Dessa forma é impossível garantir, sem uma autenticação efetiva com a unidade controladora, a execução do artefato de forma plena em quaisquer dos níveis, ou seja, não é garantida que toda a referência enviada pelo servidor irá retornar uma resposta válida e mesmo que toda resposta válida foi originária de alguma referência entregue pelo servidor.

3.5 Entrada e Saída

Inicialmente, foi considerada apenas uma entrada válida para o artefato, o texto contendo a referências. Porém, após analisar os possíveis usos do artefato, levando em consideração a sua proposta e com o intuito de facilitar o seu uso, as entradas válidas para o artefato passaram a ser três: um texto contendo as referências separadas por nova linha, ou um vetor com textos das referências, ou um endereço URL (*Uniform Resource Locator*) que responda uma requisição HTTP, com os dados das referências.

Aqui, considerou-se duas abordagens possíveis para o servidor, uma que responderia com uma referência aleatória e outra que responderia com referências do interesse do colaborador. A segunda, com intuito de aumentar a motivação para a colaboração, demanda um conhecimento prévio do colaborador, sendo por cadastro ou por análise de padrões de navegação.

Independente da fonte da entrada, o artefato considera uma referência válida, superficialmente, quando há quatro elementos ou metadados separados por ponto ou vírgula em uma mesma linha, sendo que as referências inválidas são desconsideradas. Essa verificação se faz necessária para evitar que colaboradores recebam outros tipos de dados que não condizem com o problema. Os elementos decorativos de fonte não são levados em consideração, como itálico ou negrito, pois esse tipo de informação necessita de dados adicionais para a sua indicação, e por serem elementos visuais não estão presente em todas as referências. Considerar esses elementos geraria conflitos, devido à incerteza de sua existência nos textos a serem analisados.

Considerando uma entrada válida, o artefato é executado para cada uma das referências válidas, ao menos uma, nos dados de entrada, e é finalizado quando todas as referências dos dados de entrada forem resolvidas ou ignoradas, produzindo como saída, um vetor contendo uma resposta para cada referência válida e confirmada. Essa resposta contém o texto da referência da entrada e todos os metadados com suas respectivas informações.

Ao nível do servidor, após o recebimento da resposta, se faz necessário confirmar, o formato válido da resposta e se a referência foi enviada pelo servidor.

3.6 Interface com o colaborador

A fim de manter o foco do colaborador no problema e possibilitar que ele contribua para a resolução do problema de forma eficiente, três elementos foram considerados indispensáveis. A apresentação do problema, ou seja, a referência a ser analisada sem alterações, o formulário da solução onde o colaborador transcreve os dados para resposta e as funcionalidades do artefato que auxiliam o colaborador na tarefa.

A imagem da Figura 4 demonstra que esses elementos essenciais são facilmente identificados.

The screenshot shows a web interface with a top navigation bar containing a dropdown menu for 'Matéria de Revista', a dropdown for 'Português', and buttons for 'Done', 'S', 'C', 'E', 'N', and '?'. Below this is a large text input field containing a reference: 'Andrade, I. A. ; Berti, D. W. ; Tomaél, M. I. , Corgosinho, R. J. M. (2011) . Inteligência coletiva e ferramentas WEB 2.0: a busca da gestão da informação e do conhecimento em organizações. Perspectivas em Gestão & Conhecimento, v. 2011, n. 1, p. 27-43, out, 2011.' Below the text field is a table of metadata fields:

Autor(es)	Andrade, I. A. ; Berti, D. W. ; Tomaél, M. I. , Corgosinho, R. J. M		
Título do documento	Inteligência coletiva e ferramentas WEB 2.0: a busca da gestão da informação e do conhecimento em organizações		
Nome da revista	Perspectivas em Gestão & Conhecimento		
Ano	2011	Edição	
Páginas	p. 27-43	Volume	v. 2011
URL			
DOI			
Número	n. 1		

Figura 4 – Artefato em execução.

Dentre todas as funcionalidades do artefato, a que demanda mais esforço do usuário é a indicação de quais informações do documento, estão presentes na referência. Mesmo que a apresentação do formulário com os elementos descritivos seja limitada aos elementos pertinentes para referenciar aquele documento, são no mínimo quatro elementos descritivos essenciais que precisam ser apresentados.

Com intuito de minimizar esse esforço, foram, inicialmente, considerados três padrões de interação, dados os limites da plataforma: a digitação, a seleção e o puxa e arrasta (*drag and drop*). Essas interações foram analisadas superficialmente, pelos movimentos necessários, tanto para a resolução do problema, quanto para correções de erros ou alterações.

Considerando-se que o usuário seja responsável por toda a identificação dos metadados, ou seja, sem nenhum pré-processamento, a reprodução dos metadados nos campos correspondentes, por meio da digitação, como ocorre parcialmente na ferramenta base, implicaria na transcrição da referência como um todo. Assim, o esforço seria máximo, além de ainda poderem ocorrer erros de digitação que poderiam comprometer a qualidade dos dados apresentados e os dados da solução. Por isso, esse tipo de interação, para essa funcionalidade, foi inicialmente descartado.

Outra alternativa seria a seleção. Nessa interação o usuário selecionaria o tipo de elemento descritivo e depois marcaria a parte do texto correspondente àquele metadado, ou vice-versa. Essa interação ocorre similarmente na ferramenta inicial, pela seleção, cópia e cola. Essa alternativa elimina os comandos de cópia e cola, onde o metadado em questão é automaticamente copiado para o campo selecionado do formulário. As correções finas ocorreriam após a transcrição. Apesar dessa alternativa reduzir o número de movimentos necessários, a seleção exata pode ser trabalhosa, principalmente em dispositivos móveis, somente, com suporte ao toque.

A última alternativa considerada foi a de arrasta e solta. Nessa interação, o usuário seleciona o metadado e, com o botão pressionado o arrasta, para o campo alvo do formulário. Liberando o clique, a transcrição então é feita automaticamente. Esse método, porém, necessita de um limitador do elemento a ser arrastado. A escolha desse limitador é um fator relevante para essa abordagem. Limitadores abrangentes podem unir metadados descritores diferentes, enquanto limitadores específicos podem aumentar o esforço, consideravelmente. No pior caso, o limitador seria o caractere, tornando essa alternativa inferior à digitação e, no melhor caso, o

limitador seriam metadados semelhantes, tornando essa a melhor solução. Consequentemente, sem o pré-processamento, essa alternativa não é melhor que a anterior, uma vez que a escolha desses limitadores é grande parte do problema principal e impossível de determinar sem uma análise.

Os limitadores do método anterior, ainda, poderiam ser aplicados, na alternativa de seleção, eliminando os problemas de seleção nativa. Porém, a seleção não seria mais intuitiva e o número de movimentos se igualaria à alternativa anterior.

Dada a análise apresentada, a interação implementada nessa funcionalidade foi a do arrasta e solta, uma vez que possui o menor esforço médio. Não há dificuldade adicional para usuários de dispositivos móveis com suporte ao toque. O processo de limitar os metadados é parte do escopo do artefato e as interações do tipo arrasta são habituais aos usuários por estar presentes nos sistemas operacionais. Contudo, se faz necessária, uma funcionalidade adicional, que permita alterar os limitadores, quando esses forem estipulados equivocadamente.

3.7 Motivação

Inicialmente, considerou-se que a participação na execução do artefato, ou *crowdsourcing*, seria forçada pelo bloqueio da navegação usual da página que disparou o artefato. Ou seja, o usuário seria obrigado a colaborar com o artefato, para ter acesso ao recurso que solicitou, pois, todos os elementos da página ficariam inacessíveis, limitando a navegação apenas ao artefato e ao retroceder da página.

Esse bloqueio seria liberado após a contribuição efetiva do usuário com o artefato, confirmada por meio da aprovação do envio da resposta. Porém, tal abordagem, que aborda a aspecto extrínseco da motivação, foi descartada por três motivos principais: (1) Tarefas de *crowdsourcing*, para obter resultados bons, precisam de usuários motivados e um usuário obrigado a fazer a tarefa acabaria se tornando menos motivado a contribuir, até em extremos casos buscando o mesmo recurso em outra fonte. (2) O segundo fator, em parte herdado do anterior e em parte da natureza do problema, é o fato que não foi encontrada uma verificação que considere a resposta errada ou correta com grau aceitável de confiabilidade, assim sendo, qualquer alternativa para considerar a contribuição efetiva, seria irrelevante, devido ao pré-processamento. Desta forma, o usuário contribuiria com qualquer informação apenas para ter acesso ao recurso. (3) E, por fim, é muito difícil de

assegurar o bloqueio da página devido à natureza da plataforma, e as tecnologias utilizadas, pois os dados são totalmente editáveis no cliente.

Então, optou-se por utilizar a participação facultativa/opcional, que demanda uma forte motivação intrínseca, pois, o artefato ainda é carregado sem o consentimento do usuário. Todavia, esse usuário pode finalizá-lo a qualquer momento e a navegação no *website* é retomada na forma usual mesmo quando há o envio da resposta e o artefato não verifica a validade dessa resposta.

3.8 Pré-processamento

Existem inúmeras abordagens computacionais para a resolução do escopo do problema, mesmo que essa abordagem difira das relatadas para resolução do problema, que passa a ser delegada ao colaborador. Algumas abordagens computacionais podem ser utilizadas para minimizar o esforço, transferindo o foco desse usuário para revisão e correção, ao invés da resolução do problema como um todo. Dessa forma, o tempo dispendido pelo usuário é mínimo, fazendo com que o colaborador esteja mais propenso a cooperar.

Um pré-processamento é visível na ferramenta base, visando à identificação dos metadados. Porém não é suficiente para transferir o foco do colaborador.

Para tal, a abordagem computacional deve ser utilizada em todas as tarefas necessárias para encontrar a solução. Assim, o papel do usuário passa a ser o de alterar o formulário onde achar apropriado, informar erros na referência ou no artefato e enviar a solução, assim que o resultado lhe for satisfatório.

Essas tarefas podem ser divididas na separação dos metadados descritivos da referência, na identificação do tipo de documento que a referência descreve e na identificação dos metadados descritivos. Sendo as duas últimas tratadas em conjunto, pois são interdependentes.

Tipos de Documentos	ABNT	ISO	APA	Vancouver
book	autor(es). título. edição. local: editora, ano. volume. paginação.	autor(es). título. edição. local: editora, ano. volume. paginação.	autor(es). (ano). título.(edição) local: editora. volume. paginação.	autor(es). título. edição. local: editora; ano. volume; paginação.
chapter	autor(es) do capítulo. título do capítulo. in: autor(es) do documento. título do documento. edição, local: editora, ano. volume. indicação do capítulo (páginas e/ou numeração).	autor(es) do capítulo. título do capítulo. in: autor(es) do documento. título do documento. edição. local: editora, ano. volume. indicação do capítulo (páginas e/ou numeração).	autor(es) do capítulo. (ano). título do capítulo. in autor(es) do documento, título do documento(edição, volume, indicação do capítulo). local: editora.	autor(es) do capítulo. título do capítulo. In: autor(es) do documento. título do documento. edição; local: editora; ano. volume; indicação do capítulo.
annals	autor(es). título. in: nome do evento, edição, ano, local. título do evento (anais, atas, etc.)... local: editora, ano, volume. paginação.	autor(es). título. in: TITULO DO EVENTO, edição, ano do evento, local do evento. local: editora, ano. volume. paginação.	autor(es). (ano). título. título do evento, edição, volume. paginação. local: editora.	autor(es). título. título do evento; ano de realização; local de realização. edição; local: editora; ano; volume; paginação.
newspaper	autor(es) da matéria. título da matéria. título do jornal, local, volume, mês ano, paginação.	autor(es) do matéria título da matéria. título do jornal. local. mês ano. volume. editora. paginação.	autor(es) da matéria (ano, mês). título do matéria. título do jornal, volume, (paginação).	autor(es) da matéria. título da matéria. título do jornal. mês ano; volume: páginação.
journal subject	autor(es) do artigo. título do artigo. título do periódico, edição, local: editora, volume, numeração, paginação, mês ano.	autor(es) do artigo. título do artigo. título da periódico. edição, local: editora, mês ano, volume(número), paginação.	autor(es) do artigo. (ano, mês). título do artigo. título do periódico, local: editora, edição, volume(número), páginação.	autor(es) do artigo. título do artigo. título do periódico. edição; local: editora; ano mês; volume(número);paginação.
case	autor(es) do caso. título do caso. título do conjunto, local: editora, volume, numeração, paginação, ano.	autor(es) do caso. título do caso. título do conjunto. local: editora, ano, volume(número), paginação.	autor(es) do caso. (ano). título do caso. título do conjunto, local: editora, volume(número), páginação.	autor(es) do caso. título do caso. título do conjunto. local: editora; ano ; volume(número);paginação.
report	autor(es). título. local: editora/instituição, ano. paginação.	autor(es). título.local: editora/instituição, ano.	autor(es). (ano). título. local: editora/instituição.	autor(es). título. local: editora; ano.
thesis dissertation monograph	autor(es). título. ano. paginação. curso – instituição.	autor(es). título. ano. (grau) - curso, instituição, ano.	autor(es). (ano). título (curso, instituição).	autor(es). título [grau]. instituição, curso; ano.
apostille	autor(es). título. local: curso. ano. paginação.	autor(es). título. local: curso, instituição. ano.	autor(es). (ano). título. curso, instituição, local.	autor(es). título. local: instituição, curso; ano.
web	autor(es). título. ano. url. ano de acesso.	autor(es). título. ano. url, ano de acesso.	autor(es).(ano). título. url. ano de acesso.	autor(es). título; ano. url. ano de acesso.
working paper	autor(es). título do artigo. título do conjunto, edição, local: editora, volume, numeração, paginação, ano.	autor(es) do artigo. título do artigo. título do conjunto, edição, local:editora, ano, volume(número), paginação.	autor(es). (ano). título do artigo. título do conjunto, (edição) local: editora, volume(número), páginação.	autor(es). título do artigo. título do conjunto. edição; local: editora; ano; volume(número);paginação.
generic	autor(es). título do artigo. título do periódico, edição, local: editora/instituição, volume, numeração, paginação, mês ano.	autor(es) do artigo. título do artigo. título do conjunto, edição, local: editora/instituição, mês ano, volume(número), paginação.	autor(es). (ano, mês). título do artigo. título do conjunto, (edição). editora/instituição, volume(número), páginação.	autor(es). título do artigo. título do conjunto. edição; local: editora; ano mês; volume(número);paginação.

Tabela 2 – Levantamento inicial da ordem dos metadados por estilo e documento.

Inicialmente, essas tarefas foram tratadas como uma só, onde uma expressão regular era utilizada para encontrar o tipo de documento da referência e extrair os metadados correspondentes. Porém, esse método foi descartado por não apresentar resultados satisfatórios e as expressões regulares utilizadas serem extensas e complexas. Apesar dessa abordagem ser descartada, foi observado que a expressão regular poderia ser quebrada, facilitando a compreensão, se a referência fosse tratada em frações, visto que as normas ou estilos definem limitadores semelhantes para separar os metadados descritivos.

A tarefa de separação desses metadados descritivos foi, primeiramente, tratada como uma simples subdivisão delimitada pelas pontuações referidas nas normas ou estilos. Contudo, essa divisão grosseira gerava muitas inconsistências e

dificultava a identificação dos metadados, por desconsiderar abreviações e separar sobrenomes de prenomes.

Vistos esses problemas, a abordagem adotada para essa tarefa, considera todos os possíveis limitadores entre os metadados dos elementos textuais, ou seja, sinais de pontuação, salvo em casos em que não representem uma separação entre os metadados. Como exemplo, o sinal de ponto “.”, na norma da ABNT deve ser considerado quando utilizado para separar o metadado correspondente ao título do metadado correspondente à edição, mas não quando utilizado para representar abreviações como em “ed.”.

- Os sinais de pontuação, considerados como separadores de metadados, com suas respectivas ressalvas são:
- Ponto: salvo em casos de abreviação.
- Vírgula: salvo em casos onde é utilizada para a separação do prenome e sobrenome de autor(es), quando é utilizada para a separação de múltiplos autores, ou quando faz parte do título.
- Dois pontos: salvo no caso em que são utilizados para separar o título do subtítulo do documento.
- Ponto-e-vírgula: salvo em casos em que é usado como separador, entre o nome de múltiplos autores.
- Parênteses: sem exceção.
- Colchetes: sem exceção.

Essa separação dos metadados é fundamental para efeitos visuais, ao apresentar os dados ao usuário e para a utilização na abordagem escolhida nas tarefas seguintes de identificação do tipo de documento que a referência descreve e de identificação dos metadados descritivos.

A abordagem adotada é de forma e posicionamento. Em princípio, os metadados, separados na tarefa antecedente, são analisados separadamente e em ordem para identificar, por meio de expressões regulares, quais os possíveis tipos de informações que representam. Essas informações, então, são comparadas aos grafos que representam o posicionamento dos metadados de cada tipo de documento referido.

O grafo é elaborado para cada tipo documento que a base queira identificar. Ele é definido pela agregação de todas as normas ou estilos, com relação à ordem em que os campos, ou metadados descritivos, aparecem. Os nós representam os campos e as arestas direcionadas, os limitadores.

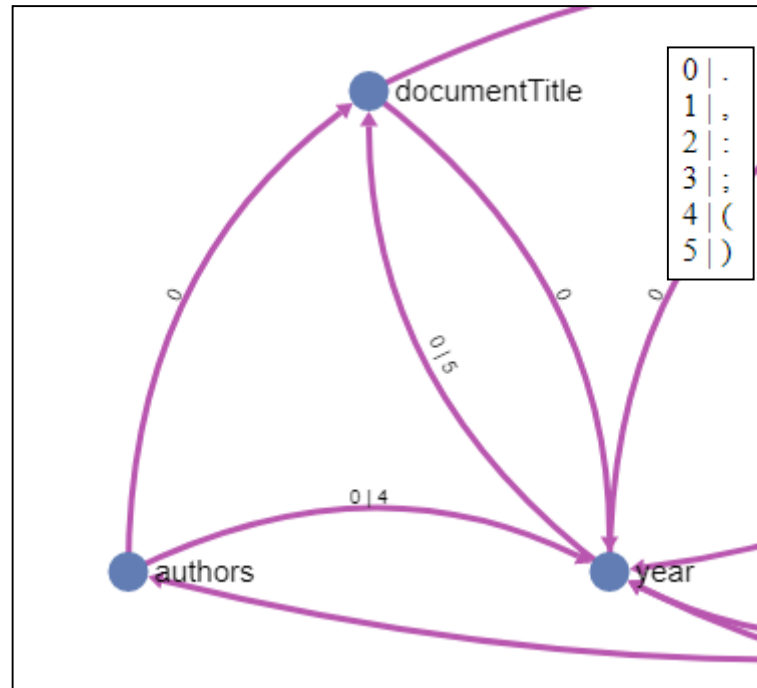


Figura 5 – Grafo exemplo.

O grafo incompleto acima evidencia que, após o metadado referente ao autor, os próximos metadados podem ser o título do documento ou o ano, se o limitador for o ponto, e apenas o ano se o limitador for “abre parênteses”.

Dessa forma os dados ponderados por expressões regulares, são comparados aos do grafo e, em caso de semelhança, o metadado é confirmado referente ao tipo de informação. Ou seja, o primeiro metadado separado é comparado ao primeiro nó do grafo, sendo que todo grafo tem um início e todo nó indica um campo da referência. Então, esse metadado é comparado por meio de expressões regulares para verificar se é compatível com o nó. Em caso de sucesso, o separador é analisado e todos os nós de saída do nó atual que possuem aresta com o valor do separador são considerados próximos nós a serem analisados. O próximo metadado é, então, verificado para ver se é compatível com os nós analisados e assim por diante.

Assim, a profundidade do caminho no grafo é referente ao número de metadados que já foram analisados, salvo em casos de previsão de falhas, onde nós que não são compatíveis são considerados verdadeiros.

Após o término dessa análise, o resultado é um caminho, em que o início se refere ao primeiro metadado. Cabe ao artefato percorrer esse caminho, transcrevendo os metadados para os referidos nós.

Em casos de ambiguidade, em que haja bifurcações no caminho, o caminho mais extenso é escolhido.

Vale lembrar que essa abordagem, baseada em uma aproximação grosseira do modelo de probabilidades de ocorrência, visa a incluir o maior número de referências possível. Esta abordagem foi adotada, pois considera os metadados de forma individual e em conjunto, assim auxiliando na classificação mesmo que o usuário altere a resposta.

3.9 Pós-processamento

O pós-processamento ao nível do cliente se faz necessário unicamente para a limpeza dos dados da resposta, uma vez que a abordagem adotada gere redundância de informações, como por exemplo: o metadado identificado como referente à edição, ainda contiver a sigla de edição junto com a informação relevante.

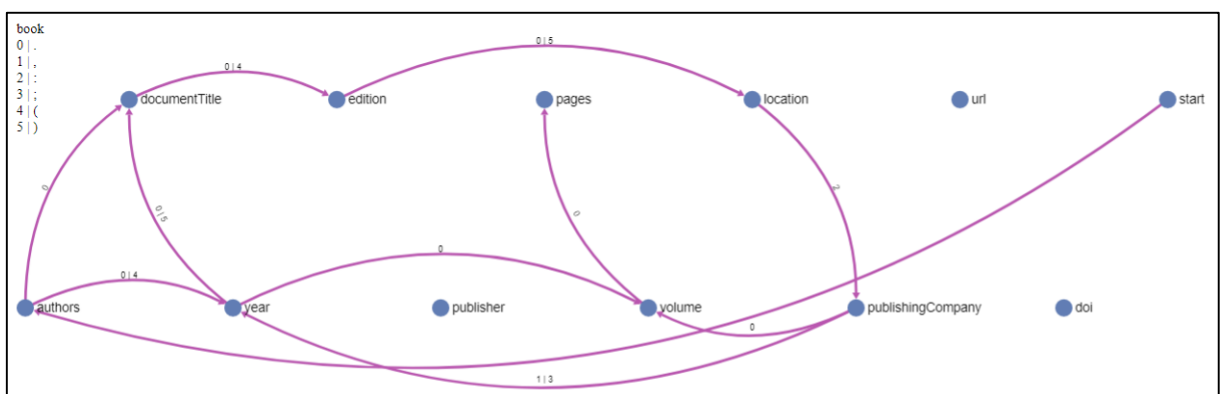


Figura 6 – Grafos de exemplo.

4 ARTEFATO

O resultado obtido foi um artefato que ao ser inserida no *website* da base, e inicializado através de algum gatilho, apresenta o problema juntamente com as funcionalidades necessárias para que o colaborador encontre uma solução.

Mesmo que inicialmente desenvolvido para ser executado em conjunto ao servidor da base de periódicos, uma aplicação auxiliar foi improvisada para o teste do artefato. Essa aplicação, então, passou a ser considerada parte do artefato uma vez que sem ela seria impossível testar a execução do mesmo.

Essa aplicação auxiliar não foi incluída nos capítulos anteriores devido ao seu grau de superficialidade, uma vez que apenas exemplifica a utilização do artefato em ambiente local através de uma página simples, ou seja, os dados de entrada e saída não passam a ser gerenciadas pelo servidor de controle. Cabe então, ao usuário, o papel que seria dessa unidade, isto é, fornecer os dados de entrada que passa a ser o texto contendo uma ou múltiplas referências, e ainda, deliberar o destino dos dados de saída, metadados com as informações classificadas pelo artefato.

Assim sendo, essa aplicação auxiliar é descrita a seguir, quase que completamente, na descrição da instalação, as funcionalidades adicionais presentes nessa aplicação são meramente visuais, unicamente para a captura e exibição das informações, com a finalidade de facilitar a compreensão do funcionamento do artefato principal.

Nesse tópico o artefato é titulado como “metaCrowdApp”, para fins de implementação, ao evitar conflitos com nomenclaturas comuns, que poderiam ocorrer devido a plataforma.

4.1 Dependências

Foram utilizados no desenvolvimento do artefato, bibliotecas populares de códigos para a plataforma. Utilizadas somente em parte, a fragmentação da parte necessária não foi agregada ao código para evitar redundâncias devido à ampla utilização dessas bibliotecas em *websites*, e mesmo previamente utilizadas pela base de periódicos que estava em questão.

Conseqüentemente essas bibliotecas, listadas abaixo juntamente com as versões empregadas, são dependências para o funcionamento correto do artefato.

- JQuery 3.1.0
- JQuery UI 1.12
- JQuery UI Touch Punch 0.2.3 (somente para o suporte ao *touch screen*)

4.2 Instalação

Com a proposta de facilitar a inserção do artefato em qualquer *website* independente da fonte origem da entrada ou destino da saída, e permitir a edição dos dados de configuração, foi adotada uma implementação que considera o carregamento dinâmico, como apresentados no tópico que arquitetura.

Essa implementação não é a melhor opção para uma distribuição do código do artefato, nem mesmo para a plataforma, mas foi utilizada e mantida devido a limitação do conhecimento dos autores com relação as tecnologias adotadas e para facilitar a compreensão.

Assim, ao invés de um formato agrupado e comprimido, a estrutura de arquivos que agrega todos os dados necessários ao artefato em um diretório, que deve ser respeitado para que o artefato funcione corretamente.

Para a execução do artefato é necessário que o código principal seja carregado na página em que irá ser apresentado, como no exemplo. Uma documentação breve, contendo todas as funções públicas do artefato está exposta no apêndice C.

```
<script type="text/javascript" rc="crowdMetaApp/script.js"></script>
```

O artefato não será inicializado imediatamente. Para tal, deverá ser evocado por meio da função “run”, dados os parâmetros necessários. Esses parâmetros que irão definir se o artefato será executado localmente ou de forma distribuída, vide apêndice C.

```
CrowdApp.run(
    "http://crowdmetaapp.atwebpages.com/metaCrowdApp/samples.json",
    "http://crowdmetaapp.atwebpages.com/metaCrowdApp/config.json",
    "http://crowdmetaapp.atwebpages.com/postSolution");
```

Em caso de sucesso a tela única do artefato deve ser exibida, e se assemelha a imagem abaixo.

Journal paper Not sure

LYKOURENTZOU, Ioanna et al. Collective intelligence systems: Classification and modeling. Journal of Emerging Technologies in Web Intelligence, v. 3, n. 3, p. 217-226, 2011.

Author(s)	LYKOURENTZOU, Ioanna et al		
Document title	Collective intelligence systems: Classification and modeling		
Journal name	Journal of Emerging Technologies in Web Intelligence		
Year	2011	Edition	
Publisher			
Pages	p. 217-226	Volume	v. 3
Place		Number	n. 3
Publishing company			
Institution			
URL			
DOI			
Month			

Figura 7 – Sucesso no carregamento do artefato.

Caso ocorra algum erro, o artefato não será inicializado, e não informará erro algum ao usuário.

4.3 Auxiliar

Como a unidade controladora não foi implementada e foi abordada de forma breve nesse trabalho, consideramos que o artefato, nesse exemplo, trabalha de forma local.

ARQUIVO ?

EXECUTAR TESTE

EXPORTAR GRAPH CONFIG 0

Figura 8 – Aplicação auxiliar

A aplicação auxiliar, conforme dito anteriormente, foi desenvolvida para auxiliar os testes no artefato em ambiente local. Para isso o colaborador, que nesse caso passa a ser apenas um usuário por não está contribuindo de forma coletiva para a

solução do problema, deve fornecer um arquivo em formato texto ou diretamente o texto que contenha uma ou mais referências separadas por caractere de nova linha. Após o comando “executar”, a aplicação auxiliar inicializará o artefato, passando os dados fornecidos pelo usuário como parâmetro, e esperará uma resposta.

As outras funcionalidades da aplicação auxiliar apenas apresentam os dados do artefato: “*graph*”, exibe o grafo dado ao índice do tipo de documento (Apêndice A), e “*config*” exibe o arquivo em formato JSON que contém as configurações que serão utilizados pelo artefato.

Após o usuário informar dados válidos (Apêndice C “*run*”), o artefato será inicializado aplicando a primeira referência informada.

4.4 Execução

Ao ser inicializado o artefato, bloqueará a navegação normal do *website* e apresentará ao colaborador uma interface semelhante à Figura 9.

Author(s)			
Document title			
Journal name			
Year		Edition	
Publisher			
Pages		Volume	
Place		Number	
Publishing company			
Institution			
URL			
DOI			
Month			

Figura 9 – Artefato e as funcionalidade.

Na figura acima estão numeradas as funcionalidades do artefato: (1) seleção do tipo de referencia, (2) confirmar e enviar a resposta, (3) indicador de incerteza na classificação da referência por parte do usuário, (4) solicitar uma nova referência, descartando a atual, (5) auto classificação da referencias pelas regras do artefato e modificações feitas em 10, (6) limpar o formulário 11, (7) explicação das funcionalidade do artefato, (8) fecha o artefato e descarta a solução, (9) mudar a linguagem de exibição do artefato, (10) apresenta a referência atual já separada nos possíveis metadados, (11) representa o formulário com as informações relevantes para o tipo de referência apontado por 1.

Há ainda duas funcionalidades de correção, que permitem o usuário recuperar o artefato em caso de erro: (r1) ao usar clique duplo esquerdo em caracteres separadores, o usuário é capaz de concertar erros na determinação dos separadores, agregando ou separando os metadados apresentados em (10). E (r2) que permite o usuário a editar o texto dos metadados para corrigir erros mais finos, como grafia.

Porem diferentemente da figura apresentada acima, o artefato, ao iniciar, tentará classificar os metadados separados em (10), com relação ao tipo de referência em (1) e aos campos em (11). Assim a duas alternativas possíveis.

The figure consists of two side-by-side screenshots of a web application interface. Both screenshots show a reference text at the top: "LYKOURENTZOU, Ioanna et al. Collective intelligence systems: Classification and modeling. Journal of Emerging Technologies in Web Intelligence, v.3, n.3, p.217-226, 2011." Below the text are several input fields for metadata. In the left screenshot (10a), these fields are populated with the data from the reference. In the right screenshot (10b), the same reference text is shown, but the metadata fields are empty. Below the reference text in the right screenshot is a section titled "WHAT KIND OF DOCUMENT DOES THE REFERENCE ABOVE DESCRIBE?" with a grid of buttons for different document types: Chapter, Proceedings, Newspaper, Journal paper, Case, Report, Academic work, Apostille, and Working paper. A note at the bottom of this section says "*IF NONE OF THE ABOVE, USE THE SELECT BOX IN THE TOP MENU".

Figura 10a e 10b – cenários ao executar.

Ocorre (10a), um cenário em que o artefato conseguiu identificar o tipo de referência, assim encontrando metadados suficientes para classificar a referência como pertencente ao um tipo de documento em específico, portanto, carregando o

formulário (11) com as informações pertinentes ao tipo da referência estipulada, e o preenchendo com as metadados sem ambiguidade.

Ou, um cenário (10b) em que o artefato não conseguiu identificar o tipo de documento que a referência descreve com exatidão, seja por ambiguidades ou erros de estipulação dos separadores dos metadados. Assim é necessário que o usuário indique o tipo de documento que a referência descreve antes do artefato prosseguir com a execução, implicando em (10a) mas passando a utilizar o tipo de documento indicado pelo usuário.

Ainda podem ocorrer cenários parecidos com (a), em que o tipo de documento da referência foi estipulado errado, e que seja necessária a correção do usuário por meio de (1).

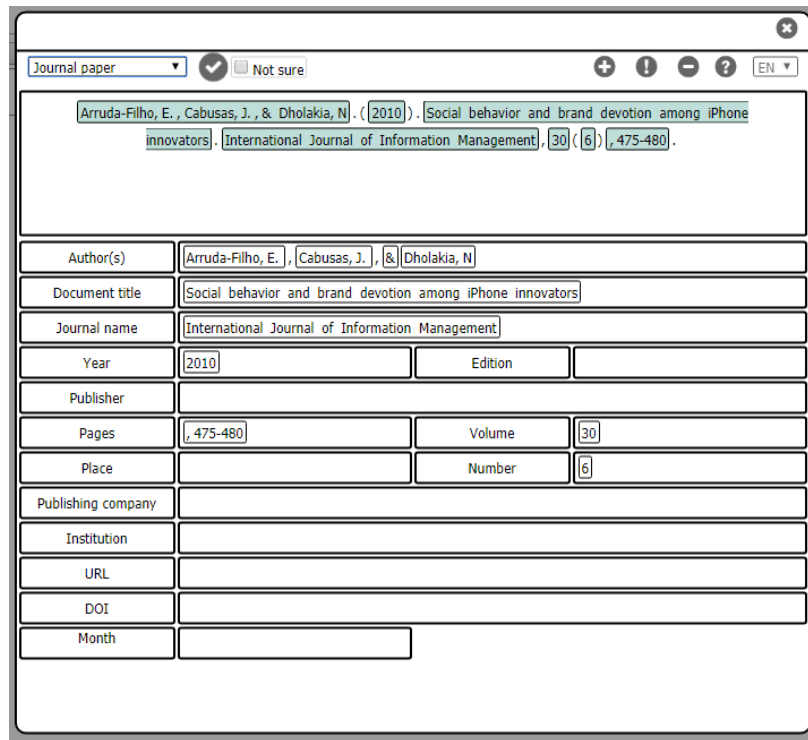
Ou, e não excludente as anteriores, que haja ambiguidades erros ao transferir os metadados em (10) para o formulário (11).

Case	<input checked="" type="checkbox"/> Not sure		
Arruda-Filho, E., Cabusas, J., & Dholakia, N. (2010). Social behavior and brand devotion among iPhone innovators. International Journal of Information Management, 30(6), 475-480.			
Author(s)	Arruda-Filho, E., Cabusas, J., & Dholakia, N		
Document title			
Journal name	Social behavior and brand devotion among iPhone innovators		
Publisher			
Year	2010	Pages	475-480
Volume	30	Place	
Publishing company			
Institution			
URL			
DOI			
Number	6		

Figura 11 – erro de estimativa do tipo.

Na figura acima o artefato erra em estipular o tipo de documento, um “Artigo de Periódico”, o indicando como um “Caso”. E erra também, por consequência da escolha prévia errada, o título do periódico e o título do artigo.

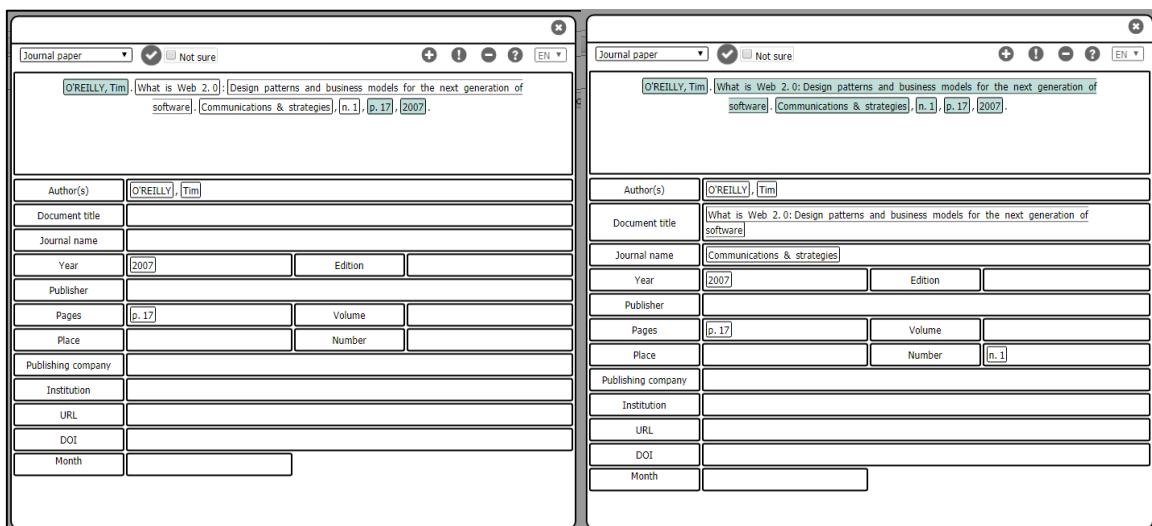
Todavia, ao corrigir o tipo de documento por meio de (1), o artefato se recupera e classifica corretamente a referência.



Journal paper	<input checked="" type="checkbox"/> Not sure	<input type="button" value="+"/>	<input type="button" value="!"/>	<input type="button" value="-"/>	<input type="button" value="?"/>	EN
Arruda-Filho, E., Cabusas, J., & Dholakia, N. (2010). Social behavior and brand devotion among iPhone innovators. International Journal of Information Management, 30(6), 475-480.						
Author(s)	Arruda-Filho, E., Cabusas, J., & Dholakia, N					
Document title	Social behavior and brand devotion among iPhone innovators					
Journal name	International Journal of Information Management					
Year	2010	Edition				
Publisher						
Pages	475-480	Volume	30			
Place		Number	6			
Publishing company						
Institution						
URL						
DOI						
Month						

Figura 12 – correção do erro de estimativa do tipo.

Existe ainda erros que ocorrem devido a dificuldades no estabelecimento dos separadores de metadados. Como ocorre na figura 13a, que através de (r1) e com a funcionalidade de (5), são corrigidos em 13b.



Journal paper	<input checked="" type="checkbox"/> Not sure	<input type="button" value="+"/>	<input type="button" value="!"/>	<input type="button" value="-"/>	<input type="button" value="?"/>	EN
O'REILLY, Tim What is Web 2.0: Design patterns and business models for the next generation of software Communications & strategies n. 1 p. 17 2007.						
Author(s)	O'REILLY, Tim					
Document title	What is Web 2.0: Design patterns and business models for the next generation of software					
Journal name	Communications & strategies					
Year	2007	Edition				
Publisher						
Pages	p. 17	Volume				
Place		Number	n. 1			
Publishing company						
Institution						
URL						
DOI						
Month						

Figuras 13a e 13b – erro de estimativa de metadado.

É facultativo o uso das ferramentas de correção, cabe ao usuário decidir se opta por corrigir a solução utilizando as ferramentas automáticas ou se prefere, o esforço de resolver de modo convencional, que na maioria dos casos implica em um esforço maior. Contudo, se após a correção dos erros, o artefato não for capaz de identificar, a referência ou os metadados, cabe ao usuário indicar as informações dos metadados de maneira usual pelo “*drag and drop*”.

É papel do usuário selecionar cada metadado não classificado em (10), e arrastá-lo para o campo correspondente no formulário, sucessivamente até que todos os dados sejam classificados, estado alcançado quando todos os dados de (10) obtiverem coloração.

A partir desse ponto, o usuário tem a responsabilidade de decidir se a resposta que foi formulada está correta, em caso de dúvida o usuário deve informar, em (3), que não tem certeza sobre a resposta obtida. E aqui entram dúvidas não somente em relação a resposta, mais também com relação a referência, ao artefato ou qualquer outra incerteza que comprometa a qualidade dos dados da resposta.

Na sequência, ocorre o envio da resposta a unidade controladora, que nesse caso, é o próprio usuário.

```
{
  "input": "Arruda-Filho, E., Cabusas, J., & Dholakia, N. (2010). Social behavior and brand devotion among iPhone innovators. Int
  "output": {
    "type": "case",
    "assurance": "false",
    "fields": {
      "authors": [
        "Arruda-Filho,E.",
        "Cabusas,J.",
        "&",
        "Dholakia,N"
      ],
      "documentTitle": "Social behavior and brand devotion among iPhone innovators",
      "documentSetTitle": "International Journal of Information Management",
      "publisher": "",
      "year": "2010",
      "pages": ",475-480",
      "volume": "30",
      "location": "",
      "publishingCompany": "",
      "institution": "",
      "url": "",
      "doi": "",
      "number": "6"
    }
  }
},
```

Figura 14 – exemplo dos dados de saída do artefato.

Com relação a classificação automática dos metadados, foi utilizado um tratamento simples, relativamente herdado da abordagem utilizada pela transcrição dos dados, onde a separação da referência é aplicada para minimizar o esforço do colaborador.

Assim, a divisão da referência sempre foi adotada como passo inicial do artefato, uma vez que, executada através de expressões regulares, é feita por funções de “substituição” e “separação”, visto que o uso de funções de substituição antes das funções de separação, facilitou a representação das expressões regulares para o problema.

Então, para cada metadado extraído é feita uma comparação com todos os possíveis campos que ele possa representar. O resultado dessa comparação então é relacionado os grafos que representam os tipos de referências.

A análise dessa relação ocorre por forma de um caminho no grafo dado o seu início e o início da referências. Assim o grafo que obtiver maior acerto com relação aos metadados passa a ser o provável tipo da referência. Esse grafo então, é apresentado no formulário e a transcrição automática para o formulário é efetuada com relação ao metadados.

Essa abordagem foi desenvolvida, pensando em tolerância a falhas, ou seja, que um nó, que representa o campo para descrever a referência, possa não estar presente na mesma, assim, o artefato poderia continuar a buscar por correspondências nos próximos nós.

Porem ao permitir a falha, um grande número de ambiguidades surge, assim metadados que possam ser correspondentes a mais de um campo, necessitam de tratamento. Dessa forma metadados com ambiguidades são resolvidos pela maior pontuação na etapa de ponderação.

Já com relação a essa abordagem, foi adotada unicamente dado a natureza do problema e da solução proposta, onde a resolução do problema iria ser aplicada aos colaboradores que de maneira distribuída e coletiva, contribuiriam para a sua resolução. Se considerarmos o artefato, devido a falta do controlador, como o único responsável pela a resolução do problema, os resultados são muito inferiores à aquelas abordagens apresentadas no tópico 2.1.1.

Porém o artefato desenvolvido é altamente configurável, e pode ser aplicado a qualquer base de periódicos ou análogo, e por reduzir o esforço necessário para a resolução do problema pode ser considerado uma alternativa viável para utilização em trabalhos semelhantes.

5 CONSIDERAÇÕES FINAIS

Conforme o desenvolvimento desta proposta, as questões teóricas de inteligência coletiva demonstraram-se cada vez mais compreensíveis. Ao se aprofundar no assunto por meio da revisão de literatura foi possível notar que inteligência coletiva, se utilizada da maneira correta, pode resolver não só problemas de origem computacional, mas também tem aplicação em diversas áreas, e em várias formas, quando um problema pode ser fragmentado em micro-tarefas e entregue ao indivíduo de forma simples, que possibilite uma resolução rápida.

Como vimos anteriormente, existem soluções puramente tecnológicas, que podem ser aplicadas para tentar solucionar o problema de agregação, mas ainda não chegamos a um ponto onde esta solução consiga resolver totalmente o problema, sem apresentar erros ou inconsistências. Dada variedade de formatos e a falta de padronização nas referências, este não pode ser resolvido de forma eficaz, por meios puramente tecnológicos. Mas, ao colocarmos a mente humana para fazer o trabalho da máquina, e também identificar divergências nos dados, muitos dos obstáculos que anteriormente dificultavam o trabalho são facilmente contornados. Nota-se que isso não exclui o trabalho da máquina, mas afunila a sua área de atuação em algo mais específico. Tendo isto em vista, e mediante a literatura e autores estudados, a perspectiva de inteligência coletiva adotada foi: utilização de pequenos esforços individuais para, manualmente, classificar metadados de referências a serem agregados em uma grande base de dados.

Deste modo, quando um determinado recurso do *website* do repositório de artigos fosse acessado, este enviaria uma solicitação para decodificar os dados de uma ou mais referências. E após realizado a separação dos metadados o *website* continuaria para acesso ao recurso. Com este processo em mente, uma ferramenta foi criada.

O artefato final, é uma ferramenta construída visando ao esforço coletivo, mesmo que este não tenha sido utilizado em um repositório empregando o esforço da multidão, ele facilita e agiliza todo o processo de separação dos metadados. Assim, o usuário que fazia manualmente o processo de decodificação, pode utilizar-se da ferramenta para que ela faça um pré-processamento nos dados, simplificando esse trabalho.

O projeto agora concluído, fica disponível para trabalhos futuros. Todo o código está disponibilizado na ferramenta GitHub, acessível em: <https://github.com/ht568cf/crowdMetaApp/>, e está aberto para qualquer tipo de utilização, mas, para continuação desta proposta, é necessário focar no segmento responsável pela inteligência coletiva, que deve ser aplicada para a resolução do problema.

Este trabalho abrange apenas a criação da ferramenta que será aplicada aos colaboradores. Agora é necessário um repositório, ou um *site* relacionado, que forneça o problema, ou seja, as referências, e uma base para tratar e armazenar os metadados classificados. A utilização da inteligência coletiva será a chave dessa implementação, decidir qual a melhor abordagem e qual o melhor incentivo aos usuários a participarem do projeto, pela obrigatoriedade ou não, é necessário para, ao final, concluir se está realmente é a melhor abordagem para a solução do problema.

Independentemente do resultado obtido, é possível afirmar que, se um grupo de indivíduos tem um interesse em comum e é estimulado da maneira correta a realizar pequenas tarefas que não tomem muito tempo, nem exijam grande esforço, pode-se agregar o resultado dessas tarefas, quase insignificantes em um contexto local, de modo que, ao serem somadas em um contexto global, se tornem significativas, proporcionando uma solução viável ao atual problema de coleção/agregação de metadados.

REFERÊNCIAS

- ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS). NBR 6032: abreviação de títulos de periódicos e publicações seriadas. Rio de Janeiro, 1989
- ABNT (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS). NBR 6023: informação e documentação – referências – elaboração. Rio de Janeiro, 2002. 24p.
- ACAR, Oguz Ali; VAN DEN ENDE, Jan. Motivation, reward size and contribution in idea crowdsourcing. In: **Dime-Druid Academy Winter Conference**. 2011.
- ANDRADE, Ilza Almeida et al. Inteligência coletiva e ferramentas web 2.0: a busca da gestão da informação e do conhecimento em organizações. **Perspectivas em Gestão & Conhecimento**, n. 1, p. 27-43, 2011.
- BONABEAU, Eric. Decisions 2.0: The power of collective intelligence. **MIT Sloan management review**, v. 50, n. 2, p. 45, 2009.
- BRABHAM, Daren C. Crowdsourcing as a model for problem solving: An introduction and cases. **Convergence**, v. 14, n. 1, p. 75-90, 2008.
- CHEN, Kay-Yut; PLOTT, Charles R. Information aggregation mechanisms: Concept, design and field implementation. **California Institute of Technology Social Science Working Paper**, v. 1131, p. 161-194, 2002.
- FUNARO, Vânia Martins Bueno de Oliveira et al. Diretrizes para apresentação de dissertações e teses da USP: documento eletrônico e impresso. **Sistema Integrado de Bibliotecas da USP**, v. 2, 2009.
- HAMMON, Larissa; HIPPER, Hajo. Crowdsourcing. **Business & Information Systems Engineering**, v. 4, n. 3, p. 163-166, 2012.
- HOWE, Jeff. The rise of crowdsourcing. **Wired magazine**, v. 14, n. 6, p. 1-4, 2006.
- Howe, Jeff. **Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business**. New York, USA: Crown Publishing Group New York, 2008.
- INTRONE, Joshua et al. Solving wicked social problems with socio-computational systems. **KI-Künstliche Intelligenz**, v. 27, n. 1, p. 45-52, 2013.
- LEIMEISTER, Jan Marco. Collective intelligence. **Business & Information Systems Engineering**, v. 2, n. 4, p. 245-248, 2010.
- LÉVY, Pierre. A inteligência coletiva: por uma antropologia do ciberespaço. São Paulo: Loyola, 1994. __. **As tecnologias da inteligência**, 1999.
- LIPINSKI, Mario et al. Evaluation of header metadata extraction approaches and tools for scientific PDF documents. In: **Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries**. ACM, 2013. p. 385-386.

LYKOURENTZOU, Ioanna et al. Collective Intelligence Systems: Classification and Modeling. **Journal of Emerging Technologies in Web Intelligence**, v. 3, n. 3, 2011.

MALONE, Thomas W.; LAUBACHER, Robert; DELLAROCAS, Chrysanthos. The collective intelligence genome. **MIT Sloan Management Review**, v. 51, n. 3, p. 21, 2010.

NOUBEL, Jean-François. Collective intelligence: the invisible revolution. **The Transitioner.org**. 2004. Disponível em: <http://www.thetransitioner.org/Collective_Intelligence_Invisible_Revolution_FNoubel.pdf>. Acesso em: 20 maio 2015.

NOV, Oded; ARAZY, Ofer; ANDERSON, David. Crowdsourcing for science: understanding and enhancing SciSourcing contribution. In: **ACM CSCW 2010 Workshop on the changing dynamics of scientific collaborations**. 2010.

O'REILLY, Tim. What is web 2.0. 2005. Disponível em: <<https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Acesso em: maio 2015.

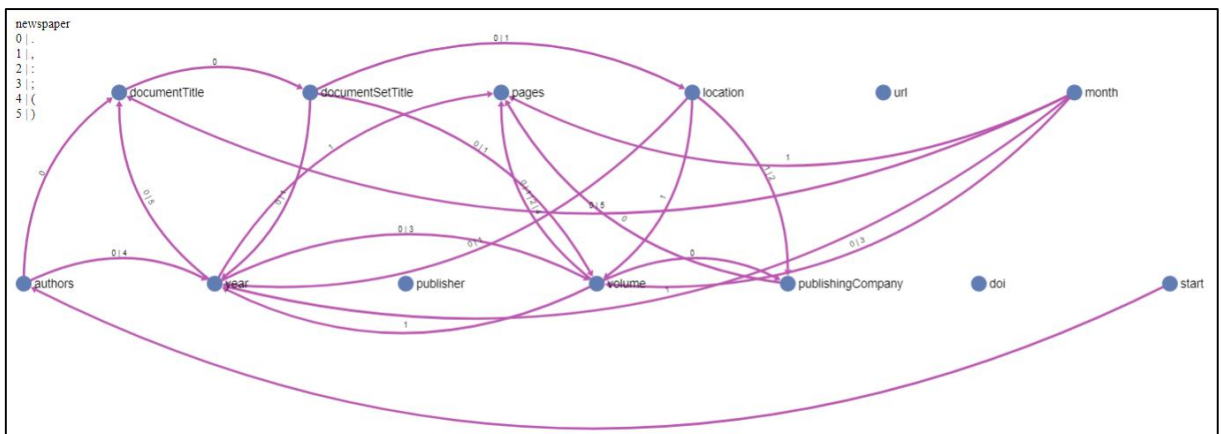
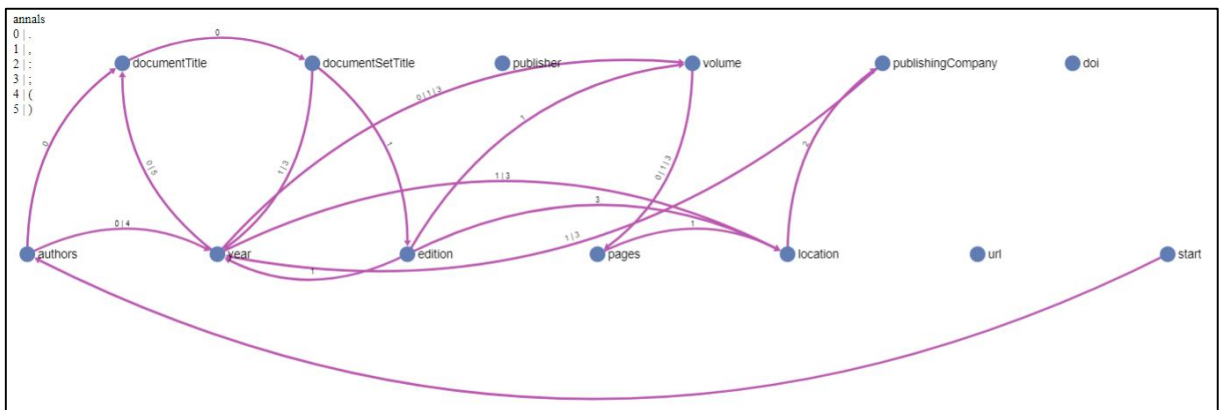
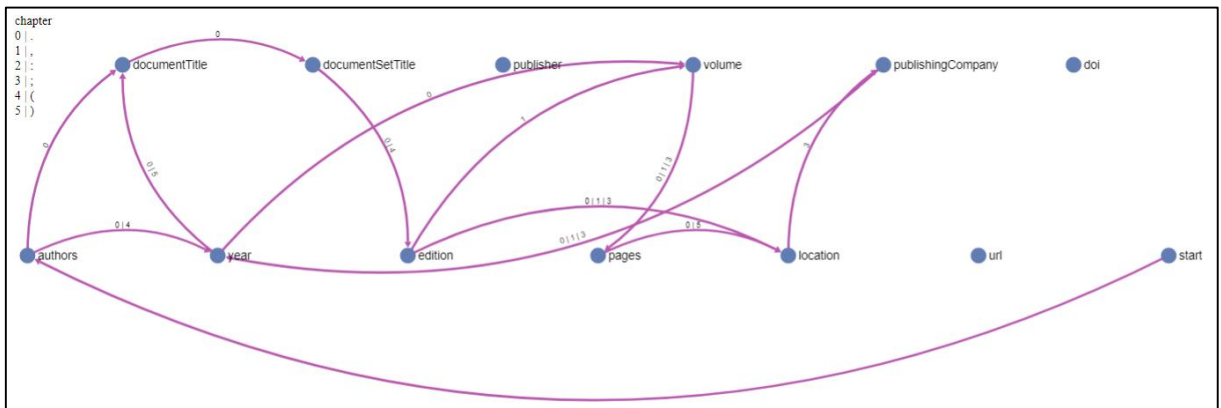
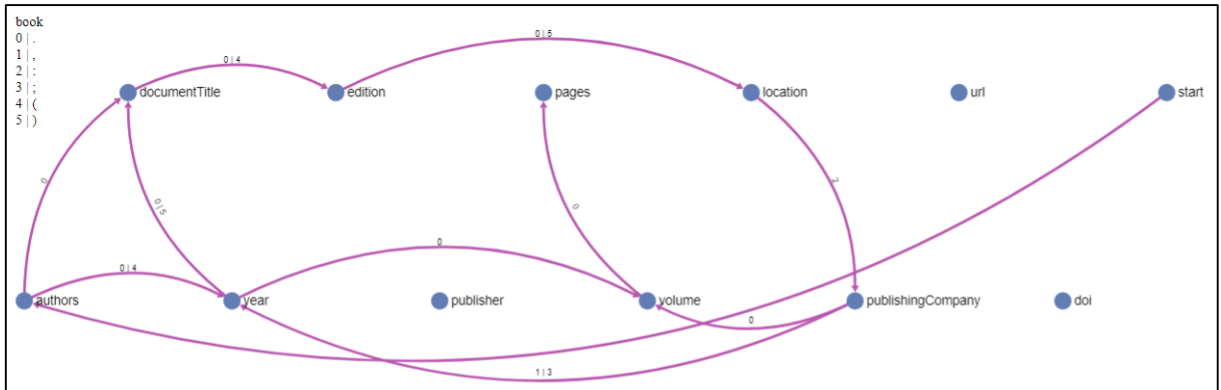
PENG, Fuchun; MCCALLUM, Andrew. Information extraction from research papers using conditional random fields. **Information processing & management**, v. 42, n. 4, p. 963-979, 2006.

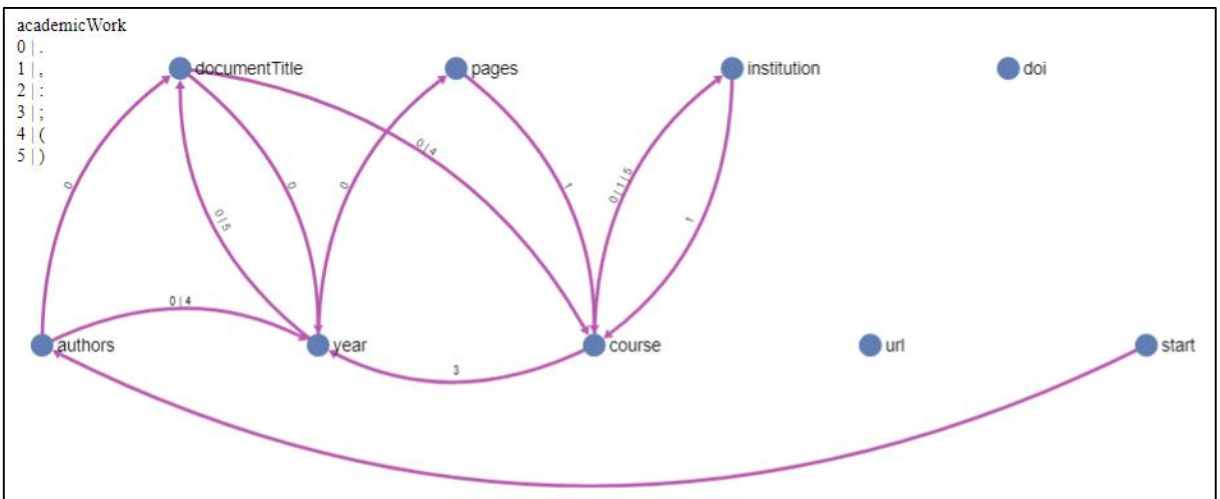
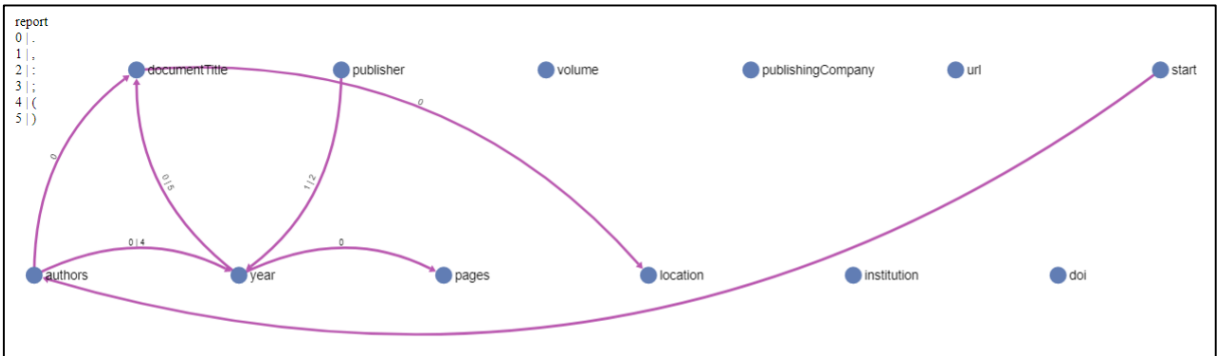
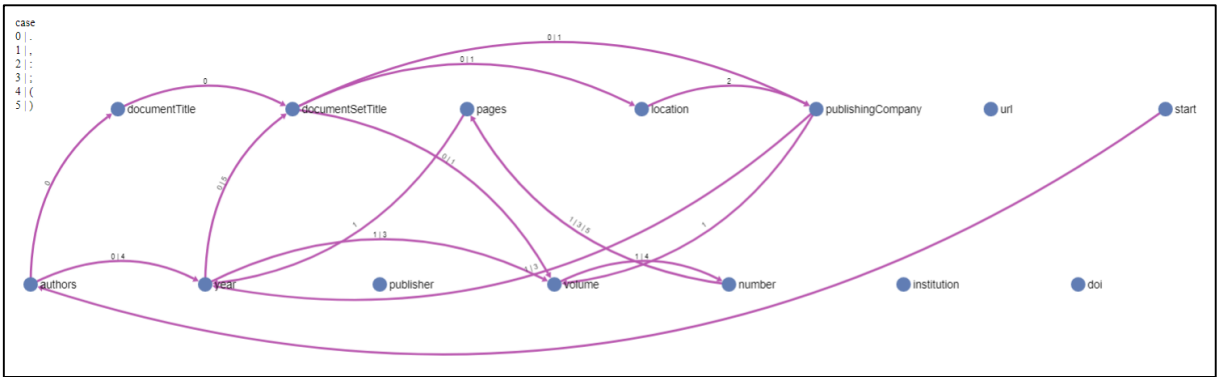
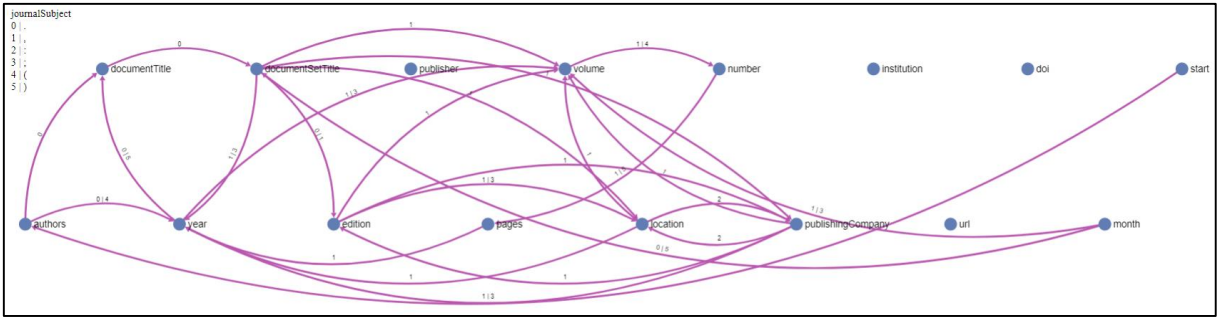
SHANG, Shari SC et al. Understanding Web 2.0 service models: A knowledge-creating perspective. **Information & Management**, v. 48, n. 4, p. 178-184, 2011.

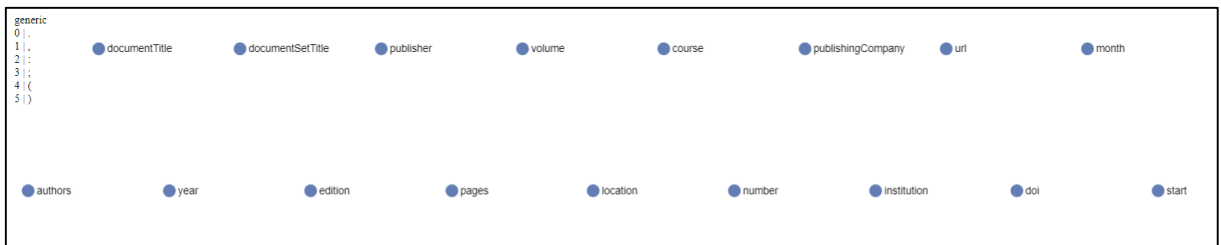
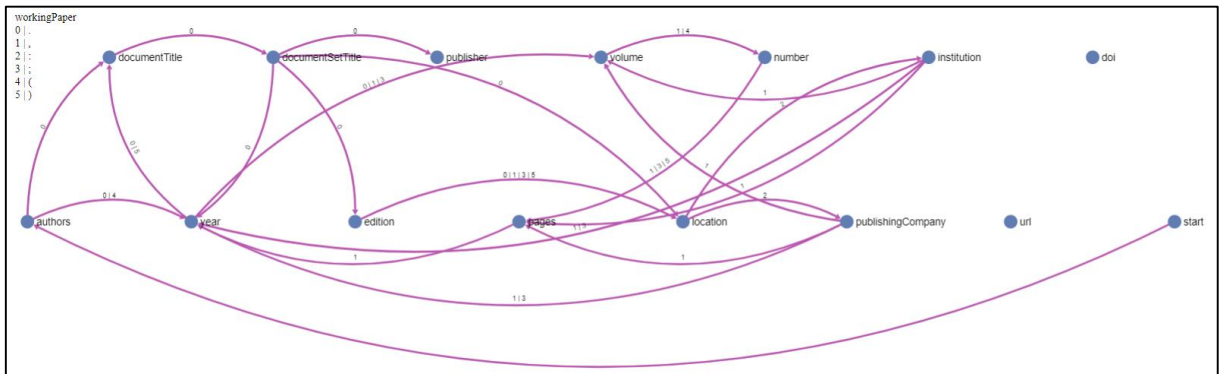
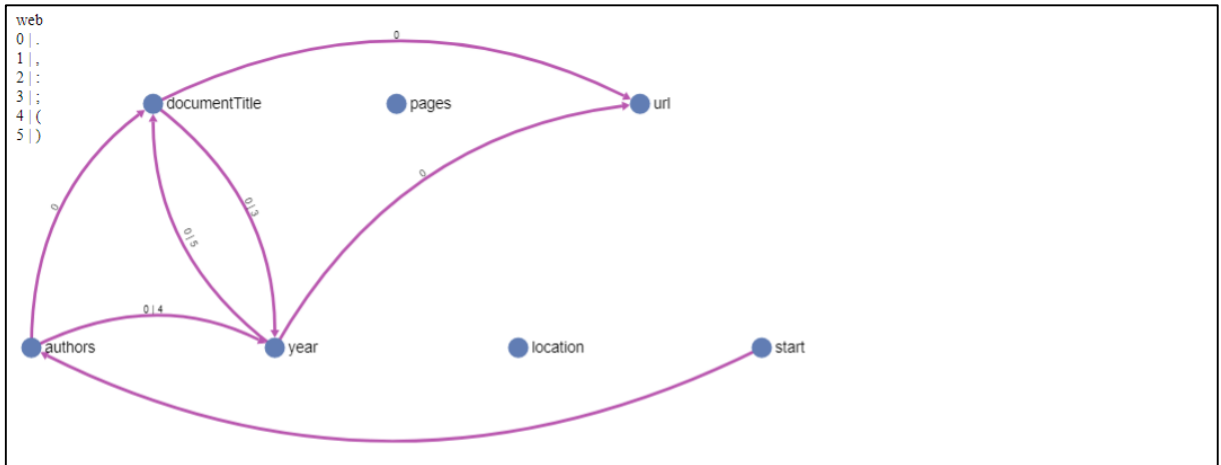
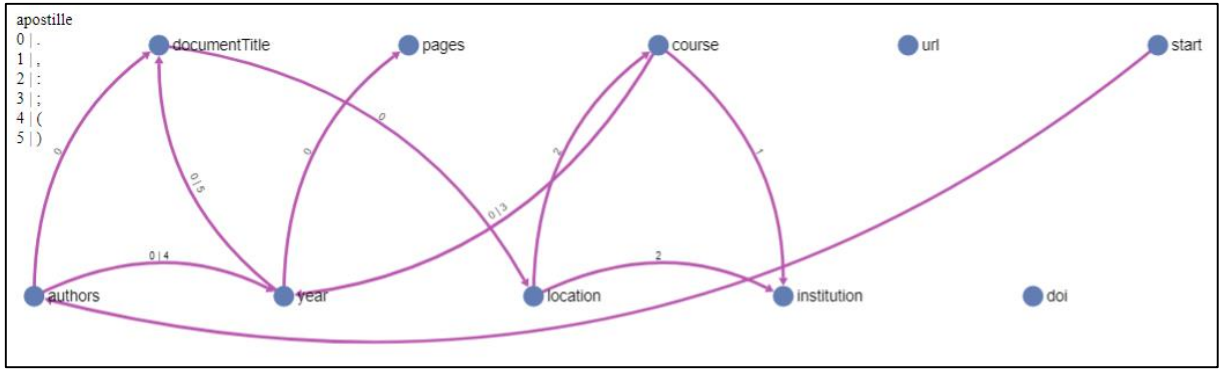
VALLERAND, Robert J. Toward a hierarchical model of intrinsic and extrinsic motivation. In: **Advances in experimental social psychology**. Academic Press, 1997. p. 271-360.

WOOLLEY, Anita Williams et al. Evidence for a collective intelligence factor in the performance of human groups. **science**, v. 330, n. 6004, p. 686-688, 2010.

APÊNDICE A – Grafos por tipo de documento.







APÊNDICE B – Código fonte e arquivos necessários ao artefato.

- Código do Artefato - <https://github.com/ht568cf/crowdMetaApp/>
- Aplicação Auxiliar - <https://github.com/ht568cf/auxCrowdMetaApp/>
- Exemplo do Artefato - <http://crowdmetaapp.atwebpages.com/>

APÊNDICE C – Descrição das funções publicas do artefato.

CrowdApp.run(input, configAddress, outputAddress)

Executa a aplicação, implicando em bloqueio do fluxo normal de navegação do site, retorna verdadeiro se foi executado e falso caso ocorra erro.

Input : String: contendo uma ou mais referências, ou endereço que retorne a uma requisição GET uma String contendo uma ou mais referencias.

Array[String]: referencias separadas em um array de Strings.

configAddress: String: endereço do arquivo de configuração.

outputAddress: String: endereço para requisição POST da resposta.

CrowdApp.getConfigJson()

Retorna o arquivo de configuração caso haja um e nulo caso contrário.

CrowdApp.isUrl(url)

Testa a validade de um endereço url, retorna verdadeiro se a url for válida, e falso caso contrário.

url: String: endereço.

CrowdApp.validateReferences(referencies)

Testa se o conteúdo de referencies é uma referência válida para a aplicação.

referencies: String: uma ou mais referencias separadas por caractere de nova linha.

CrowdApp.close()

Encerra a aplicação e descarta todos os dados.

ANEXO A – Resumo das normas referenciais

Nessa seção, são analisadas as normas ABNT, Vancouver, APA e ISO, com relação à elaboração do referencial bibliográfico. Dentre as orientações contidas nas normas, apesar de englobarem os diversos tipos de fontes, serão descritas apenas as orientações para as fontes textuais, relevantes para a elaboração do artefato, dado o escopo do trabalho.

Os subtópicos dessa seção sumarizam a norma ABNT, utilizada como referência para os demais padrões, e suas diferenças, uma vez que, há poucas diferenças entres os modelos e é a norma mais utilizada no Brasil.

A.1 - ABNT

A norma NBR 6023 (ABNT, 2002), homologada pela ABNT, tem como um dos seus objetivos orientar a preparação e compilação de referências dos materiais utilizados na elaboração de documentos. Para tal, a regra define tipos, elementos, formatação e ordem, a fim de padronizar os componentes das referências de documentos.

Dentre as definições apresentadas pela regra e utilizadas nesse trabalho, podem-se considerar referencias como:

“Conjunto padronizado de elementos descritivos, retirados de um documento, que permite sua identificação individual.” (ABNT, 2002).

E documentos, como:

“Qualquer suporte que contenha informação registrada, formando uma unidade, que possa servir para consulta, estudo ou prova. Inclui impressos, manuscritos, registros audiovisuais, sonoros, magnéticos e eletrônicos, entre outros.” (ABNT, 2002).

Como referido na definição, a regra trata referências como um conjunto de elementos descritivos. Esses elementos são apresentados na subseção 2.3.1.1, e para diferentes tipos de documentos textuais, agrupados em semelhança, a regra define uma formatação específica da referência, resumida na subseção 2.3.1.2.

A.1.1 - Elementos

A NBR 6023 distingue dois tipos de elementos na composição das referências, os essenciais, indispensáveis à identificação do documento, e os elementos complementares, que quando necessários, são adicionados aos elementos essenciais, a fim de completar a descrição do documento (ABNT, 2002).

Os elementos essenciais são, pela norma, obrigatórios na elaboração das referências e variam com relação ao tipo do documento. Já os elementos complementares são facultativos, variantes em relação às características do documento, e aparecem depois dos elementos obrigatórios.

Essa seção apresenta a definição de cada elemento em unidade, juntamente com sua formatação, sem considerar o seu tipo. Para tal, vide subseção 2.3.1.2, como também, para análise do conjunto de elementos.

• Autor(es)

A norma define que deve ser utilizado o Código de Catalogação Anglo-Americano (CCAA, 2002) vigente para entrada de nomes de pessoas física ou entidades (ABNT, 2002). Os nomes dos autores são indicados pelo sobrenome, em letras maiúsculas, e os prenomes abreviados ou não, separados por uma vírgula e espaço. Em caso de múltiplos autores, devem ser separados por ponto-e-vírgula, seguido de espaço.

Quando existirem mais de três autores, indica-se apenas o primeiro, acrescentando-se a expressão “*et al*” em itálico.

Se for necessário destacar o tipo de responsabilidade pelo conjunto da obra (organizador, coordenador, compilador, editor, etc.) deve ser apresentado de forma abreviadas entre parênteses após o nome do responsável.

Em documentos em que o autor seja uma entidade, o nome da entidade por extenso, em caixa alta, é utilizado como autor, e em sequência hierárquica descendente.

Em caso de autoria desconhecida, a entrada é feita pelo título. O termo anônimo não deve ser usado em substituição ao nome do autor desconhecido.

Quando referenciados múltiplos documentos em sequência de um mesmo autor, pode-se substituir as seguintes por 6 traços sublineares seguidos de um ponto final.

• Título e subtítulo

O título e o subtítulo (se for usado) devem ser reproduzidos tais como figuram no documento, separados por dois-pontos.

Em caso de títulos muito grandes pode-se suprimir as últimas palavras, substituindo-as por "...", sem que se altere o sentido do título.

Títulos de periódicos podem ser abreviados, segundo a norma. Documentos sem título podem ser identificados por palavra ou frase que indique seu conteúdo entre colchetes.

O título passa a ser o primeiro elemento da referência quando se referir a periódicos como um todo.

• Edição

Deve ser transcrita, utilizando abreviatura dos numerais ordinais e da palavra edição, ambas na mesma língua do documento. Ementas e acréscimos à edição também devem ser abreviados.

• Local

A cidade de publicação deve ser indicada no documento e, em casos de homônimos de cidades, acrescentam-se, informações adicionais, como estado, país, etc.

Locais transcritos entre colchetes indicam que o documento não contém o local, mais é possível de se identificar. Caso contrário, a abreviação [S.I.] indica que não é possível determinar o local.

• Editora

Deve ser apresentada com abreviatura de prenome e supressão de palavras que indicam a natureza comercial ou jurídica. A expressão [s.n.] *sine nomine*, pode ser utilizada em referências onde a editora não pode ser identificada.

• Datas

Anos devem sempre ser representados por quatro algarismos arábicos, meses devem ser abreviados, sendo apresentados separados por espaço simples. As

horas, minutos, segundos devem ser apresentados com separador dois-pontos, se necessário.

A data de publicação, deve sempre estar presente. E, em caso de conjunto de documentos ou periódicos, o ano do mais novo e do mais velho, separados por hífen, devem ser indicados, caso o conjunto ainda esteja em curso o ano do mais novo é suprimido.

Quando uma data não puder ser identificada, uma data em colchetes é indicada.

- **Volume**

É apresentado pela sigla de volume na língua do documento, seguido por um espaço e a numeração em algarismos arábicos.

- **Paginação**

O número total de páginas pode ser descrito utilizando algarismos arábicos depois de p. ou f.

Em caso de fragmento de documento, é indicada a página de início e a página de fim, separadas por hífen ou indicação de capítulo.

- **Numeração**

Números que expressam ordem são indicados pela abreviação de número seguida do algarismo romano que representa o documento. Porém, números que expressam a identificação do documento em um conjunto ou base, são relacionados pela sigla da base em maiúsculas seguida do número do documento na base.

- **Meio eletrônico**

Quando disponibilizadas ou acessadas em meio eletrônico, as informações do documento, relativas à descrição física do meio, são adicionadas ao fim da referência. Ou em caso de consultas *online*, é utilizada a expressão: “Disponível em:”, seguida da *url* do endereço e da expressão “Acesso em:” seguida da data de acesso, opcionalmente acrescida dos dados referentes à hora, minutos e segundos.

O *Digital Object Identifier*, DOI, tem como objetivo localizar e acessar materiais publicados na web. É apresentado como uma sequência de números única para cada

documento, disponibilizada pelo responsável pela publicação. Se disponível, deve ser apresentado, precedido da palavra “DOI” seguida de dois pontos.

- **Série**

Deve ser apresentada pelo nome ou algarismos romanos, entre parênteses, ao final da referência.

A.1.2 - Tipos

As subseções desse tópico trazem os componentes dos diversos tipos de referências que a norma descreve, os tipos de documento que se aplicam à regra e a ordem, juntamente com sua pontuação específica. Informações relativas em especificidade ao campo, como formatação e características.

- **Monografia completa**

Inclui livros, folhetos, trabalhos acadêmicos (dissertações, teses etc.), manual, guia, catálogo, enciclopédia, dicionário etc. Os elementos essenciais são autor(es), título, edição, local, editora e data de publicação. A pontuação e ordem são apresentadas no formato:

Autor(es). **Título**. Edição. Local: Editora, Data de publicação. Paginação. Série. Notas. ISBN.

Em caso de trabalhos acadêmicos, deve-se informar o grau, informando o tipo, Mestrado, Doutorado, TCC, Monografia, etc., a unidade de defesa, o local, e a data de defesa nessa ordem como apresentado abaixo.

Grau (Mestrado, Doutorado, TCC, Monografia, etc.) – Unidade de defesa, local, data de defesa.

- **Parte de monografia**

Inclui capítulo, volume, fragmento, parte de obra com autor e título próprio. Os elementos essenciais são autor(es), título da parte, seguido da expressão “In:”, e da

referência completa da monografia. No final da referência, deve-se informar a paginação ou outra forma de individualizar a parte referenciada.

Autor(es) do capítulo. Título do capítulo. In: Autor(es) do documento. **Título do documento**: Edição, Local de publicação: Editora, data da publicação. Indicação do capítulo (páginas ou numeração).

Em caso de o(s) autor(es) do documento, e do capítulo ser o(s) mesmo(s), usam-se seis traços seguidos de ponto, no local de “autor(es) do documento”.

- **Publicações periódicas**

Inclui toda coleção de um título de periódico. Os elementos essenciais são: título, local de publicação, editora, datas de início e de encerramento da publicação, se houver.

TITULO DO DOCUMENTO. Local de publicação: Editora, ano. Periodicidade.

- **Parte de revista**

Inclui volume, fascículo, números especiais e suplementos sem título próprio. Os elementos essenciais são: título da publicação, local de publicação, editora, numeração do ano e/ou volume, numeração do fascículo, informações de períodos e datas de publicação.

TITULO DO DOCUMENTO. Local de publicação: Editora, ano. Periodicidade.

- **Artigo ou matéria de revista**

Inclui publicações ou parte, com títulos próprios, de periódicos. Os elementos essenciais são: autor(es), título da parte, artigo ou matéria, título da publicação, local de publicação, numeração correspondente ao volume e/ou ano, fascículo ou número, paginação inicial e final, quando se tratar de artigo ou matéria, data ou intervalo de publicação e particularidades que identificam a parte (se houver).

Autor(es) do artigo. Título do artigo/matéria: **Título da revista**, Local de publicação, volume, numeração, paginação, ano de publicação.

- **Artigo ou matéria de jornal**

Inclui comunicações, editorial, entrevistas, resenhas, reportagens, resenhas e outros.

Os elementos essenciais são: autor(es) (se houver), título, título do jornal, local de publicação, data de publicação, seção, caderno ou parte do jornal e paginação correspondente. Quando não houver seção, caderno ou parte, a paginação do artigo ou matéria precede a data.

Autor(es) do artigo (se houver). Título do artigo: **Título do jornal**, local de publicação, data de publicação, seção, paginação, ano de publicação.

- **Evento**

Inclui atas, anais, resumos entre outras. Os elementos essenciais são: nome do evento, numeração (se houver), ano e local (cidade) de realização. Em seguida, deve-se mencionar o título do documento (anais, atas, tópico temático etc.), seguido dos dados de local de publicação, editora e data da publicação.

NOME DO EVENTO, numeração do evento (arábico), ano, local de realização. Título do documento. (Anais, Atas, etc.)... Local de publicação: Editora, ano de publicação. Paginação.

- **Trabalho apresentados em evento**

Os elementos essenciais são: autor(es), título do trabalho apresentado, seguido da expressão In: nome do evento, numeração do evento (se houver), ano e local (cidade) de realização, título do documento (anais, atas, tópico temático etc.), local, editora, data de publicação e página inicial e final da parte referenciada.

Autor(es). Título do documento. In: NOME DO EVENTO, numeração do evento (arábico)., ano, local de realização. **Título do documento** (Anais, Atas, etc.)... Local de publicação: Editora, ano de publicação. Paginação.

- **Patente**

Os elementos essenciais são: entidade responsável e/ou autor, título, número da patente e datas (do período de registro).

ENTIDADE RESPONSÁVEL. Autor(es), **Título**. Número da patente, datas (período de registro).

- **Legislação**

Compreende a Constituição, as emendas constitucionais e os textos legais infraconstitucionais (lei complementar e ordinária, medida provisória, decreto em todas as suas formas, resolução do Senado Federal) e normas emanadas das entidades públicas e privadas (ato normativo, portaria, resolução, ordem de serviço, instrução normativa, comunicado, aviso, circular, decisão administrativa, entre outros).

Os elementos essenciais são: jurisdição (ou cabeçalho da entidade, no caso de se tratar de norma), título, numeração, data e dados da publicação. No caso de Constituições e suas emendas, entre o nome da jurisdição e o título, acrescenta-se a palavra Constituição, seguida do ano de promulgação, entre parênteses.

- **Jurisprudência**

Compreende súmulas, enunciados, acórdãos, sentenças e demais decisões judiciais. Os elementos essenciais são: jurisdição e órgão judiciário competente, título (natureza da decisão ou ementa) e número, partes envolvidas (se houver), relator, local, data e dados da publicação.

- **Doutrina**

Inclui toda e qualquer discussão técnica sobre questões legais (monografias, artigos de periódicos, etc.), referenciada conforme o tipo de publicação.

- **Materiais especiais**

Inclui filmes, gravações, objetos, etc. Em caso de materiais científicos, os elementos essenciais são: autor(es), título, ano e especificação do objeto.

Autor(es). **Título** (denominação, ou indicação {sem título}). Ano. Especificação do objeto.

A.2 - ISO

ISO 690 é um padrão internacional para elaborar referências bibliográficas, desenvolvido pela Organização Internacional de Normalização, que especifica elementos, sua a ordem e orientações para produção de referências para produção de documentos e inclusão em bibliografias, resumos, etc. (Funaro *et al.*, 2016).

Foi a base para a elaboração da norma 6023. Por isso não há grandes diferenças com relação à norma ABNT descrita acima, os pontos de discrepância são mínimos, citados a seguir.

Este modelo de referência segue os modelos de citação, numerado e autor-ano. Enquanto no primeiro a norma é semelhante à NBR 6023, no segundo, o ano de publicação é apresentado logo após os autores.

Documentos consultados *online*, são apresentados pela expressão “[online]” após o título e a expressão “[visitado dia mês ano]”, após a editora ou ano, dependendo do modelo de citação, juntamente com “Disponível em: url.” como último elemento da referência.

A indicação do volume, número e paginas pode ocorre na sua forma resumida segundo a expressão, “edição (numero), páginas”.

Os títulos dos documentos não são destacados em negrito, os títulos dos documentos são destacados em itálico, caso o documento não seja parte de um conjunto (revista, periódico, jornal, etc.), caso contrário, somente o título do conjunto passa a ser apresentado em itálico.

Documentos com mais de dois autores, para separação entre os dois últimos listados é utilizado o “e”.

Para referências em outras línguas o equivalente das expressões é utilizado.

A.3 - Vancouver

Estilo ou norma de Vancouver, foi elaborado pelo Comitê Internacional de Editores de Revistas Médicas (ICMJ, sigla em inglês), em 1978, para ditar as regras de apresentação de documentos da área médica.

Dentre estas regras (Funaro *et al.*, 2016), encontram-se as que explicam como preparar, organizar e formatar dados de referências para elaboração de documentos.

Levando em consideração a norma da ABNT como padrão, a norma de Vancouver possui semelhanças e discrepâncias. Para o escopo desse trabalho, serão destacadas nessa seção apenas informações que diferem da norma adotada como padrão.

Com relação a autoria, de uma ou múltiplas pessoas físicas, os sobrenomes dos autores não são expressos em letras maiúsculas, e sim adotando a nomenclatura normal de nomes próprios. Para prenomes, é optativa a abreviação, utilizando a primeira letra dos prenomes que se optar por abreviar, e caso ocorra, é apresentada logo após o sobrenome, separada por um espaço, sem pontuações entre múltiplos prenomes e pontos finais, sendo utilizada vírgula para a separação entre múltiplos autores. Caso não haja abreviação, a pontuação continua a ser a padrão.

A indicação de volume, se houver, é feita em numerais arábicos seguidos do número da edição entre parênteses, sem espaços.

Em caso de documento obtido na Internet, deve constar a expressão “[internet]” ao lado do título do documento ou entidade, seguida da última data de atualização do documento, juntamente com a data da citação, como o modelo “atualizado ano mês dia; citado ano mês dia], após o título do fragmento. A indicação “Disponível em: endereço *online*”, deve ser acrescentada no fim da referência.

Em caso de documentos não textuais, o tipo do documento é indicado entre colchetes, após o título.

A ordem e os elementos essenciais dessa norma se assemelham muito à norma da ABNT.

A.4 - APA

A Associação Americana de Psicologia (*American Psychological Association* [APA]), é uma instituição que atua nos Estados Unidos e no Canadá, representando a área de psicologia.

Desde de 1928, a APA publica diretrizes para orientar o desenvolvimento de documentos com qualidade (FUNARO, 2016).

Dentre essas diretrizes, a orientação para a organização e compilação do referencial, como os anteriores, não difere muito da NBR 6023. Porém, há alguns aspectos que merecem ser destacados.

Na indicação de autoria, a separação entre sobrenomes e prenomes é indicada pela vírgula, onde os sobrenomes são expressos em nomenclatura normal de nomes próprios e os prenomes são abreviados normalmente. Em caso de múltiplos autores, a indicação de cada um é separada por vírgula, exceto o último, que é separado do anterior por uma vírgula e o “&”. Caso o número de autores seja maiores que sete, são indicados os 6 primeiros e o último, onde a indicação do último é precedida por reticências, “...”, e um espaço.

A data de publicação é indicada logo após a autoria. Caso ausente no documento, a data de impressão ou *copyright* é indicada. Porém se nenhuma data puder ser encontrada, a notação “*no date*” abreviada, entre parênteses “(n.d.)” é indicada. Em caso de livros e artigos aceitos, mas não publicados ainda, deve-se indicar a expressão *no prelo*, ou equivalente em outras línguas, entre parênteses.

O título do documento ou em caso de partes, o título do conjunto, deve ser apresentado em itálico. A indicação de documentos não convencionais, é igual ao modelo de Vancouver. A série é indicada entre parênteses, logo após o título.

A indicação do volume é apresentada antes da paginação, utilizando a quantidade seguidas das abreviaturas “Vol.” ou “Vols.”, salvo em periódicos, onde o volume é indicado logo após o título do periódico, também destacado em itálico e separado por vírgula, dispensando a abreviação de volume. Quando houver número do fascículo, deve ser apresentado entre parênteses logo após o volume, seguido de vírgula e espaço, antes da indicação de páginas.

Quando a editora for a mesma responsável pela autoria, é utilizado a palavra “Autor”, após o local de publicação. Entre a indicação do local de publicação e a editora são utilizados dois pontos.

A paginação é indicada logo após o título em destaque entre parênteses e precedida da abreviatura “pp.”.

Caso múltiplas informações sejam apresentadas entre parênteses após o título, elas devem ser agrupadas em um título só e separadas entre si por vírgulas.

Para documentos *online*, se o número DOI não estiver indicado, o endereço eletrônico deve ser apresentado, precedido da expressão “Recuperado de”.

ANEXO B – Abreviaturas de meses – ABNT NBR 6023.

Abreviatura dos meses

<u>Português</u>		<u>Espanhol</u>		<u>Italiano</u>	
janeiro	jan.	enero	enero	gennaio	genn.
fevereiro	fev.	febrero	feb.	febbraio	febr.
março	mar.	marzo	marzo	marzo	mar.
abril	abr.	abril	abr.	aprile	apr.
maio	maio	mayo	mayo	maggio	magg.
junho	jun.	junio	jun.	giugno	giugno
julho	jul.	julio	jul.	luglio	luglio
agosto	ago.	agosto	agosto	agosto	ag.
setembro	set.	septiembre	sept.	settembre	sett.
outubro	out.	octubre	oct.	ottobre	ott.
novembro	nov.	noviembre	nov.	novembre	nov.
dezembro	dez.	diciembre	dic.	dicembre	dic.
<u>Francês</u>		<u>Inglês</u>		<u>Alemão</u>	
janvier	janv.	January	Jan.	Januar	Jan.
février	févr.	February	Feb.	Februar	Feb.
mars	mars	March	Mar.	März	März
avril	avril	April	Apr.	April	Apr.
mai	mai	May	May	Mai	Mai
juin	juin	June	June	Juni	Juni
juillet	juil.	July	July	Juli	Juli
août	août	August	Aug.	August	Aug.
septembre	sept.	September	Sept.	September	Sept.
octobre	oct.	October	Oct.	Oktober	Okt.
novembre	nov.	November	Nov.	November	Nov.
décembre	déc.	December	Dec.	Dezember	Dez.