

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO  
MESTRADO EM ADMINISTRAÇÃO

DISSERTAÇÃO DE MESTRADO

A INTELIGÊNCIA COLETIVA NA BIOINFORMÁTICA:  
UMA REVISÃO SISTEMÁTICA DA LITERATURA

RODRIGO DE FREITAS

CURITIBA  
2019

RODRIGO DE FREITAS

A INTELIGÊNCIA COLETIVA NA BIOINFORMÁTICA:  
UMA REVISÃO SISTEMÁTICA DA LITERATURA

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Administração, Curso de Mestrado em Administração, Programa de Pós-graduação em Administração, Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Alexandre Reis Graeml

CURITIBA

2019

### **Dados Internacionais de Catalogação na Publicação**

Freitas, Rodrigo de

A inteligência coletiva na bioinformática [recurso eletrônico] : uma revisão sistemática da literatura / Rodrigo de Freitas.-- 2019.

1 arquivo eletrônico (97 f.) : PDF ; 1,28 MB.

Modo de acesso: World Wide Web.

Texto em português com resumo em inglês.

Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná.

Programa de Pós-graduação em Administração. Linha de Pesquisa: Tecnologia e Desenvolvimento Organizacional, Curitiba, 2019.

Bibliografia: f. 92-97.

1. Administração - Dissertações. 2. Inteligência coletiva. 3. Bioinformática. 4. Web 2.0 (Sistema de recuperação da informação). 5. Web Semântica. 6. Bioinformática - Projetos - Desenvolvimento. 7. Biotecnologia - Pesquisa. 8. Publicações científicas. 9. Sistemas de recuperação da informação. 10. Integração de dados (Computação). I. Graeml, Alexandre Reis, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Administração. III. Título.

CDD: Ed. 23 -- 658

**Biblioteca Central do Câmpus Curitiba - UTFPR**  
**Bibliotecária: Luiza Aquemi Matsumoto CRB-9/794**

Ministério da Educação  
**Universidade Tecnológica Federal do Paraná**  
**Campus Curitiba**  
Diretoria de Pesquisa e Pós-Graduação  
Programa de Pós-Graduação em Administração

## **TERMO DE APROVAÇÃO**

### **A INTELIGÊNCIA COLETIVA NA BIOINFORMÁTICA: UMA REVISÃO SISTEMÁTICA DA LITERATURA**

Por

**Rodrigo de Freitas**

Esta dissertação foi apresentada às **16h00, dia 06 de setembro de 2019** como requisito parcial para a obtenção do título de MESTRE EM ADMINISTRAÇÃO, na Linha de Pesquisa **Tecnologia e Desenvolvimento Organizacional**, do Programa de Pós-Graduação em Administração da Universidade Tecnológica Federal do Paraná. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após a deliberação, a Banca Examinadora considerou o trabalho **APROVADO**.

---

Prof. Dr. Alexandre Reis Graeml  
(PPGA/UTFPR)  
Orientador

---

Prof. Dr. Jurandir Peinado  
(PPGA/UTFPR)  
Membro Interno

---

Profa. Dra. Rita Cristina Galarraga Berardi  
(UTFPR)  
Membro Interno

---

Prof. Dr. Paulo Sergio Macuchen Nogas  
(PUC-PR)  
Membro Externo

---

Prof. Dr. Thiago Cavalcante Nascimento  
(PPGA/UTFPR)  
Coordenador do PPGA

**Programa de Pós-Graduação em Administração (PPGA)**

*Dedico este trabalho a Deus, a Nossa Senhora Aparecida, a minha esposa, minha  
filha, meus familiares e amigos.*

## Agradecimentos

Não vejo de outra forma senão começar agradecendo a Deus pelo dom da minha vida, tudo o que eu tenho e todos a minha volta devo a Deus, um pai que me dá forças para seguir, que caminha ao meu lado e nos momentos difíceis me carrega no colo. Estendo este agradecimento a Nossa Senhora Aparecida, a quem carinhosamente tenho o privilégio de chamar de “mãezinha do céu”, pelo zelo e cuidado, um cuidado de mãe, buscando interceder pelo filho, me cobrindo com seu manto nos momentos difíceis. Agradeço imensamente a Deus e a Nossa Senhora Aparecida.

Quero agradecer a minha família, começando pela minha esposa, uma companheira que tem me dado apoio de longa data, uma mulher incrível, que suportou junto comigo os percalços desta caminhada, em vários momentos executando tarefas da minha responsabilidade ou até mesmo saindo de casa para que eu pudesse me dedicar aos estudos, sendo compreensiva e me motivando. Esposa que, em certos momentos de reclamação da minha parte, me motivava com frases do tipo “*Você não quer ser mestre? Então aguenta e não reclama!*” ou “*Agora que você começou, você vai terminar, vai até o fim!*”

Ainda sobre a família, não poderia deixar de agradecer a minha filha, uma benção em nossa vida, que nasceu enquanto eu cursava as disciplinas, que foi concebida no mesmo período em que realizava minha inscrição no processo seletivo do mestrado. Tive que abdicar de momentos juntos para estudar, mas ela sempre me serviu de motivação, dando-me força para continuar, nos momentos difíceis, com seu sorriso sincero e palavras ainda mal pronunciadas. Representou uma força imensa para continuar e, juntamente com minha esposa, foi a inspiração do meu dizer “*Por elas e para elas!* “. Para minha esposa Kateline e minha filha Manuela, o meu amoroso obrigado.

Agradeço aos meus pais por tudo que me proporcionaram, pela criação que a mim foi dada. Hoje, se sou quem eu sou, devo a eles. Têm participação direta nisso, mesmo sabendo de forma superficial do que se trata um mestrado, eles têm me

apoiado para não parar de estudar. Aos meus pais, muito obrigado. Aproveito para agradecer ao restante dos meus familiares que me apoiaram durante todo esse período, seja com palavras de carinho e compreensão da minha ausência em vários momentos. Muito obrigado.

Direciono um agradecimento aos professores da UTFPR pelo conhecimento compartilhado durante as disciplinas, discussões que despertaram a reflexão e o pensar diferente, todos de certa forma têm uma participação neste trabalho, mesmo que indireta, em algum momento foram lembrados. Gostaria de dar uma ênfase ao Prof. Jurandir, que além de ter compartilhado seu conhecimento durante o período das disciplinas, o fez na banca de qualificação, fazendo ótimas sugestões em torno da estrutura do trabalho, propondo reflexões que vieram a somar neste trabalho. Estendo o agradecimento à Prof.<sup>a</sup> Rita, que também muito contribuiu na banca de qualificação. Uma de suas observações mudou a trajetória do trabalho, ao sugerir uma abordagem focada dentro da biotecnologia, uma delimitação mais apurada, enfim, uma intervenção “cirúrgica”. Aos professores, o meu muito obrigado.

Ao falar de professores não poderia deixar de agradecer a alguém que, sem dúvida, mudou a minha vida, que me fez ver a ciência e a academia de outra forma, meu orientador Prof. Alexandre Graeml. Hoje o considero como “um pai acadêmico”, alguém que, frente a dificuldade, procura ajudar de forma incondicional. Ao se deparar com as minhas limitações, as apontava e me apoiava na melhoria, alguém que sabe reconhecer as minhas qualidades, tais atitudes comparadas a de um pai. Desde o início do mestrado, procurou meu aprimoramento e, nos últimos tempos, dedicou uma atenção especial para a conclusão desse trabalho. Não seria exagero da minha parte pontuar que nos últimos anos, em alguns momentos, falei mais vezes com o Prof. Graeml do que com o meu pai. Por tudo que me proporcionou e o que ainda pode me proporcionar, a minha admiração e meu agradecimento especial.

Gostaria de agradecer aos meus amigos e colegas que, por meio de palavras e gestos, me apoiaram desde a tomada da decisão de me inscrever no processo seletivo para o programa de mestrado até os momentos de angústia, destinando-me palavras de apoio e conforto. Em vários momentos se fez necessária minha

ausência para que eu pudesse estudar, ausência física, em eventos ou até mesmo nos grupos de conversa, para eles o meu muito obrigado.



*“Senhor, dai-me força para mudar o que pode ser mudado...  
Resignação para aceitar o que não pode ser mudado...  
E sabedoria para distinguir uma coisa da outra.”*

*São Francisco de Assis*

## Resumo

A presente pesquisa buscou mapear os trabalhos científicos que tratam do uso de tecnologias da informação, com ênfase para a Internet, para fomentar a inteligência coletiva e obter resultados de pesquisas em biotecnologia que não seriam possíveis sem os esforços coordenados de contingentes de pesquisadores. Na busca do entendimento da relação entre *Web 2.0* e bioinformática foi realizada uma revisão sistemática da literatura a partir de uma busca no *GoogleScholar* dos trabalhos que abordam a temática. O estudo teve uma visão epistemológica positivista, descritiva e de caráter quantitativo. A pesquisa consistiu na análise de 22 trabalhos que atenderam os critérios de inclusão no *corpus* da pesquisa, os quais foram classificados com base na temática principal e na temática secundária (ou acessória), conforme a preocupação e o foco dado pelos autores aos trabalhos. Como resultado, foram obtidas 6 categorias de interesse, sendo elas *integração de dados*, *pesquisa colaborativa*, *serviço web*, *catálogo de dados*, *disciplina web*, e *web semântica*. A grande maioria dos trabalhos demonstrou interesse, principal ou secundário, por *integração de dados*, *pesquisa colaborativa* ou *serviços web*, que representaram 82% das temáticas tratadas e evidenciam as principais contribuições que a *Web 2.0* trouxe para a pesquisa em biotecnologia. Tal resultado mostra que os pesquisadores veem com grande atenção a utilização da inteligência coletiva em estudos da área, preocupando-se com a integração de grandes bases de dados, acessíveis e com boa qualidade, estimulando a pesquisa colaborativa e o engajamento dos pesquisadores. O desenvolvimento de serviços *web* por meio de ferramentas que facilitem e reforcem a inteligência coletiva no campo da bioinformática também mereceu atenção. A pesquisa revela que os recursos tecnológicos que facilitam o surgimento da inteligência coletiva têm proporcionado ganhos significativos para estudos da área de biologia, consolidando-se a *Web 2.0* e os seus possíveis desdobramentos em uma *web* semântica, como recursos imprescindíveis ao desenvolvimento de pesquisa colaborativa na área.

### Palavras-chave

*Web 2.0*; bioinformática; inteligência coletiva.

### **Abstract**

The objective of this work is to map scientific papers that discuss the use of information technologies, with special emphasis to the Internet, to foster collective intelligence and to obtain better results from bioinformatics research that would not be possible without the coordinated efforts of researchers through the *web*. In order to understand the relationship between *Web 2.0* and bioinformatics, a systematic literature review was performed based on a research *corpus* obtained from a Google Scholar search of papers that address the theme. The study had a positivist, descriptive and quantitative epistemological view. The research consisted of the analysis of twenty-two papers that met the inclusion/exclusion criteria to be considered, which were classified based on the main theme and the secondary (or accessory) theme, according to the concern and focus given by the authors to the papers. As a result, six categories of interest were obtained, namely *data integration*, *collaborative research*, *web service*, *data catalog*, *web discipline* and *semantic web*. The majority of the studies showed interest (main or secondary) in *data integration*, *collaborative research* or *web services*, which represent 82% of the addressed themes, and highlight the main contributions that *Web 2.0* has brought to bioinformatics research. This result shows that researchers pay close attention to the use of collective intelligence in studies of the field, worrying about the integration of large, accessible and good quality databases, stimulating collaborative research and the engagement of researchers. The development of *web services* through tools that facilitate and reinforce collective intelligence in the field of bioinformatics also deserved attention. The research reveals that the technological resources that facilitate the emergence of collective intelligence have provided significant gains for biology studies, consolidating *Web 2.0* and its possible developments in a *semantic web*, as indispensable resources for the development of collaborative research in the field.

### **Keywords**

*Web 2.0; bioinformatics; collective intelligence.*

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>18</b>
1.1	CONTEXTUALIZAÇÃO.....	18
1.2	FORMULAÇÃO DO PROBLEMA DE PESQUISA .....	20
1.3	OBJETIVOS DA PESQUISA.....	21
1.4	JUSTIFICATIVA TEÓRICA E PRÁTICA .....	21
1.5	ESTRUTURA DA DISSERTAÇÃO.....	23
<b>2</b>	<b>QUADRO TEÓRICO DE REFERÊNCIA</b> .....	<b>24</b>
2.1	WEB 1.0.....	24
2.2	INTELIGÊNCIA COLETIVA.....	25
2.3	WEB 2.0.....	26
2.4	WEB 3.0 OU WEB SEMÂNTICA.....	35
2.5	BIOINFORMÁTICA .....	37
<b>3</b>	<b>METODOLOGIA</b> .....	<b>42</b>
3.1	ESPECIFICAÇÃO DO PROBLEMA.....	42
3.1.1	Apresentação das perguntas ou hipóteses de pesquisa .....	42
3.1.2	Apresentação/definição das categorias analíticas ou variáveis .....	42
3.2	DELIMITAÇÃO E DELINEAMENTO DA PESQUISA .....	42
3.2.1	População e amostra .....	43
3.2.2	Delineamento e etapas da pesquisa .....	43
3.2.3	Procedimentos de coleta de dados .....	45
3.2.4	Procedimentos de tratamento e análise dos dados .....	52
3.2.5	Facilidades e dificuldades na coleta e tratamento dos dados .....	52
3.2.6	Limitações da pesquisa.....	53
<b>4</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b> .....	<b>54</b>
4.1	APRESENTAÇÃO DOS RESULTADOS.....	54
4.1.1	Artigos selecionados para o <i>corpus</i> .....	54
4.1.2	Temporalidade .....	55
4.1.3	Instituições e países envolvidos na pesquisa sobre <i>Web 2.0</i> na bioinformática.....	57
4.1.4	Referências seminais dos trabalhos do <i>corpus</i> da pesquisa.....	63
4.1.5	Referência a outros artigos do <i>corpus</i> da pesquisa .....	67
4.1.6	Categorização dos artigos.....	68
4.2	ANÁLISE DOS RESULTADOS .....	75
4.2.1	Categorização principal dos trabalhos .....	75
4.2.2	Categorização secundária dos trabalhos .....	78
4.2.3	Categorização agregada dos trabalhos .....	80
4.3	DISCUSSÃO DOS RESULTADOS .....	81

<b>5</b>	<b>CONCLUSÃO E RECOMENDAÇÕES .....</b>	<b>86</b>
	<b>REFERÊNCIAS.....</b>	<b>92</b>

## Lista de figuras

Figura 1	Modelo dos 4 fatores da <i>Web 2.0</i> .....	32
Figura 2	Associação de conhecimentos da biologia molecular à informática, que culminou com o advento da bioinformática.....	40
Figura 3	Etapas de realização de um levantamento bibliográfico .....	45
Figura 4	Número de trabalhos publicados x ano de publicação.....	56
Figura 5	Número de trabalhos publicados x países de origem dos periódicos .....	57
Figura 6	Número de universidades x países .....	61
Figura 7	Departamentos x quantidade de pesquisadores .....	62
Figura 8	Quantidade total de vezes que um artigo foi usado como referência pelos trabalhos do <i>corpus</i> .....	64

## Lista de tabelas

Tabela 1	Quantidades de retornos para buscas no Google Scholar.....	48
Tabela 2	Classificação dos artigos contidos nas páginas de busca utilizando os termos de pesquisa “ <i>Bioinformatics</i> ” e “ <i>Web 2.0</i> ”.....	50
Tabela 3	Classificação dos artigos contidos nas páginas de busca utilizando os termos de pesquisa “ <i>Bioinformatics</i> ” e “ <i>Web 3.0</i> ”.....	51
Tabela 4	Trabalhos <i>corpus</i> x quantidade de referências bibliográficas apresentadas .....	63
Tabela 5	Quantidade de vezes que um artigo foi utilizado como referência para trabalhos contidos no <i>corpus</i> .....	65
Tabela 6	Categoria principal/secundária x quantidade de trabalhos.....	74

## Lista de quadros

Quadro 1	Artigos selecionados para a RSL .....	54
Quadro 2	Universidades e/ou centros de pesquisas x quantidade artigos.....	58
Quadro 3	Universidades e centros de pesquisa em que foram realizados os trabalhos incluídos no <i>corpus</i> da pesquisa.....	59
Quadro 4	Quantidade de referências a outros trabalhos contidos no <i>corpus</i> .....	67
Quadro 5	Artigo x categorias principais e secundárias .....	70



## Lista de siglas e abreviaturas

API *Application Programming Interface*

DNA *Ácido desoxirribonucleico*

OECD *Organization for Economic Cooperation and Development*

RNA *Ácido ribonucleico*

RSL *Revisão Sistemática de Literatura*

# 1 INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO

Dentre os avanços tecnológicos de maior destaque na atualidade estão os que se referem à ciência da vida, com particular destaque à biotecnologia.

A palavra biotecnologia foi utilizada pela primeira vez em 1919 pelo engenheiro húngaro Karl Ereky. Contudo, uma definição formal somente foi produzida em 1992, na Convenção sobre Diversidade Biológica (FERRO, 2010).

Em uma perspectiva etimológica, o termo biotecnologia é derivado da união de três palavras gregas: *bio* (vida), *tecnos* (utilização prática) e *logos* (conhecimento) (CUNHA E MELO, 2006). Cunha e Melo (2006) preconizam que a biotecnologia, em um espectro estrito, pode ser definida como a aplicação e utilização de conhecimentos científicos relacionados à vida, embora seja conveniente ressaltar que o termo tem, normalmente, sido aplicado de forma mais ampla.

Assim, Figueiredo, Penteado e Medeiros (2006) proporcionam uma definição de biotecnologia mais ampla, considerando-a como a manipulação de seres vivos, ou parte destes, na produção de bens ou serviços, envolvendo tecnologias de diferentes níveis. A partir desta definição, esses autores afirmam que biotecnologia pode envolver desde a utilização da fermentação na preparação de alimentos até a manipulação genética e os avanços no campo da biologia molecular, representados por alimentos transgênicos, os quais têm, como característica fundamental, a alteração genética de sua composição.

Faleiro e Andrade (2011) fazem uma distinção entre a biotecnologia clássica (ou tradicional) e a biotecnologia moderna. Para esses autores, a biotecnologia clássica tem como característica principal um conjunto de atividades de menor complexidade, já desenvolvidas pelos seres humanos há milhares de anos, incluindo, por exemplo, a fermentação, na qual o papel dos microrganismos vivos é transformar alimentos. Já a biotecnologia moderna teria, como sua característica principal, o envolvimento da tecnologia genética, em que existe a possibilidade de

aplicar uma cultura de células e embriões em DNA recombinante, para utilização na produção de novos produtos ou no desenvolvimento de processos de maior rentabilidade (FALEIRO E ANDRADE, 2011).

Faleiro e Andrade (2011) também dão destaque aos estudos genéticos e à necessidade de tecnologia direcionada para esses estudos, entrando em cena a bioinformática, responsável por prover as ferramentas tecnológicas que apoiam tais estudos.

Bioinformática tem como significado a aplicação da informática em estudos de biologia. O termo foi utilizado pela primeira vez por Hesper e Hogeweg em meados de 1970, descrevendo estudos informáticos em sistemas bióticos (HOGEWEG, 1978).

Conforme explanado por Muniz (2003), a bioinformática pode ser considerada multidisciplinar, uma vez que faz uso de outros recursos, dentre eles, computacionais, físicos, matemáticos e estatísticos, voltados para a análise da biologia molecular. Ao abordar a aplicação da bioinformática na biologia, o autor destaca que esta pode ser dar de duas formas: para a apresentação de grandes massas de dados ou para a análise desses dados. Em ambas as formas, é possível identificar a necessidade de recursos tecnológicos.

Lesk (2002) propõe uma forma diferente de se analisar a aplicação da informática na biologia, enfatizando que o desenvolvimento de algoritmos estruturados e recursos informáticos avançados contribui para a resolução das questões apresentadas pela biologia.

Ao enfatizar a utilização de grandes bases de dados, a bioinformática propõe forma de utilização e, neste ponto, é possível verificar uma relação direta com a *Web 2.0*, uma vez que dados podem ser produzidos coletivamente pela multidão. A *Web 2.0* pode ser entendida como um fruto da coletividade e seus resultados atribuídos à multidão (HANDSFIELD *et al.*, 2009).

O termo *Web 2.0* foi utilizado por O'Reilly e Dale Dougherty em meados de 2005, a fim de contextualizar o método utilizado por algumas empresas na era

“ponto com”, evidenciando o envolvimento dos usuários na *web* (O'REILLY, 2007). Procurando explicar a *Web 2.0*, O'Reilly (2007) reforça o desenvolvimento de uma inteligência coletiva, evidenciando tal feito como característica fundamental para proporcionar a construção do conhecimento de forma descentralizada, tendo a comunidade no papel central.

Uma vez indicado o papel central da multidão na construção coletiva, Boulos e Wheeler (2007) reforçam a importância da participação dos usuários, trazendo como necessidade a existência de uma interface que proporcione tal participação. Em uma perspectiva mais social, Abram (2007) dá um enfoque ao aspecto humano da *Web 2.0*, enfatizando a interação social que resulta na criação de distribuição de conteúdo próprio.

A inteligência coletiva, conforme descrita por Lévy (2003), pode ser entendida como uma inteligência distribuída de forma coordenada e valorizada, resultando em uma mobilização efetiva de diferentes competências. Maleewong *et al.* (2008) e Yuan *et al.* (2008) esclarecem que a inteligência coletiva se refere à habilidade de resolução de problemas por meio da colaboração e integração entre inúmeros indivíduos.

Uma vez que a bioinformática tem sido aprimorada para atender os estudos da biologia, buscou-se entender a aplicação da inteligência coletiva por meio da utilização da *Web 2.0* e a forma como tal aplicação se deu.

## 1.2 FORMULAÇÃO DO PROBLEMA DE PESQUISA

Autores têm apresentado exemplos do uso, e conjecturas sobre o potencial de aplicação futura, de inteligência coletiva por meio de ferramentas da *Web 2.0* para a realização de projetos na área de bioinformática. Mas como essas possibilidades são exploradas pelos autores da área? De que forma tem sido utilizada a inteligência coletiva na bioinformática?

Em um ambiente turbulento, configurado por mudanças rápidas nas tecnologias envolvidas, mas também nas formas de se organizar para a pesquisa, é importante compreender como a comunidade científica percebe essas novas

possibilidades. Para tanto, foi formulada a seguinte questão estruturante para esse projeto de pesquisa: **De que forma a inteligência coletiva tem sido utilizada por meio da aplicação de ferramentas da *Web 2.0* em projetos de bioinformática?**

### 1.3 OBJETIVOS DA PESQUISA

O objetivo principal dessa pesquisa é compreender de que forma a inteligência coletiva tem sido utilizada por meio de ferramentas da *Web 2.0* em projetos na área de bioinformática.

Para atingir esse objetivo maior, foram definidos alguns objetivos específicos acessórios:

- a) Identificar publicações científicas que tratem do uso da inteligência coletiva por meio das ferramentas da *Web 2.0* em projetos na área de bioinformática;
- b) Mapear as áreas de estudos que estão relacionadas com o uso da inteligência coletiva por meio das ferramentas da *Web 2.0* em projetos na área de bioinformática nas publicações científicas identificadas;
- c) Identificar as formas de aplicação da inteligência coletiva por meio das ferramentas da *Web 2.0* em projetos na área de bioinformática nas publicações científicas identificadas.

### 1.4 JUSTIFICATIVA TEÓRICA E PRÁTICA

Muniz (2003) caracteriza a bioinformática como uma área multidisciplinar, a qual, fazendo uso de recursos vários, tem aplicação direta na biologia molecular, trazendo embasamento para novas descobertas sobre organismos vivos. Os autores Hultquist *et al.* (2002) apontam a bioinformática como uma das grandes responsáveis pelo desenvolvimento no campo da biotecnologia, campo este que está ligado diretamente a genética, segmento que se mostra em grande evidência na atualidade.

Com o surgimento de inúmeros estudos distribuídos ao redor do mundo, hoje se dispõe de dados gerados por pesquisadores que podem nem sequer se conhecer pessoalmente, mas estão conectados e interagem na geração de conhecimento novo, ainda que atuando como uma “multidão”. Ao abordar a “multidão de pesquisadores”, Solomon e Schrum (2007) destacam a característica das ferramentas da *Web 2.0* na conexão desses indivíduos, as quais são voltadas explicitamente para a colaboração e manipulação de informações de forma descentralizada.

O presente estudo se demonstra atrativo, a partir de uma perspectiva prática devido aos seguintes fatores:

- a) o rápido crescimento da área da biotecnologia moderna e, conseqüentemente, da bioinformática, demanda estudos que ajudem a compreender de que forma a inteligência coletiva viabilizada pela utilização de ferramentas da *Web 2.0* pode levar a maior eficiência na pesquisa científica da área. De acordo com Hultquist *et al.* (2002), a bioinformática pode ser considerada uma das responsáveis pelo desenvolvimento de tecnologias inovadoras no campo da biotecnologia, a partir de investimentos que crescem ano a ano. Tal fenômeno se deve ao potencial de retorno econômico proporcionado pela inovação decorrente da sua utilização na pesquisa.
  
- b) a possibilidade de tornar a bioinformática ainda mais eficiente, em alguns casos, a partir do uso da inteligência coletiva por meio de ferramentas da *Web 2.0* como forma de suporte em tarefas envolvidas. Uma vez aplicada a tecnologia da computação em nuvem, devido à sua característica de utilização sem restrições geográficas, ou seja, de forma desmembrada e conectada, pesquisadores podem implementar ferramentas que atingem elevado paralelismo em sua atuação, de maneira simplificada (SALDANHA, 2012).

- c) assim, os possíveis resultados podem levar a incremento no bem-estar da sociedade, resultantes da criação de alternativas promissoras para avanços na bioinformática.

## 1.5 ESTRUTURA DA DISSERTAÇÃO

Este trabalho é composto por cinco capítulos, os quais estão distribuídos e estruturados da seguinte forma:

- a) Capítulo 1 – contextualiza o tema, define o problema, os objetivos e justifica a realização desta pesquisa;
- b) Capítulo 2 – é composto por um quadro referencial contendo uma explanação teórica sobre bioinformática, inteligência coletiva, *Web 2.0* e *Web 3.0 / Web semântica*;
- c) Capítulo 3 – descreve os procedimentos metodológicos adotados;
- d) Capítulo 4 – apresenta os trabalhos selecionados para a composição do *corpus* da revisão sistemática da literatura, os resultados obtidos, com mapeamentos e categorizações, a análise e a discussão dos resultados;
- e) Capítulo 5 – apresenta as conclusões obtidas e sugere encaminhamentos para novos estudos.

## 2 QUADRO TEÓRICO DE REFERÊNCIA

Este capítulo tem como propósito trazer à tona alguns conceitos e definições da *Web 1.0*, *Web 2.0*, *Web 3.0* / *Web* semântica e bioinformática.

### 2.1 WEB 1.0

Para entender o que é a *Web 2.0*, se faz necessário em um primeiro momento ilustrar determinadas características da *Web 1.0*, que foi a antecessora e pode ser classificada como uma base para a *Web 2.0*.

Choudhury (2014) propõe que a primeira apresentação da *Web 1.0* se deu em 1989, caracterizada como conexão de informações. Segundo Rosen e Nelson (2008), a *Web 1.0* tem como característica ser estática. Ou seja, como ela permite a alimentação de informações em sentido único, feita por parte de pessoas pré-definidas, sem pretensão de qualquer colaboração por parte dos usuários, o conteúdo tende a não ser modificado com frequência. Esses autores reforçam a importância da *Web 1.0*, salientando que, em meados dos anos 90, os *sites* criados na Internet permitiram a disponibilização de grande quantidade de dados e informações, mesmo que sem participação massiva da comunidade na criação de conteúdo.

Segundo Cormode e Krishnamurthy (2008), a *Web 1.0* possui poucos criadores de conteúdos e uma filosofia diferente da *Web 2.0*. Tal filosofia, resultado do número restrito de criadores de conteúdo, se caracteriza por uma posição unicamente receptiva por parte de quem consome tal conteúdo.

Como Hedberg e Brudvik (2008) explicaram, a *Web 1.0* permite, em alguns casos, uma criação limitada de produtos e suas respectivas impressões. Contudo, tais produtos não podem ser salvos ou alterados após sair dos *sites*, de modo que toda e qualquer manipulação de conteúdos em plataformas da *Web 1.0* é de responsabilidade dos proprietários das plataformas.



## 2.2 INTELIGÊNCIA COLETIVA

Uma definição de inteligência de grande aceitação, atribuída a Wechsler (1964), remete à capacidade composta ou global do indivíduo de nortear suas ações com determinado propósito, com razoabilidade e interação efetiva com o ambiente. Muitas vezes se trata da inteligência de cada indivíduo, mas ela também pode ser observada na forma como coletivos definem e executam suas ações em busca de atingir seus intentos.

Segundo o *Handbook of Collective Intelligence* (2011), confeccionado pelo *MIT Center for Collective Intelligence*, a inteligência coletiva é definida como indivíduos organizados em grupos, os quais realizam atividades de maneira coletiva que parecem ser inteligentes.

A fim de um entendimento etimológico, Leimeister (2010) explica o termo inteligência coletiva separando os termos. Refere-se à “inteligência” como a capacidade de aprendizado, entendimento e adaptação ao ambiente, fazendo uso do próprio conhecimento. O termo “coletivo” é descrito pelo autor como um complemento que faz com que tal inteligência possa ser entendida como o resultado da interação de indivíduos reunidos em grupos, que não precisam necessariamente ter os mesmos pontos de vista. Na verdade, as diferenças de perspectiva dos membros pode ser um fator preponderante para se chegar à solução de determinados problemas.

Ainda segundo Leimeister (2010), a inteligência coletiva teve seu significado atualizado, em especial, devido ao surgimento de aplicativos com características da *Web 2.0* e a possibilidade de geração e compartilhamento de conteúdo pelo usuário, o que requer um envolvimento mais crítico e direto do usuário na *web* de forma coletiva. Ao exemplificar tais ações, Leimeister (2010) cita classificações de produtos, que tendem a influenciar processos de criação de opinião pública, ações executadas coletivamente, gerando assim, inteligência coletiva.

Seragan (2007) enfatiza a característica da aplicação do conhecimento por parte dos grupos para buscar a adaptação de um ambiente em mudança. A composição do comportamento, preferências e ideias diferentes dentro de um grupo,

a partir da diversidade encontrada, proporciona um ambiente favorável à criação de novas percepções.

Assim, Yuan *et al.* (2008) consideram que a inteligência coletiva é obtida a partir de inteligências individuais atuando em um sistema adaptativo complexo composto por indivíduos, onde os agentes adaptativos agem de forma individual e mutuamente, resultando no surgimento gradativo da inteligência coletiva.

Ao tratar da aplicação da inteligência coletiva em questões de construção do conhecimento colaborativo, Maleewong *et al.* (2008) salientam a criação do conhecimento colaborativo. A abordagem da inteligência coletiva incentiva a participação dos integrantes do grupo, emergindo discussões com aceitações e/ou rejeições, de acordo cada posição proposta.

Preece e Shneiderman (2009) focam no papel da tecnologia como mediadora na participação social da multidão, a fim de alcançar objetivos comuns por meio da colaboração na *web*.

Malone e Crowston (1994) ao abordar interação e diferentes perspectivas, fazem uma associação entre a inteligência coletiva e a biologia, ao sugerirem que, da mesma forma que na fisiologia humana diferentes partes do corpo, com diferentes propósitos e organizadas de forma coordenadas, mantêm uma pessoa viva e saudável, grupos organizados podem obter resultados sistêmicos superiores aos que poderiam ser obtidos a partir de esforços individuais, com base no conhecimento colaborativo e na inteligência coletiva gerada a partir da interação.

### 2.3 WEB 2.0

O'Reilly (2007) atribui a si próprio, juntamente com Dale Dougherty, a criação do termo *Web 2.0*, em meados de 2005, para se referir à forma de operação de várias empresas que se sobressaíram com a popularização do fenômeno “ponto com” no início desse século, as quais tinham características comuns.

Handsfield *et al.* (2009), adotando uma perspectiva de pertencimento argumenta que, diferentemente do que ocorria com as ferramentas da *Web 1.0*, as ferramentas da *Web 2.0* são “propriedade” da coletividade ou da multidão que compõe a Internet.

Aquela abordagem de *web* estática, que caracterizava a *Web 1.0* é deixada de lado ao se passar a perceber a *web* como um ambiente dinâmico, que tem em sua essência a comunicação em rede (PAROUTIS, 2009). Um exemplo prático da forma de trabalho colaborativo para a gestão do conhecimento em rede é a *Wikipédia*, conforme Paroutis (2009), a qual depende de usuários que não são apenas consumidores de informação, mas também geradores de conteúdo para ampliá-la ou melhorá-la, o que contrasta com a rigidez das enciclopédias tradicionais.

O'Reilly (2007) enfatiza o desenvolvimento da inteligência coletiva como uma característica fundamental de sucesso da *Web 2.0*, uma vez que sua essência vai ao encontro da ideia da construção descentralizada, tendo a comunidade papel central na construção do conhecimento.

Buscando entender melhor o significado de inteligência coletiva, Lévy (2003, p. 28-29) a descreve como uma “inteligência distribuída por todas as partes, coordenada em tempo real, resultando em uma mobilização efetiva de competências”, em que, segundo o autor, a participação coletiva é valorizada. Quando um grupo de indivíduos colabora ou as pessoas competem entre si tende a surgir uma inteligência coletiva (ALAG, 2009).

Szuba (2001) é mais específico, ao criar uma definição para inteligência coletiva computacional, na qual o grupo de agentes que participa da inteligência coletiva envolve pessoas e computadores. Métodos computacionais possibilitam análises e definições em grupos, por parte dos humanos envolvidos, ou viabilizam o processamento de informações de forma autônoma pelos agentes que se encontram espalhados em ambientes dispersos.

Uma característica fundamental para o sucesso da *Web 2.0*, segundo Boulos e Wheeler (2007), é a participação dos usuários. Para que possa ocorrer essa participação coletiva, os autores reforçam a necessidade de se contar com uma

interface de programação aberta estruturada, permitindo aos usuários a criação, organização, localização e compartilhamento de conteúdo (BOULOS E WHEELER, 2007). Rosen e Nelson (2008) destacam que, além da possibilidade de criação de conteúdo por parte de usuários sem conhecimento técnico significativo, a *Web 2.0* estimula a formação de redes sociais, comunidades e agrupamentos *online*, formados em torno de um determinado segmento ou conteúdo específico.

Um requisito importante para a formação dessas comunidades na rede, de acordo com Cormode e Krishnamurthy (2008), foi a necessidade de um certo pensamento crítico da comunidade, com o propósito de analisar os desafios deste ambiente tecnológico que se encontra em franca evolução.

Sobre a característica das ferramentas, Solomon e Schrum (2007) enfatizam que as ferramentas da *Web 2.0* são projetadas de forma explícita para a colaboração e manipulação das informações, diferentemente do que ocorria com as ferramentas da *Web 1.0*, que mantinham o desenvolvedor original do conteúdo no controle. Para que os usuários pudessem se transformar em produtores de conteúdo, além de consumidores, foi necessário que ferramentas com esse intuito passassem a ter mais relevância, dentro da estrutura da *Web 2.0* (HEDBERG E BRUDVIK, 2008).

Ao descrever sobre *Web 2.0*, Abram (2007) aborda um aspecto mais humano da interatividade por meio de conversas, desenvolvimento de redes interpessoais e personalização. Segundo o autor, a *Web 2.0* envolve um fenômeno social que vai além de experiências sociais, ela se consolidou como um meio de criação e distribuição de conteúdo próprio, de forma aberta, descentralizada de autoridade, com liberdade para se compartilhar e reutilizar conteúdo.

A descentralização da criação, juntamente com a possibilidade de reuso e compartilhamento, proporcionam um valor agregado ao conteúdo criado de forma coletiva, segundo Herrera Muñoz (2013), que faz com que a *Web 2.0* passe a ser percebida de forma distinta do que o era nos tempos da *Web 1.0*.

Fazendo uma relação entre Inteligência e *Web 2.0*, Michalsky (2012) argumenta que inteligência, a partir de um entendimento comum, resulta da

colaboração entre membros de um grupo, logo, o maior desafio da *Web 2.0* é o desenvolvimento de ferramentas para estimular a participação dos indivíduos, fomentando a inteligência coletiva.

Hazlett *et al.* (2005), à luz de perspectiva motivacional, argumentam que a dinâmica centrada nas pessoas precisa levar em consideração questões motivacionais na partilha do conhecimento. Ainda na linha motivacional, MALONE *et al.* (2009) propõem que existem diferentes motivações para que integrantes da comunidade participem dos esforços e iniciativas coletivos característicos da *Web 2.0*.

Malone *et al.* (2009) propõem que tais motivações podem ser de caráter financeiro, havendo oferta de uma recompensa monetária, a busca de reconhecimento entre os pares ou de *status* dentro da comunidade, ou de ordem afetiva, quando os participantes se identificam com a causa. Os autores resumem essas motivações como se relacionando a dinheiro, glória ou amor.

Algumas empresas perceberam que poderiam aproveitar o envolvimento das pessoas, proporcionado pelas tecnologias da *Web 2.0*, para a obtenção de vantagens comerciais (PAROUTIS, 2009).

Sendo mais específico sobre o aproveitamento das pessoas por parte das empresas, McKenna (1995) enfatiza que, ao se envolver os clientes no desenvolvimento do produto, existe uma tendência de fidelização do cliente e aceleração do produto até o mercado, ou seja, colocar a comunidade de clientes, em sua coletividade, para trabalhar em prol do produto, mesmo que indiretamente, pode ser uma boa estratégia para se obter diferencial de mercado.

McKenna (1995) vai além de preconizar o envolvimento do cliente no desenvolvimento do produto, o autor defende a ideia de um diálogo contínuo, a fim de construir um vínculo com o cliente que o transforme em um parceiro da empresa, constituindo um estilo de relação.

Nambisan e Nambisan (2008), por sua vez, ao sugerirem a criação de um ambiente virtual voltado ao cliente, que pode consistir em um fórum em que a

participação dos clientes, por meio de discussões *online*, proporciona acesso a informações valiosas que podem ser utilizadas para aprimorar os produtos e/ou serviços da empresa. Esses autores apresentam casos em que esses ambientes de discussão são usados, também para permitir que clientes com maior conhecimento do produto ofereçam suporte a outros clientes, prestando esclarecimentos acerca dos produtos e/ou serviços.

Algumas das características citadas por O'Reilly (2007) são: *web* como plataforma; dados como diferencial competitivo; usuários agregando valor; modelos de programação simples; software acima da plataforma; experiência do usuário rica; formação de comunidades e aproveitamento da inteligência coletiva. Cada uma dessas características foi discutida em maior detalhe por Michalsky (2012):

- Na *web* como plataforma o software deixa de ter o caráter de produto e passar ter uma conceituação de serviço, mudando o conceito de negócios das empresas. Tal característica permite que funcionalidades sejam adicionadas a todo momento, sendo desenvolvidas e testadas continuamente pelos próprios usuários;
- Os dados como um diferencial competitivo tiram o valor da aplicação e focam nas informações propriamente ditas, assim, bancos de dados que contêm informações que podem ser usadas em diferentes segmentos, como exemplo imagens de satélites entre outros, passam a ter mais valor do que as ferramentas que permitem a sua criação;
- Usuários agregando valor está diretamente relacionado à característica da plataforma como serviço, onde os participantes refinam o conteúdo;
- Modelos de programação simples permitem integrações de diversas maneiras. Um exemplo disso é o desenvolvimento de APIs, que podem ser integradas às plataformas;
- Software acima da plataforma ilustra como a transformação da *web* em plataforma eliminou a necessidade de atualizações frequentes de diversos programas e sistemas operacionais, além de permitir a

utilização por diversos meios eletrônicos (celulares, *tablets*, computadores entre outros);

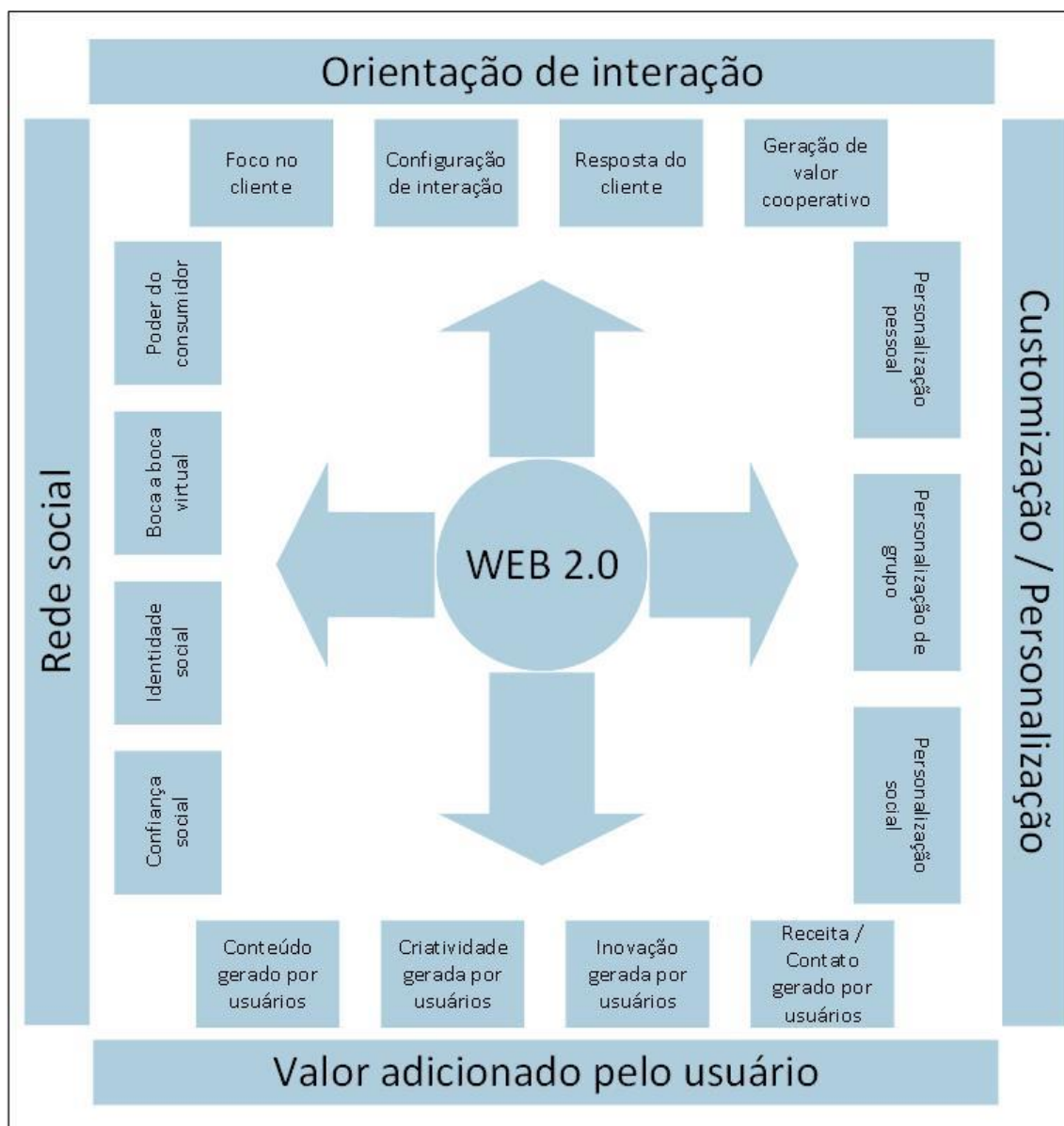
- Experiência rica do usuário na *Web 2.0* proporciona certa facilidade na criação de aplicações voltadas para *web*, com interfaces mais ricas e intuitivas, se comparadas às aplicações *desktop*;
- Formação de comunidades se torna um efeito direto da *Web 2.0*. A aglutinação de indivíduos em torno de assuntos correlatos se dá devido à facilidade de uso das ferramentas de comunicação *online*;
- Aproveitamento da inteligência coletiva pode ser considerado uma das mais importantes características da *Web 2.0*, a qual tem o seu foco na coletividade.

Trazendo a discussão para um contexto de negócios, Wirtz *et al.* (2010) argumenta que as características da *Web 2.0* mudaram o que eles denominam como “as regras do jogo” no que diz respeito a criação de valor, provocando desta forma a disruptura da eficácia dos modelos de negócios estabelecidos na Internet.

Tal fenômeno tem se dado devido ao desenvolvimento da Internet do desenvolvimento de novos modelos de negócio mais ajustados ao ambiente tecnológico por ela proporcionado, focando-se na interação do cliente com a empresa e obrigando a organização a dar a devida atenção ao “diálogo” que pode construir com seus clientes (McKenna, 1995).

Na tentativa de explicar o fenômeno da *Web 2.0*, Wirtz *et al.* (2010) propõem quatro fatores fundamentais: redes sociais; orientação à interação; customização/personalização e valor agregado pelo usuário. Cada um dos quatro fatores apresenta diferentes facetos, conforme mostrado na Figura 1.

Figura 1 Modelo dos 4 fatores da Web 2.0



Fonte: Traduzida de Wirtz, Schilke E Ullrich (2010).

Conforme proposto por Wirtz *et al.* (2010), o primeiro fator são as redes sociais, o qual comumente é descrito como arquiteturas de interações humanas *online*. É composto pelas seguintes facetas:

- **Confiança social:** representa a confiança entre os participantes e a expectativa de retribuição;
- **Identidade social:** refere-se aos usuários que buscam criar uma identidade na *web*, respondendo a um sentimento de pertencimento,



que pode ser “replicado” ou transferido do ambiente virtual para outras relações no mundo real.

- Boca a boca virtual: consiste na transferência de informações entre os envolvidos acerca de determinado assunto, fazendo uso de aplicações informatizadas;
- Aumento do poder do consumidor: o consumidor com maior poder de opinião e ação acaba ganhando “poder”.

O segundo fator proposto por Wirtz *et al.* (2010) consiste na orientação de interação. Este tem como foco a capacidade da empresa gerir eficientemente a crescente demanda por parte dos clientes, a qual resulta de diálogos mais autênticos entre empresa e cliente. Tal fator é composto pelas seguintes facetas:

- Centralidade do cliente: coloca o cliente como prioridade;
- Configuração de interação: se refere à forma como se dá o processo de interação, definindo quais as informações são trocadas, quem é o público alvo, quais os motivos envolvidos, desenhando desta forma a estrutura a ser utilizada para o processo;
- Resposta do cliente: consiste na capacidade da empresa de coletar o *feedback* do cliente, mantendo um diálogo sempre que necessário, buscando ao mesmo tempo a tratativa individual com cada cliente e a catalogação da informação para torna-la gerenciável pela organização;
- Conceito de geração de valor cooperativo: capacidade da empresa e integrar clientes a fim de que se tornem parceiros, visando à obtenção de informações a serem utilizadas para a melhoria de produtos, serviços e processos.

O terceiro fator proposto por Wirtz *et al.* (2010), denominado customização e/ou personalização, foca no individual, na busca do “exclusivo”, adequado a cada demanda. Este fator é composto pelas seguintes facetas:

- Personalização pessoal: conceito que permite ao usuário da Internet personalizar aplicativos ou *sites* de acordo com necessidades e preferências, o que tem se tornado comum em ambientes contextualizados, como a *Web 2.0*;
- Personalização do grupo: proporciona a interação e disseminação de informações por parte de grupos, atendendo a demandas de forma coletiva, construindo e inserindo novas configurações;
- Customização social: consiste na disponibilização de produtos e/ou serviços específicos para camadas sociais distintas.

No quarto e último fator proposto por Wirtz *et al.* (2010), denominado valor agregado pelo usuário, consiste em um foco no usuário e nas informações geradas por ele, estando dividido da seguinte forma:

- Conteúdo gerado pelo usuário: consiste na criação de *sites*, mídias, arquivos de áudio e vídeo, se mostrando uma tendência, obrigando as empresas a integrarem suas aplicações a essa demanda dos usuários da *Web 2.0*;
- Criatividade gerada pelo usuário: o usuário da *Web 2.0*, além de ser responsável pela geração de conteúdo, tende a interferir no processo de criação de novas ideias nas empresas envolvidas neste ambiente;
- Inovação gerada pelo usuário: esta categoria consiste no foco dado às inovações que decorrem de desenvolvimento distribuído;
- Fontes de receitas para empresa: considerada como uma consequência adicional da *Web 2.0*, permitindo que, a partir da interação dos usuários, seja possível a otimização e expansão dos produtos e serviços disponibilizados.

Ainda no âmbito organizacional, Paroutis (2009) destaca que a adoção das tecnologias da *Web 2.0* tem trazido benefícios para as empresas. Contudo, o autor demonstra certa preocupação acerca de benefícios reais advindos do uso dessas

ferramentas por parte de funcionários dessas empresas, além de sobre os fatores motivacionais que podem ser usados.

A *Web 2.0* permite que a gestão do conhecimento seja realizada de forma e em intensidade impossíveis no passado, graças aos seus múltiplos atributos. A gestão de conhecimento tradicional precisou ser completamente readequada para se beneficiar do potencial das novas tecnologias (LEVY, 2009), ainda que exista convergência entre os princípios da *Web 2.0* e da gestão de conhecimento tradicional.

Levy (2009) sugere a aplicação da *Web 2.0* na gestão de conhecimento em mais de uma etapa. A primeira consistiria na aplicação das ferramentas da *Web 2.0*. Em etapas subsequentes seriam tratados aspectos mais profundos, dentre os quais a participação de usuários ativos.

Levy (2009) alerta apenas para um ponto de divergência entre o que se pode conseguir com a *Web 2.0* e aquilo que se pretendia, tradicionalmente, com a gestão do conhecimento. Para ele, a gestão do conhecimento tradicional preconiza a centralização do gerenciamento, requerendo controle, ao passo que, na *Web 2.0*, descentralização e o não controle são características centrais.

#### 2.4 WEB 3.0 OU WEB SEMÂNTICA

A *Web 3.0*, que se refere ao que algumas vezes também é chamada de *Web* semântica, tem como característica ser um conteúdo *web* organizado e dependente de uma padronização para poder ser processável por equipamentos, fazendo uso de *frameworks* (BERNERS-LEE *et al.*, 2001).

Ainda sobre a estrutura da *web* semântica, Berners-Lee *et al.* (2001) enfatizam que ela estava estruturada para o conteúdo relevante de páginas *web*, emergindo desta forma um ambiente para que *softwares* possam realizar algumas tarefas para o usuário, no que diz respeito ao acesso e tratamento de informações.

Ao descrever a origem da *web* semântica, Berners-Lee *et al.* (2001) propõem que não se trata de uma *web* separada, mas de uma extensão da *web* vigente,

enfatizando a importância do significado recebido pela informação, reforçando a possibilidade do trabalho colaborativo entre pessoas e computadores.

Segundo esses autores, para a existência da *web* semântica, alguns passos referentes a informação devem ser executados, os quais estão em andamento, relacionados à forma aplicada na estruturação dos dados. Uma vez em vigência a *Web 3.0*, as máquinas terão uma funcionalidade significativa, a qual consiste no aprimoramento do processamento e “entendimento” dos dados (BERNERS-LEE *et al.*, 2001).

É necessário um tipo de categorização dos dados, para serem consumidos por seres humanos ou para serem consumidos por máquinas. Berners-Lee *et al.* (2001) utilizam o conteúdo mostrado na televisão como um exemplo de dados voltados para seres humanos. Sobre os dados voltados para as máquinas, os autores dão o exemplo de bases de dados.

Sobre os possíveis ganhos proporcionados pela *web* semântica, Berners-Lee *et al.* (2001) enfatizam que ela não se trata de uma “mera ferramenta”, voltada para execução de tarefas individuais. Com o devido planejamento, pode auxiliar no aprimoramento do conhecimento, em geral. A exposição de novos conceitos por atores participantes de uma rede permite sua rápida difusão, universalmente, o que é possível devido à característica de estruturação, em que o trabalho e o aprendizado ocorrem em conjunto, por seres humanos e máquinas.

Miranda, Gualtieri e Coccia (2010) atribuem o dinamismo da *web* semântica à grande integração com ferramentas e aos ganhos na indexação dos documentos, fazendo uso dos *frameworks*, ontologias e metadados, permitindo também a agentes não-humanos o acesso a tais informações. Os autores ainda dão o devido enfoque à possibilidade de interação por parte dos usuários, que podem interagir com a *web* de forma distribuída, possibilitando o acesso à informações e interações com a comunidade global, indiferente do método a ser utilizado e da sua localização, resultando nas informações alinhadas aos devidos propósitos, no momento e lugar certo, fazendo uso do dispositivo certo (MIRANDA, GUALTIERI E COCCIA, 2010).

O papel da semântica é “lidar com problemas de linguagem, interações do usuário e personalização”, ou seja, viabilizar a forma de raciocínio de um ser humano por um computador (DOSZKOCS, 2010, p. 37).

Ao fazer um paralelo com o crescimento da ciência, Hendler (2003) argumenta que o crescimento da ciência moderna leva a um proporcional crescimento da necessidade de colaboração entre cientistas de diferentes áreas de estudos, universidades e disciplinas, o que reforça a tendência colaborativa que se estabelece a partir do uso das modernas tecnologias de informação.

## 2.5 BIOINFORMÁTICA

O termo bioinformática, no sentido literal da palavra, é a junção de *bio*, relacionado à vida, ou uma abreviação de biologia, com a palavra informática. Trata-se, portanto da informática aplicada a estudos da vida, ou em biologia.

A primeira aparição do termo bioinformática se deu em 1970, quando Hesper e Hogeweg (1970) o utilizaram para se referir a estudos informáticos em sistemas bióticos. Hogeweg (1978) defende o seu pioneirismo na utilização do termo bioinformática, em sua publicação juntamente com Bem Hesper nos anos 70, contrariando várias publicações posteriores que apontavam a primeira utilização do termo em 1978. Para o autor, a dificuldade de acesso à primeira publicação de 1970 e o fato de o estudo ter sido escrito em holandês, contribuíram para a dificuldade no reconhecimento da primeira utilização do termo quase uma década antes.

Conforme explanado por Attwood *et al.* (2011), contudo, existe uma dificuldade na definição de bioinformática. Esses autores apresentam estudos e períodos em que as sementes da bioinformática foram semeadas, iniciando com o trabalho de Franklin e Gosling, em meados de 1953, buscando a elucidação da estrutura do DNA e a abordagem oportunista para a sua interpretação, adotada por Watson e Crick, que naquele mesmo ano estavam buscando interpretar o DNA fazendo uso da informática.

Dando continuidade na tentativa do mapeamento da origem da bioinformática, Attwood *et al.* (2011) classificam como inovador o trabalho de determinação das

estruturas tridimensionais de proteínas, realizado por Kendrew, em meados de 1958, e retomado por Muirhead e Pertutz, por volta de 1963.

Esses estudos destacados por Attwood refletem a essência da aplicação da informática para resolver problemas da área da Biologia, conforme preconizado por Hesper e Hogeweg (1970), ao se referirem a estudos informáticos em sistemas bióticos.

Na mesma linha de Attwood *et al.* (2011), Hagen (2000) também é cauteloso na busca por uma definição para bioinformática, admitindo que havia um ambiente convidativo para se associar o termo bioinformática à estudos do sequenciamento do DNA, projeto genoma em larga escala, utilização de supercomputadores e Internet.

Segundo Hagen (2000), cientistas entendem que a bioinformática já estava presente no início do reconhecimento da importância dos supercomputadores como ferramenta fundamental para estudos da biologia molecular uma década antes do sequenciamento molecular. O autor ressalta que, mesmo não utilizando o termo bioinformática, os cientistas da área de biologia tinham uma clara visão da forma como a tecnologia dos computadores, combinada à matemática e à biologia molecular, trazia um suporte interessante na busca das respostas para as perguntas fundamentais das ciências da vida.

Assim, a bioinformática é considerada uma área multidisciplinar da ciência que faz uso de recursos computacionais, físicos, matemáticos e estatísticos com o propósito de classificar, analisar e disponibilizar dados sobre biologia molecular, voltados para o embasamento de novas descobertas sobre organismos vivos (MUNIZ, 2003).

Há duas formas principais de uso da informática na bioinformática: (1) a utilização e apresentação de grandes massas de dados geradas por procedimentos da biologia molecular, e (2) a análise destas informações, por meio de algoritmos e procedimentos computacionais (MUNIZ, 2003; GIBAS; JAMBECK, 2001).

Indo ao encontro da ideia de multidisciplinaridade da bioinformática proposta por Muniz (2003) e da constatação de Hagen (2000) que a bioinformática é um

campo relativamente novo, Mount (2001) reforça as duas características e adiciona a necessidade de processamento de grandes volumes de dados biológicos da área como elementos importantes para a definição de bioinformática, apoiada por ferramentas computacionais, estatísticas e matemáticas para o processamento de informações.

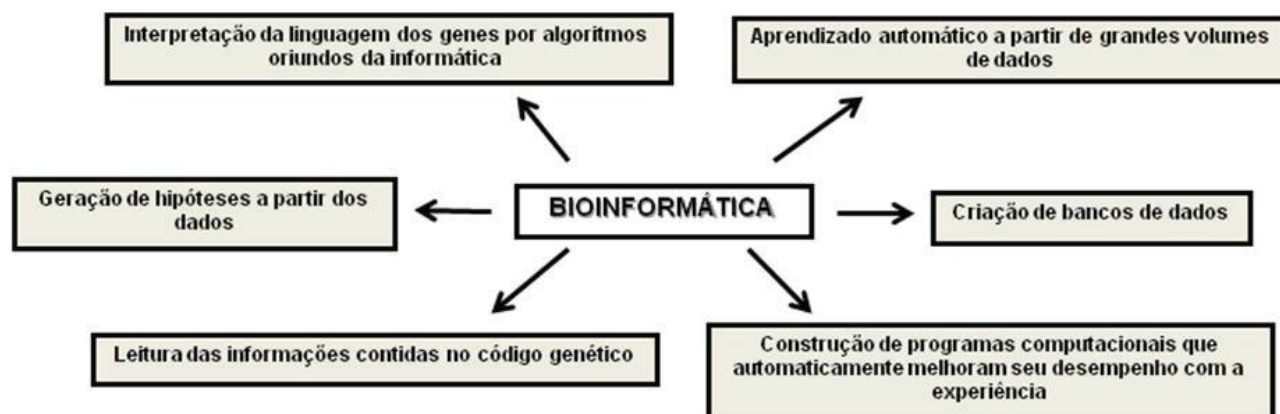
Como os avanços na bioinformática decorrem, primordialmente, dos avanços na área da ciência da computação, Lesk (2002) propõem uma outra ótica para análise da relação entre biologia e informática, atribuindo maior relevância aos aspectos computacionais.

O autor analisou a contribuição da computação enaltecendo algoritmos bem elaborados, estruturas de dados bem definidas, técnicas de recuperação de informação em grandes bases de dados e engenharia da computação, destacando sua importância na análise e resolução de problemas biológicos.

De acordo com Hultquist *et al.* (2002), a bioinformática pode ser considerada uma das responsáveis pelo desenvolvimento de tecnologias inovadoras no campo da biotecnologia, desfrutando de crescimentos anuais no nível de investimentos, em virtude do potencial retorno econômico decorrente das inovações obtidas.

Dentre os vários projetos em que se faz uso da bioinformática, mapeamentos genéticos estão em grande evidência. Para exemplificar a utilização da bioinformática, De Araujo (2008) mostra a associação entre a biologia molecular e a informática (ver a Figura 2).

Figura 2 Associação de conhecimentos da biologia molecular à informática, que culminou com o advento da bioinformática



Fonte: Araújo *et al.* (2008).

Ao abordar mapeamentos genéticos, Saldanha (2012) enfatizou que o projeto genoma contribuiu para uma maior propagação sobre a evolução do sequenciamento genético nos meios comuns, proporcionando maior conhecimento acerca do tema. Ainda sobre este, Saldanha (2012) explicou que o projeto genoma é executado por uma equipe multidisciplinar, composta por biólogos, nos laboratórios de biologia molecular, e cientistas da computação, nos laboratórios de bioinformática.

Ainda sobre a participação de biólogos e cientistas da computação em suas respectivas estruturas laboratoriais, Saldanha (2012) descreve como acontecem as etapas dos estudos multidisciplinares que envolvem esses pesquisadores. A primeira etapa consiste na produção de fragmentos de DNA ou RNA, por meio de sequenciadores, os quais no passo seguinte serão convertidos em sequências de caracteres (*strings*) compostos pelas letras (A, C, G e T ou U), atividades estas executadas pelos biólogos dentro dos laboratórios de biologia molecular.

Realizada essa primeira etapa, as fases subsequentes são realizadas em laboratórios de bioinformática, responsáveis pelo armazenamento, gerenciamento e processamento dos dados. A computação em nuvem é uma alternativa para o tratamento da grande quantidade de dados produzida pelos projetos genoma, tecnologia esta que oferece uma estrutura computacional flexível e adaptável à demanda (SALDANHA, 2012).



Uma vez aplicada a tecnologia da computação em nuvem, que elimina restrições geográficas, ou seja, permite a colaboração de forma desmembrada e conectada, pesquisadores podem implementar ferramentas que atingem elevado paralelismo de maneira simplificada (SALDANHA, 2012).

Devido à grande quantidade de dados gerados pelos projetos genoma, com necessidade de gerenciamento e processamento, a bioinformática agrega o grande poder computacional em ambiente de larga escala altamente colaborativo (JITHESH *et al.*, 2005).

A computação em nuvem, somada à necessidade de ambiente altamente colaborativo, viabiliza a existência de diferentes frentes de trabalho, em diferentes lugares, promovendo a colaboração entre equipes (ROSA *et al.*, 2006).

### 3 METODOLOGIA

Este capítulo é voltado para o detalhamento dos procedimentos metodológicos aplicados na consecução dos objetivos propostos para o estudo.

#### 3.1 ESPECIFICAÇÃO DO PROBLEMA

##### 3.1.1 Apresentação das perguntas ou hipóteses de pesquisa

Vistos os objetivos anteriormente listados, as seguintes questões foram pontuadas:

- Que aspectos da utilização da inteligência coletiva por meio das ferramentas da *Web 2.0* estão sendo discutidos nas publicações científicas identificadas em projetos relacionados à área de bioinformática?

- Quais áreas de estudo estão relacionadas com uso da inteligência coletiva por meio das ferramentas da *Web 2.0* em projetos relacionados à área de bioinformática?

- Quais são as formas de aplicação da inteligência coletiva por meio das ferramentas da *Web 2.0* para a área de bioinformática, identificadas nas publicações científicas?

##### 3.1.2 Apresentação/definição das categorias analíticas ou variáveis

As duas categorias analíticas principais que se pretende relacionar neste estudo são *pesquisa em bioinformática* e *Web 2.0/inteligência coletiva*, buscando-se identificar como a segunda pode ser utilizada para facilitar o avanço da primeira.

#### 3.2 DELIMITAÇÃO E DELINEAMENTO DA PESQUISA

Este tópico é dedicado à apresentação, em detalhe, do desenho da pesquisa, considerando a definição da população e amostra, etapas de pesquisa, delineamento do estudo, perspectiva temporal, nível e unidade de análise,

procedimentos de coleta e análise de dados e, por fim, dificuldades e limitações da pesquisa.

### 3.2.1 População e amostra

O presente estudo tem como *corpus* de análise artigos incluídos em uma revisão sistemática de literatura (RSL) sobre a aplicação da inteligência coletiva por meio de ferramentas da *Web 2.0* para a realização de estudos e desenvolvimentos na área de bioinformática, procurando levantar o estado da arte a este respeito.

### 3.2.2 Delineamento e etapas da pesquisa

A pesquisa fez uso da RSL para a execução de um levantamento bibliográfico, cuja meta é compreender como publicações científicas abordam a aplicação da inteligência coletiva por meio de ferramentas da *Web 2.0* na bioinformática.

Sua visão epistemológica é positivista, uma vez que se tem uma definição construída previamente do objeto de estudo, supondo a utilização de ferramentas da *Web 2.0* no segmento de bioinformática, buscando a compreensão mais profunda do potencial de sua aplicação aos estudos em bioinformática.

A pesquisa relacionada ao trabalho é descritiva. A perspectiva temporal utilizada foi longitudinal, uma vez que se buscou entender a relação entre os constructos durante o período em que ocorreram as publicações. Vale ressaltar que não se teve a intenção de caracterizar as variáveis ao longo do período das publicações.

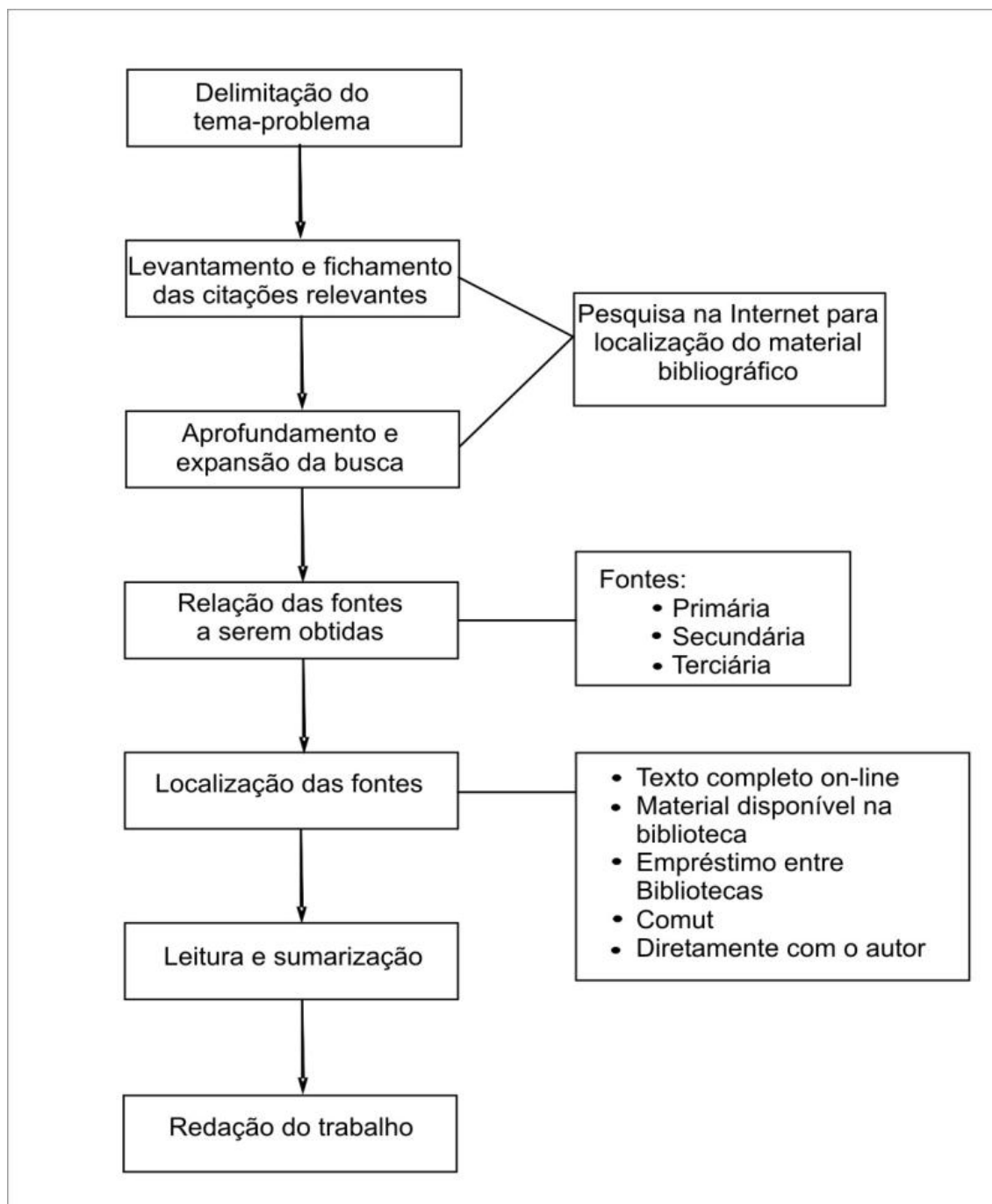
O levantamento bibliográfico foi escolhido com o intuito de sanar o problema de pesquisa fazendo uso de referenciais teóricos publicados, proporcionando uma “discussão” entre as várias contribuições científicas identificadas na literatura (BOCCATO, 2006). Segundo Boccato (2006), é de suma importância realizar um planejamento detalhado para este tipo de pesquisa, o qual consiste na definição da temática, passando pela estruturação lógica da coleta de dados e finalizando com a definição da forma da comunicação dos resultados.

Reforçando a ideia do planejamento para o levantamento bibliográfico, Volpato (2000) preconiza a importância da formulação da temática, que resultará nos termos que expressem o conteúdo e, conseqüentemente, a busca dos estudos a serem utilizados no *corpus* de análise.

Com um enfoque na aplicação de boas técnicas, Salomon (2004) enfatiza o rigor do levantamento bibliográfico, definindo-se, de forma metódica, como se dá a identificação, obtenção e fichamento da informação, além da redação do trabalho científico.

Pizzani *et al.* (2012) sugerem uma forma de executar determinados passos, incluindo o levantamento e fichamento das citações relevantes, aprofundamento e expansão da busca, relação das fontes a serem obtidas e a localização das fontes, conforme pode ser visto na Figura 3.

Figura 3 Etapas de realização de um levantamento bibliográfico



Fonte: Pizzani *et al.* (2012).

### 3.2.3 Procedimentos de coleta de dados

Foi realizada uma RSL, para levantar os tópicos que estão sendo discutidos em publicações científicas envolvendo *Web 2.0* e bioinformática.

A RSL foi escolhida como opção para o levantamento preliminar de dados, devido ao rigor da sua aplicação como estratégia de levantamento e análise de dados, conforme salientado por Botelho *et al.* (2011). A RSL é eficaz na identificação, interpretação e avaliação de estudos relevantes acerca de um tópico, com a característica de ser fidedigna, confiável, passível de reprodução e auditoria (KITCHENHAM, 2004).

Foi adotado o protocolo de RSL criado por Kitchenham (2004), tendo como procedimento preliminar a busca de artigos em meio eletrônico, a qual resultou na seleção dos artigos que compuseram o *corpus* da pesquisa e foram, posteriormente, analisados.

Sobre a busca dos artigos por meio eletrônico, Severino (2002) reforça a ideia de que a Internet é hoje indispensável, devido a possibilidade de acesso a um excelente acervo com facilidade.

Ao abordar os repositórios utilizados para a busca dos artigos, Cunha (2001) sugere o acesso a bases de dados bibliográficos, uma vez que tais bases têm como característica a qualidade nas informações nelas contidas. Inúmeros repositórios eletrônicos estão hoje disponíveis, incluindo bibliotecas virtuais de universidades, congressos entre outros. Dentre os canais voltados para a publicação de produção científica é importante focar em revistas e jornais de cunho científico.

O *Google Scholar* foi escolhido como base de dados para a busca e seleção dos artigos a serem utilizados na pesquisa. A base se mostra interessante devido à quantidade de trabalhos disponíveis e à possibilidade de parametrização de seleção excluindo citações e patentes, possibilitando foco à pesquisa, visto que a ferramenta realiza a busca de palavras chave pré-definidas em diferentes partes do trabalho (JACSO, 2005).

Padilha e Graeml (2015) e Vianna, Graeml e Peinado (2017), inspirados no protocolo proposto por Kitchenham (2004) para a realização de RSL, recomendam que os seguintes procedimentos sejam adotados ao definir o *corpus* da pesquisa a partir de busca no *Google Scholar*:

- Desmarcar as opções “incluir citações” e “incluir patentes”;
- Aceitar resultados, preliminarmente, em qualquer idioma;
- Não aplicar período de tempo das publicações, salvo se houver motivo para estabelecer uma restrição temporal;
- Considerar para inclusão no *corpus* de pesquisa, a depender de serem satisfeitos outros critérios posteriores, artigos que contenham ao menos uma das expressões usadas como critério de filtragem no título, nas palavras-chave ou no resumo do artigo;
- Analisar os resultados fornecidos pelo *Google Scholar* nas páginas de respostas à busca até se deparar com duas páginas inteiras sem resultados relevantes para o propósito da pesquisa;
- Focar exclusivamente em artigos publicados em jornais, revistas científicas e capítulos de livros, excluindo artigos apresentados em congressos, teses e dissertações;
- Ler os resumos dos artigos ou introdução dos capítulos pré-selecionados e descartar aqueles que indiquem que o assunto tratado é distinto do que se pretende analisar;
- Realizar a leitura integral dos artigos ou capítulos de livros remanescentes visando à confirmação de que o conteúdo atende os interesses da RSL.

Em um primeiro momento buscou-se analisar a relação entre *biotecnologia* e *inteligência coletiva*. Contudo, as buscas revelaram que nos artigos com discussões próximas à temática, os pesquisadores pouco faziam uso do termo *biotecnologia*, expondo estudos envolvendo segmentos da biotecnologia com maior delimitação e fazendo uso de termos mais específicos.

A busca resultou em 42 artigos dentre 170 entradas. Após a aplicação do procedimento executado por Padilha e Graeml (2015) e Vianna, Graeml e Peinado

(2017), inspirados no protocolo proposto por Kitchenham (2004), resultaram apenas 7 artigos que tinham uma relação aproximada sobre a temática.

Dentre esses poucos artigos resultantes da busca, deparou-se com a citação de um documento da OECD (um *framework* para estatísticas de biotecnologia) que apresentava uma categorização das técnicas da biotecnologia, documento que foi então utilizado como referência para a redefinição dos termos de busca utilizados para avançar com a definição do *corpus* da pesquisa aqui proposta.

Segundo o documento da OECD (2005), as técnicas de biotecnologia estão distribuídas da seguinte forma:

- DNA/RNA;
- Proteínas e outras moléculas;
- Cultura e engenharia de células e tecidos;
- Técnicas de processamento biotecnológico;
- Genes e vetores RNA;
- Bioinformática;
- Nano biotecnologia.

A partir dos segmentos acima listados, foram extraídos alguns termos que foram combinados com o termo “*Web 2.0*” e realizadas as pesquisas no *Google Scholar*, das quais resultaram as quantidades de retornos apresentadas na Tabela 1.

Tabela 1 Quantidades de retornos para buscas no Google Scholar

<b>Combinações utilizadas</b>	<b>Resultados</b>
"DNA" e "Web 2.0"	9780
"RNA" e "Web 2.0"	1670
"Proteins" e "Web 2.0"	4500
"Cell culture" e "Web 2.0"	204



"Tissue culture" e "Web 2.0"	109
Biotechnology techniques	7
"Gene" e "Web 2.0"	18000
"RNA vectors" e "Web 2.0"	1
"Bioinformatics" e "Web 2.0"	5630
"Nanobiotechnology" e "Web 2.0"	67

Fonte: elaborada pelo autor.

Foram realizadas análises preliminares nas combinações com maiores quantidades de resultados, sendo que as duas primeiras ("*DNA*" e "*Web 2.0*" e "*Gene*" e "*Web 2.0*") não apresentaram o devido enfoque com inteligência coletiva, que apareceu como temática acessória ou sequer foi mencionada.

Acabou-se optando por analisar mais profundamente os resultados da busca por "*bioinformatics*" e "*Web 2.0*" devido ao fato de "*bioinformatics*" ser um termo mais abrangente e a combinação ter apresentado a terceira maior quantidade de resultados gerais. Além disso vários estudos, mesmo a partir de uma análise preliminar, já se demonstravam voltados à discussão do uso de inteligência coletiva para o avanço das pesquisas nas áreas tratadas. Por fim, outro fator que contribuiu para a escolha da utilização do termo "*bioinformatics*" como expressão de busca foi a sua ligação com a tecnologia da informação, possivelmente levando os pesquisadores envolvidos com a área a serem "*early adopters*" da "*Web 2.0*" se comparados a outros pesquisadores da área de biologia.

Feita a definição dos termos a serem utilizados, a busca no *Google Scholar* se deu na segunda quinzena de março de 2019, envolvendo as expressões "*bioinformatics*" e "*Web 2.0*".

Na parametrização dos filtros de exibição da ferramenta de busca foram eliminadas patentes e citações, sem qualquer filtro de relevância de exibição. O *Google Scholar* apresentou 5460 entradas.

Seguindo as recomendações mencionadas acima, a busca foi encerrada na página 8, uma vez que tanto na página 8 quanto na página 7 não foram identificados artigos relevantes para o propósito da pesquisa. Logo, foram analisadas, de forma preliminar, 80 entradas (eram 10 por página).

Foi montada uma tabela contendo informação sobre o tipo de trabalho representado por cada uma das 80 entradas analisadas preliminarmente (ver a Tabela 2). Durante a coleta de dados, nenhuma das 80 entradas contidas nas 8 páginas analisadas de resultados do *Google Scholar* estava inacessível ou bloqueada. Foi possível obter o conteúdo integral de todas as fontes diretamente dessa base de dados *online*, sem a necessidade de se lançar mão de outras bases ou repositórios, complementarmente. Isto permitiu que a análise preliminar do conteúdo de cada artigo ocorresse conforme descrito no detalhamento do protocolo utilizado, sem qualquer restrição.

Tabela 2 Classificação dos artigos contidos nas páginas de busca utilizando os termos de pesquisa “*Bioinformatics*” e “*Web 2.0*”

Nº página Google Scholar	Artigos em periódico	Capítulos de livro	Artigos em anais de congresso	Dissertações ou teses	Outros
1	6				4
2	5	1		1	3
3	1		1		8
4	3				7
5	2		2		6
6	3				7
7					10
8					10
Total	20	1	3	1	55

Fonte: elaborada pelo autor.

Dentre os trabalhos selecionados na primeira busca, o termo *Web 3.0* surgiu algumas vezes, onde os autores mesmo que sem fazer uma abordagem direta, indicavam a *Web 3.0* como uma continuidade da *Web 2.0*.

Considerando-se isso, decidiu-se realizar uma nova busca no *Google Scholar*, fazendo uso do mesmo protocolo utilizado anteriormente, mas com as expressões “*bioinformatics*” e “*Web 3.0*”. A busca se deu em meados de julho de 2019.

O *Google Scholar* apresentou 831 entradas. Seguindo as mesmas recomendações e procedimentos adotados para a busca anterior, a segunda busca foi encerrada na página 6, uma vez que tanto na página 5 quanto na página 6 não foram identificados novos artigos ou capítulos de livro relevantes para o propósito da pesquisa. Logo, foram analisadas, de forma preliminar, sessenta entradas (eram 10 por página).

Essa segunda busca resultou em 12 artigos que atendiam aos requisitos da pesquisa, dentre os quais, 9 já tinham sido selecionados por meio do procedimento anterior de busca, resultando desta forma 3 artigos novos para a composição da RSL.

Foi montada uma tabela contendo informação sobre o tipo de trabalho representado por cada uma das 60 entradas analisadas previamente (ver a Tabela 3). Durante a coleta de dados nenhuma das 60 entradas contidas nas 6 páginas de resultados analisadas do *Google Scholar* estava inacessível ou bloqueada. Foi possível obter o conteúdo integral de todas as fontes diretamente dessa base de dados *online*, sem a necessidade de se lançar mão de outras bases ou repositórios, complementarmente. Isto permitiu que a análise preliminar do conteúdo de cada artigo que ainda não tinha aparecido na busca anterior ocorresse conforme descrito no detalhamento do protocolo utilizado, sem qualquer restrição.

Tabela 3 Classificação dos artigos contidos nas páginas de busca utilizando os termos de pesquisa “*Bioinformatics*” e “*Web 3.0*”

Nº página <i>Google Scholar</i>	Artigos novos em periódicos	Artigos repetidos	Capítulos de livros	Artigos em anais de congresso	Dissertações ou teses	Outros
1		5				5
2	1	2				7
3	1	1		1		7
4	1	1		1		7
5				3		7
6				2		8
Total	3	9	0	7	0	41

Fonte: elaborada pelo autor.

Em poder dos 20 artigos e do capítulo de livro resultantes da primeira busca (vide Tabela 2) e dos 3 artigos resultantes da segunda busca (vide Tabela 3), totalizando desta forma 24 trabalhos, foi realizada a análise por meio da leitura completa, a fim de confirmar se atendiam aos requisitos do estudo para comporem o *corpus* da RSL. Após a análise completa dos trabalhos, constatou-se que 2 artigos não tratavam da relação entre a *Web 2.0/Web 3.0* e a bioinformática, de modo que foram eliminados.

Depois da eliminação dos 2 artigos mencionados acima, que não contemplavam a proposta do estudo, restaram 22 trabalhos, os quais compuseram o *corpus* definitivo da RSL.

#### 3.2.4 Procedimentos de tratamento e análise dos dados

Na confecção do estudo, foi realizado um mapeamento dos 22 trabalhos que compuseram o *corpus* da RSL, classificando-os de acordo com uma categoria principal e uma secundária.

Foram identificadas 6 categorias, que serviram de agrupadoras como categoria principal ou categoria secundária, dentro de cada artigo. Essas categorias emergiram da análise do próprio *corpus* de análise, ou seja, não havia qualquer predefinição de quais deveriam ser. À medida que uma característica ou argumento relacionado à relação entre *Web 2.0* e *bioinformatics* aparecia em algum dos trabalhos analisados, era feita uma anotação, para ver se ela se repetiria em outros estudos. Como diferentes autores usaram termos distintos, em alguns casos, para tratar das mesmas coisas, houve um esforço de agrupamento de ideias durante todo o processo de categorização.

A categoria principal foi a relacionada à temática central do artigo e a categoria secundária envolveu temática acessória, relativa a uma premissa, um resultado ou um desdobramento da temática central.

A título de ilustração, e para aumentar a transparência e segurança da classificação realizada, foi separado um ou mais trechos de cada trabalho que tenha sido usado para justificar sua classificação em uma determinada categoria (ver Quadro 5). Vale ressaltar que a terminologia utilizada para nomear as 6 categorias se deu com base na aproximação dos termos utilizados pelos autores.

#### 3.2.5 Facilidades e dificuldades na coleta e tratamento dos dados

Os resultados e as inferências obtidos a partir da categorização dos trabalhos precisam ser considerados com alguma cautela, uma vez que os termos e enfoques adotados nos diferentes trabalhos podem não representar o entendimento de toda a

comunidade de pesquisadores da área de bioinformática ou podem não ter sido precisamente capturados, nas ideias que pretendiam transmitir, pelo processo de categorização.

### 3.2.6 Limitações da pesquisa

Conforme citado anteriormente, as pesquisas selecionadas para o *corpus* da RSL em sua totalidade estavam disponíveis para livre acesso. Como todos os trabalhos estavam no idioma inglês, no momento da leitura e para o embasamento da categorização, foi realizada a tradução dos trechos para o português. Desta conversão pode decorrer alguma interpretação diferente daquela pretendida pelo autor no idioma original do artigo, ponto este que pode ser considerado uma limitação da pesquisa.

Outra limitação do estudo é que partiu de expressões de busca previamente definidas (“*Web 2.0*” / “*Web 3.0*” e “*bioinformatics*”) e não explorou outras variações que possam ter sido utilizadas por autores para tratar dos mesmos assuntos. Assim, existe a possibilidade de que uma parte dos estudos que poderiam ser incluídos no *corpus* tenham escapado à análise.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

### 4.1 APRESENTAÇÃO DOS RESULTADOS

#### 4.1.1 Artigos selecionados para o *corpus*

Conforme visto anteriormente, nos procedimentos metodológicos, após a aplicação de alguns filtros, foram selecionados 22 trabalhos para a composição do *corpus* final de análise. No Quadro 1 estão listados os artigos que compuseram o *corpus* de análise e seus respectivos autores.

Quadro 1 Artigos selecionados para a RSL

	<b>Artigo</b>	<b>Autor(es)</b>	<b>Ano</b>
1	<i>Bringing Web 2.0 to bioinformatics</i>	Zhang Zhang, Kei-Hoi Cheung e Jeffrey P. Townsend	2008
2	<i>HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0</i>	Kei-Hoi Cheung, Kevin Y. Yip, Jeffrey P. Townsend e Matthew Scotch	2008
3	<i>State of the nation in data integration for bioinformatics</i>	Carole Goble e Robert Stevens	2008
4	<i>Conducting research on the web: 2007 update for the bioinformatics links directory</i>	Joanne A. Fox, Scott McMillan e B. F. Francis Ouellette	2007
5	<i>A Dictyostelium bioinformatics resource update</i>	Petra Fey, Pascale Gaudet, Tomaz Curk, Blaz Zupan, Eric M. Just, Siddhartha Basu, Sohel N. Merchant, Yulia A. Bushmanova, Gad Shaulsky, Warren A. Kibbe e Rex L. Chisholm.	2008
6	<i>BioCatalogue: a universal catalogue of web services for the life sciences</i>	Jiten Bhagat, Franck Tanoh, Eric Nzuobontane, Thomas Laurent, Jerzy Orłowski, Marco Roos, Katy Wolstencroft, Sergejs Aleksejevs, Robert Stevens, Steve Pettifer, Rodrigo Lopez and Carole A. Goble.	2010
7	<i>Web 3.0 and medicine</i>	Dean Giustini	2007
8	<i>Semantic Web for Health Care and Life Sciences: a review of the state of the art</i>	Kei-Hoi Cheung, Eric Prud'hommeaux, Yimin Wang and Susie Stephens.	2009
9	<i>At the intersection of public-health informatics and bioinformatics: using advanced web technologies for phylogeography.</i>	Matthew Scotch, Changjiang Mei, Cynthia Brandt, Indra Neil Sarkar e Kei Cheung	2010
10	<i>Data integration in bioinformatics: current efforts and challenges</i>	Zhang Zhang, Vladimir B. Bajic, Jun Yu, Kei-Hoi Cheung e Jeffrey P. Townsend	2011

11	<i>Tools and collaborative environments for bioinformatics research</i>	Paolo Romano, Rosalba Giugno e Alfredo Pulvirenti	2011
12	<i>Biomashups: the new world of exploratory bioinformatics?</i>	James M. Hogan, Jiro Sumitomo, Paul Roe e Felicity Newell.	2011
13	<i>Experience using web services for biological sequence analysis</i>	Heinz Stockinger, Teresa Attwood, Shahid Nadeem Chohan, Richard Côté, Philippe Cudré-Mauroux, Laurent Falquet, Pedro Fernandes, Robert D. Finn, Taavi Hupponen, Eija Korpelainen, Alberto Labarga, Aurelie Laugraud, Tania Lima, Evangelos Pafilis, Marco Pagni, Steve Pettifer, Isabelle Phan e Nazim Rahman	2008
14	<i>Biological knowledge management: the emerging role of the Semantic web technologies.</i>	Erick Antezana, Martin Kuiper e Vladimir Mironov	2009
15	<i>Building biomedical web communities using a semantically aware content management system</i>	Sudeshna Das, Lisa Girard, Tom Green, Louis Weitzman, Alister Lewis-Bowen e Tim Clark	2008
16	<i>ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data</i>	Vincent Gardeux, Fabrice P. A. David, Adrian Shajkofci, Petra C. Schwalie e Bart Deplancke.	2017
17	<i>BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature</i>	Guillermo de la Calle, Miguel García-Remesal, Stefano Chiesa, Diana de la Iglesia e Victor Maojo.	2009
18	<i>The evolution of the web and implications for eResearch</i>	Wendy Hall, David de Roure e Nigel Shadbolt.	2008
19	<i>Web 3.0 and health librarians: an introduction</i>	Allan Cho e Dean Giustini	2008
20	<i>Neuropsychology 3.0: Evidence-based science and practice</i>	Robert M. Bilder	2011
21	<i>A semantic web management model for integrative biomedical informatics</i>	Helena F. Deus, Romesh Stanislaus, Diogo F. Veiga, Carmen Behrens, Ignacio I. Wistuba, John D. Minna, Harold R. Garner, Stephen G. Swisher, Jack A. Roth, Arlene M. Correa, Bradley Broom, Kevin Coombes, Allen Chang, Lynn H. Vogel e Jonas S. Almeida	2008
22	<i>Semantic web may be cancer information's next step forward</i>	Mike Martin	2011

Fonte: elaborado pelo autor.

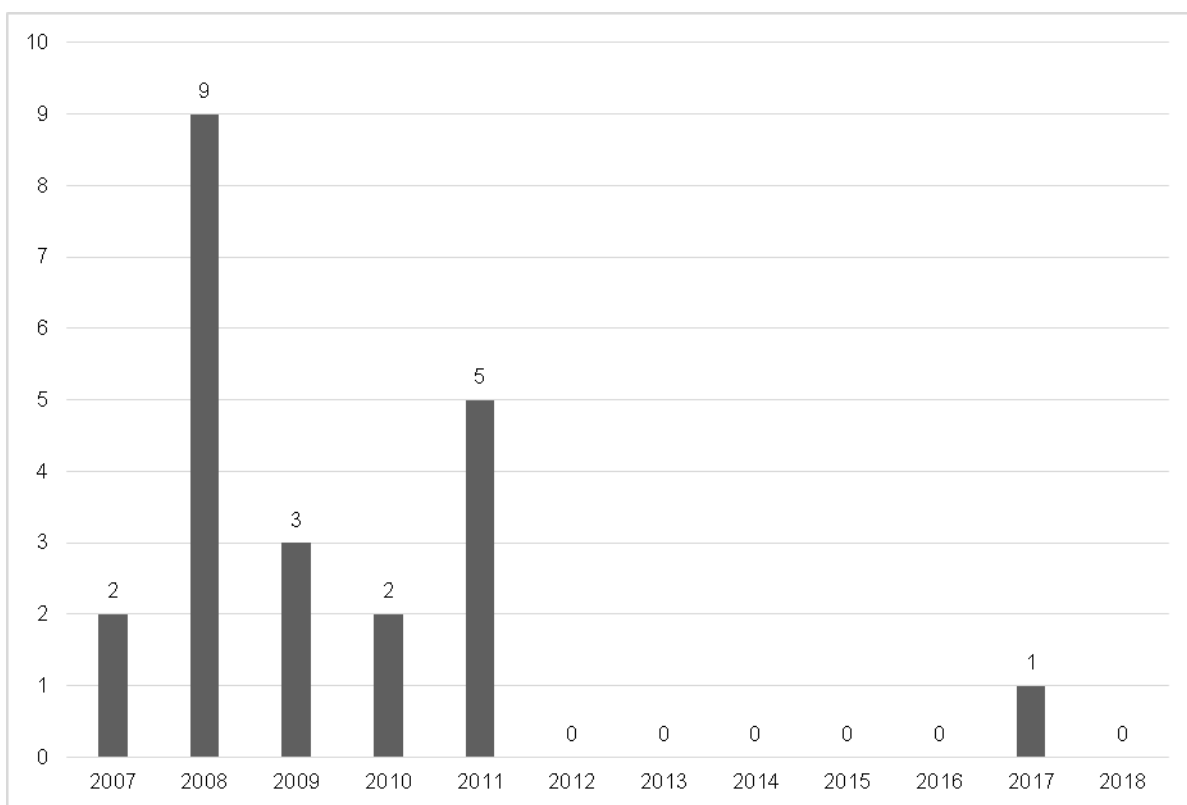
#### 4.1.2 Temporalidade

Ao realizar a busca dos trabalhos no *Google Scholar* a serem utilizados na RSL, não foi aplicado filtro de data, permitindo que trabalhos de qualquer ano de

publicação fossem contemplados pela busca. Os trabalhos selecionados foram publicados entre o período de 2007 e 2017.

Constatou-se que no período de 2012 a 2016 não houve trabalhos que tratassem do assunto da pesquisa e ao mesmo atendessem os requisitos para serem incluídos nesta RSL. A maior concentração de trabalhos se deu de 2007 e 2009, provavelmente por se tratar de período em que a área estava compreendendo os impactos de uma *web* mais interativa (*Web 2.0*) sobre a forma de realizar ciência. Também é digno de nota o fato de que o tema deixou de despertar o mesmo nível interesse da primeira década deste século. Na Figura 4 é possível ver o número de publicações e seus respectivos anos.

Figura 4 Número de trabalhos publicados x ano de publicação



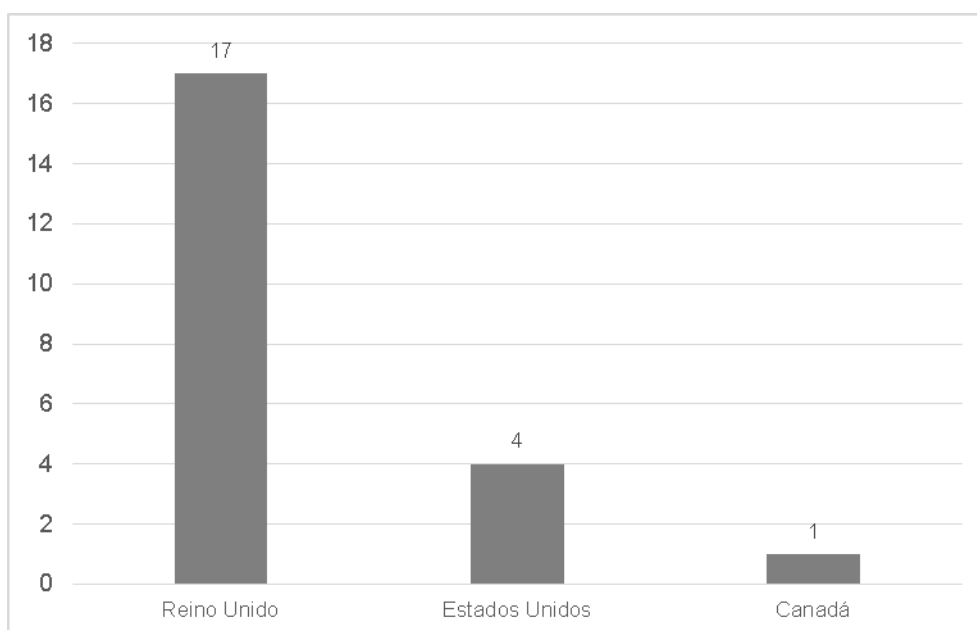
Fonte: elaborada pelo autor.



#### 4.1.3 Instituições e países envolvidos na pesquisa sobre *Web 2.0* na bioinformática

Visando a mapear os países onde se tem publicado a respeito do uso da *Web 2.0* e inteligência coletiva na bioinformática, constatou-se que os periódicos onde ocorreram as publicações estão situados em 3 localidades (Reino Unido, Estados Unidos e Canadá), conforme é possível ver na Figura 5.

Figura 5 Número de trabalhos publicados x países de origem dos periódicos



Fonte: elaborada pelo autor.

Dezessete publicações ocorreram em periódicos britânicos, seguidas de 4 trabalhos publicados em periódicos dos Estados Unidos e um do Canadá.

Embora essa informação tenha alguma relevância. Talvez mais importante seja identificar a origem dos pesquisadores que estiveram discutindo o tema, já que muitos dos periódicos, notoriamente, são publicados por editoras e universidades de países desenvolvidos. A fim de identificar os locais de discussões sobre a temática desta pesquisa, buscou-se, então, mapear as universidades e centros de pesquisa a que estão afiliados os pesquisadores, além dos departamentos de que fazem parte, a fim de entender a multidisciplinaridade em torno da temática.

Iniciou-se o mapeamento identificando os autores dos trabalhos selecionados para o *corpus* da RSL e suas respectivas universidades, centros de pesquisa e departamentos. Foram identificados 107 autores para os 22 artigos contidos no *corpus* do estudo, distribuídos em 36 universidades ou outras organizações diversas. Feito o devido mapeamento, buscou-se mensurar a participação das universidades nos trabalhos, resultando na quantidade de artigos relacionada a cada universidade e/ou centro de pesquisa, com base no número de autores afiliados, conforme pode ser visto no Quadro 2.

Quadro 2 Universidades e/ou centros de pesquisas x quantidade artigos

<b>Universidades e/ou centro de pesquisa</b>	<b>Quantidade artigos</b>
<i>Yale University</i>	5
<i>University of Manchester</i>	3
<i>University of British Columbia</i>	3
<i>Swiss Institute of Bioinformatics</i>	2
<i>European Bioinformatics Institute</i>	2
Outros	2
<i>The University of Texas M.D. Anderson Cancer Center</i>	1
<i>University of Ljubljana</i>	1
<i>Universidad Politécnica de Madrid</i>	1
<i>Harvard University</i>	1
<i>University of Southampton</i>	1
<i>Northwestern University Biomedical Informatics Center</i>	1
<i>Norwegian University</i>	1
<i>University of Texas Southwestern Medical Center</i>	1
<i>Baylor College of Medicine</i>	1
<i>Broad Institute of MIT and Harvard</i>	1
<i>Catholic Bolivian University</i>	1
<i>Chinese Academy of Sciences</i>	1
<i>Discovery IT, Eli Lilly</i>	1
<i>Lilly Singapore</i>	1
<i>EMBL in Heidelberg</i>	1
<i>Gulbenkian Institute</i>	1
<i>Institute of Information Technology</i>	1
<i>International Institute of Molecular and Cell Biology</i>	1
<i>Jane and Terry Semel Institute for Neuroscience &amp; Human Behavior at UCLA</i>	1
<i>King Abdullah University of Science and Technology (KAUST)</i>	1
<i>MIT</i>	1
<i>National Cancer Research Institute (IST)</i>	1
<i>Queensland University of Technology</i>	1

<i>Reference Librarian, Humanities and Social Sciences</i>	1
<i>The University of Zurich</i>	1
<i>University of Amsterdam</i>	1
<i>University of Catania</i>	1
<i>University of Maryland and New York University</i>	1
<i>University of Vermont</i>	1
<i>World Wide Web Consortium</i>	1

Fonte: elaborado pelo autor.

Conforme os dados apresentados na Tabela 4, foi possível identificar 5 universidades e centros de pesquisas cujos pesquisadores participaram de 2 ou mais trabalhos incluídos no *corpus* da pesquisa, participação esta que, em alguns casos se restringiu a um único pesquisador. Além das universidades e centros de pesquisas, foi possível identificar outras organizações que não são da área acadêmica, mas tiveram autores participando em 2 trabalhos.

Mapeadas as universidades e centros de pesquisas, buscou-se identificar as universidades e centros de pesquisas mais atuantes. Foram evidenciados os países das universidades e centros de pesquisas, a fim de entender onde aconteceu a discussão acerca da temática da pesquisa, o que é apresentado no Quadro 3.

Quadro 3 Universidades e centros de pesquisa em que foram realizados os trabalhos incluídos no *corpus* da pesquisa

<b>Artigo</b>	<b>Universidade</b>	<b>País</b>
<i>Bringing Web 2.0 to bioinformatics</i>	<i>Yale University</i>	EUA
<i>HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0</i>	<i>Yale University</i>	EUA
<i>State of the nation in data integration for bioinformatics</i>	<i>University of Manchester</i>	Reino Unido
<i>Conducting research on the web: 2007 update for the bioinformatics links directory</i>	<i>University of British Columbia</i>	EUA
<i>A Dictyostelium bioinformatics resource update</i>	<i>Northwestern University Biomedical Informatics Center</i>	EUA
	<i>University of Ljubljana, Slovenia</i>	Eslovênia
	<i>Baylor College of Medicine, Houston</i>	EUA
<i>BioCatalogue: a universal catalogue of web services for the life sciences</i>	<i>University of Manchester</i>	Reino Unido
	<i>European Bioinformatics Institute, Hinxton, Cambridge</i>	Reino Unido
	<i>International Institute of Molecular and Cell Biology in Warsaw</i>	Polônia
	<i>University of Amsterdam</i>	Bélgica

	<i>European Bioinformatics Institute, Hinxton, Cambridge</i>	Reino Unido
<i>Web 3.0 and medicine</i>	<i>University of British Columbia</i>	EUA
<i>Semantic Web for Health Care and Life Sciences: a review of the state of the art</i>	<i>Yale University</i>	EUA
	<i>WorldWideWeb Consortium</i>	EUA
	<i>Lilly Singapore</i>	Cingapura
	<i>Discovery IT, Eli Lilly, Indianapolis</i>	EUA
<i>At the intersection of public-health informatics and bioinformatics: using advanced web technologies for phylogeography.</i>	<i>Yale University</i>	EUA
	<i>University of Vermont</i>	EUA
<i>Data integration in bioinformatics: current efforts and challenges</i>	<i>Yale University</i>	EUA
	<i>King Abdullah University of Science and Technology (KAUST)</i>	Arábia Saudita
	<i>Chinese Academy of Sciences, Beijing</i>	China
<i>Tools and collaborative environments for bioinformatics research</i>	<i>National Cancer Research Institute (IST)</i>	Itália
	<i>University of Maryland and New York University</i>	EUA
	<i>University of Catania</i>	Itália
<i>Biomashups: the new world of exploratory bioinformatics?</i>	<i>Queensland University of Technology</i>	Austrália
<i>Experience using web services for biological sequence analysis</i>	<i>Swiss Institute of Bioinformatics</i>	Suíça
	<i>University of Manchester</i>	Reino Unido
	<i>Institute of Information Technology, Islamabad</i>	Paquistão
	<i>European Bioinformatics Institute, Hinxton, Cambridge</i>	Reino Unido
	<i>MIT</i>	EUA
	<i>Gulbenkian Institute</i>	Portugal
	<i>EMBL in Heidelberg</i>	Alemanha
<i>Biological knowledge management: the emerging role of the semantic web technologies.</i>	<i>Catholic Bolivian University</i>	Bolívia
	<i>Norwegian University</i>	Noruega
<i>Building biomedical web communities using a semantically aware content management system</i>	<i>Harvard University</i>	EUA
	<i>Broad Institute of MIT and Harvard</i>	EUA
	<i>Staff Engineer at VMware</i>	EUA
<i>ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data</i>	<i>Swiss Institute of Bioinformatics</i>	Suíça
<i>BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature</i>	<i>Universidad Politécnica de Madrid</i>	Espanha
<i>The evolution of the web and implications for eResearch</i>	<i>University of Southampton</i>	Reino Unido
<i>Web 3.0 and health librarians: an introduction</i>	<i>Reference Librarian, Humanities and Social Sciences</i>	EUA

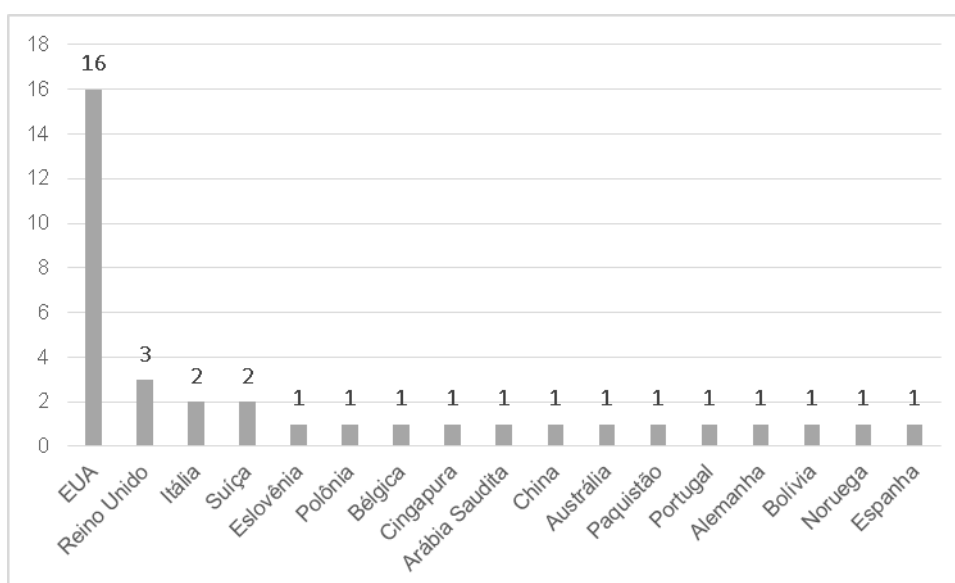
	<i>University of British Columbia</i>	EUA
<i>Neuropsychology 3.0: Evidence-based science and practice</i>	<i>Jane and Terry Semel Institute for Neuroscience &amp; Human Behavior at UCLA</i>	EUA
<i>A semantic web management model for integrative biomedical informatics</i>	<i>The University of Texas M.D. Anderson Cancer Center</i>	EUA
	<i>University of Texas Southwestern Medical Center</i>	EUA
<i>Semantic web may be cancer information's next step forward</i>	<i>The University of Zurich</i>	Suíça

Fonte: elaborado pelo autor.

Foi possível constatar, em alguns trabalhos, a participação de pesquisadores pertencentes a diferentes instituições, trabalhando em conjunto no mesmo artigo. Em alguns casos, foram identificadas diferentes instituições do mesmo país. Em outros casos, constatou-se a participação de pesquisadores de universidades e centros de pesquisas situados em diferentes países, participando do mesmo trabalho.

A Figura 6 apresenta a consolidação do número de autores por país. Observou-se que os EUA detiveram o maior número de instituições envolvidas na discussão de bioinformática e inteligência coletiva, totalizando 16 instituições. As diferentes universidades e centros de pesquisas dos outros países que tiveram participação nos trabalhos totalizaram 20 instituições, das quais 14 estão localizadas na Europa.

Figura 6 Número de universidades x países

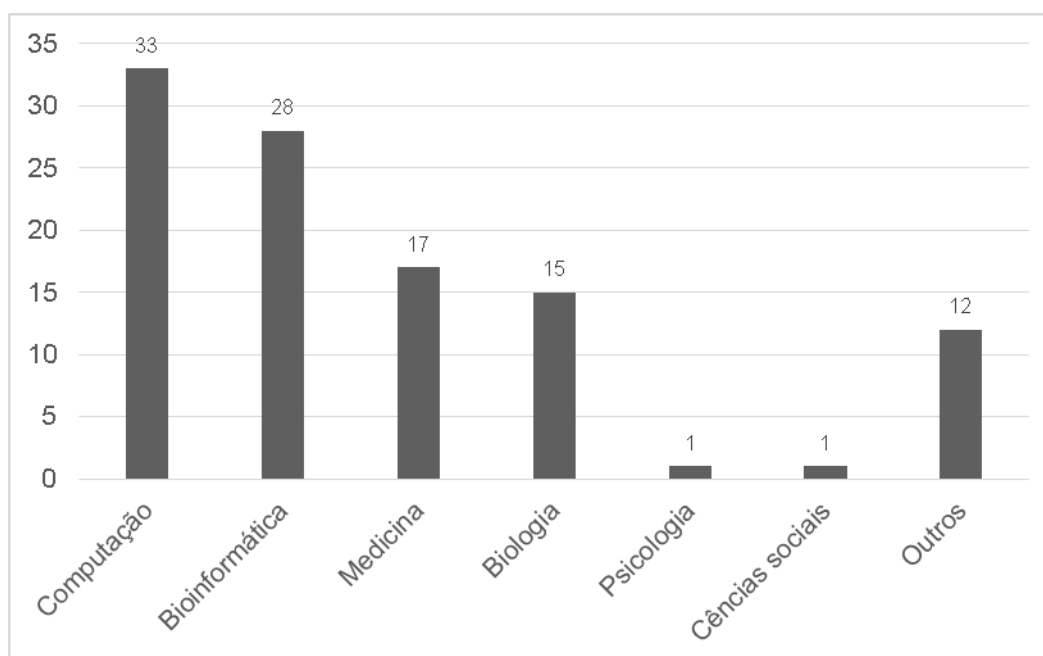


Fonte: elaborada pelo autor.

Depois de realizado o mapeamento das universidades e/ou centros de pesquisa, procurou-se identificar os departamentos a que os pesquisadores estavam vinculados. Considerou-se isso importante para se saber se o vínculo era com departamentos de informática, de biologia, ou de outras áreas.

Foram classificados os departamentos dos quais os autores fazem parte. Foram identificados e classificados os departamentos em sete áreas “*Computação*”, “*Bioinformática*”, “*Medicina*”, “*Biologia*”, “*Psicologia*”, “*Ciências sociais*” e “*Outros*”, conforme é possível ver na Figura 7.

Figura 7 Departamentos x quantidade de pesquisadores



Fonte: elaborada pelo autor.

A partir da Figura 7, é possível identificar que a maior parte dos autores envolvidos nos trabalhos que compuseram a RSL fazem parte de departamentos de “*Computação*” ou “*Bioinformática*”, totalizando 61 autores, dentre 107, o que representa 57% do total.

Departamentos de “*Medicina*”, “*Biologia*”, “*Psicologia*” e “*Ciências sociais*” contribuíram com trinta e quatro autores, totalizando, aproximadamente, 32% do total.

A classificação “*outros*” contém autores que não necessariamente fazem parte de uma universidade e/ou centro de pesquisa, totalizando, aproximadamente, 11% do total. Tais autores fazem parte de empresas privadas de segmento comercial ou de organizações que não possuem em seu *core* a pesquisa científica.

Conforme os números mostrados na Figura 7, é possível identificar que a maior parte das discussões tem participação de autores que estão ligados à “*Computação*” e à “*Bioinformática*”. Contudo, “*Medicina*” e “*Biologia*” também apresentam números relevantes, somando elementos para a discussão em torno da *Web 2.0* na bioinformática.

#### 4.1.4 Referências seminais dos trabalhos do *corpus* da pesquisa

Ao analisar os 22 trabalhos utilizados no *corpus*, procurou-se mapear as referências bibliográficas utilizadas. Foram encontradas 888 referências bibliográficas, conforme apresentado na Tabela 4.

Tabela 4 Trabalhos *corpus* x quantidade de referências bibliográficas apresentadas

	<b>Artigo</b>	<b>Quantidade de referências</b>
1	<i>Bringing Web 2.0 to bioinformatics</i>	73
2	<i>HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0</i>	20
3	<i>State of the nation in data integration for bioinformatics</i>	58
4	<i>Conducting research on the web: 2007 update for the bioinformatics links directory</i>	5
5	<i>a Dictyostelium bioinformatics resource update</i>	21
6	<i>BioCatalogue: a universal catalogue of web services for the life sciences</i>	24
7	<i>Web 3.0 and medicine</i>	12
8	<i>Semantic Web for Health Care and Life Sciences: a review of the state of the art</i>	3
9	<i>At the intersection of public-health informatics and bioinformatics: using advanced web technologies for phylogeography.</i>	18
10	<i>Data integration in bioinformatics: current efforts and challenges</i>	128
11	<i>Tools and collaborative environments for bioinformatics research</i>	52
12	<i>Biomashups: the new world of exploratory bioinformatics?</i>	31
13	<i>Experience using web services for biological sequence analysis</i>	32
14	<i>Biological knowledge management: the emerging role of the semantic web technologies.</i>	181

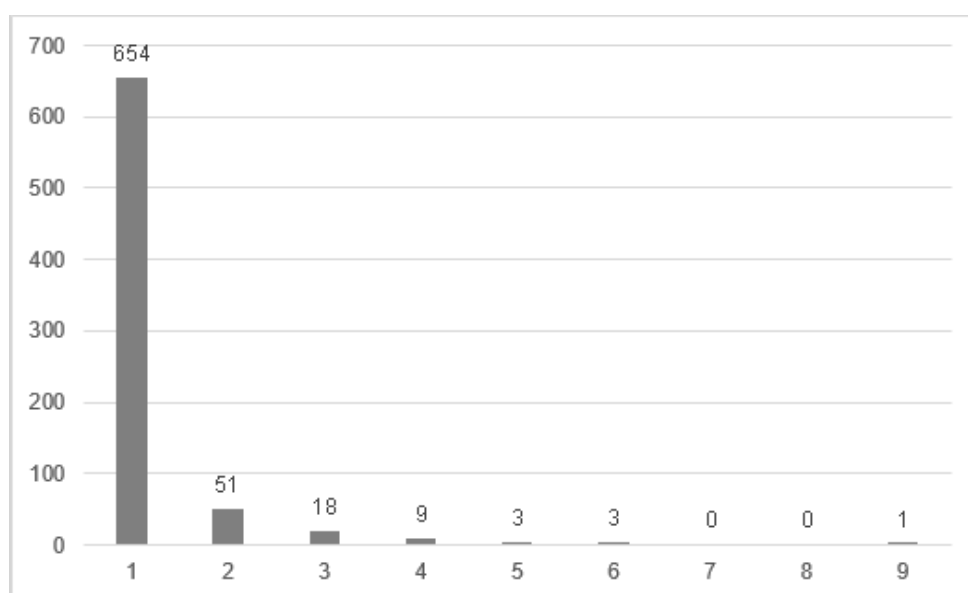
15	<i>Building biomedical web communities using a semantically aware content management system</i>	27
16	<i>ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data</i>	13
17	<i>BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature</i>	44
18	<i>The evolution of the web and implications for eResearch</i>	24
19	<i>Web 3.0 and health librarians: an introduction</i>	49
20	<i>Neuropsychology 3.0: Evidence-based science and practice</i>	25
21	<i>A semantic web management model for integrative biomedical informatics</i>	48
22	<i>Semantic web may be cancer information's next step forward</i>	0
		888

Fonte: elaborada pelo autor.

Feita a apresentação dos artigos com a respectiva quantidade de referências, buscou-se mapear possível repetições de referências dentre os trabalhos que compuseram o *corpus* da RSL.

O número de vezes que uma determinada referência foi utilizada nos diversos trabalhos variou entre um e nove. Artigos que serviram de referência para somente um dos 22 trabalhos incluídos no *corpus* de análise totalizaram 654 artigos. As quantidades podem ser vistas na Figura 8.

Figura 8 Quantidade total de vezes que um artigo foi usado como referência pelos trabalhos do *corpus*



Fonte: elaborada pelo autor.



De acordo a Figura 8, é possível constatar que 73% dos artigos foram citados somente 1 vez, enquanto 27% foram citados 2 ou mais vezes, o que mostra que há trabalhos que inspiraram vários dos trabalhos incluídos no *corpus* da RSL.

Ainda sobre os números ilustrados na Figura 8, destaca-se que somente um artigo que foi citado 9 vezes dentre 22 trabalhos, o que indica que foi usado como referência para 41% dos trabalhos analisados. 6 artigos foram usados como referência em 6 trabalhos que compuseram o *corpus*, 3 artigos que foram mencionados em 5 trabalhos, 9 artigos apareceram como referência de 4 dos trabalhos, e 18 artigos mencionados em 3 trabalhos, totalizando 37 artigos que foram referência para pelo menos 3 dos trabalhos contidos no *corpus*. Os artigos que foram citados no mínimo 3 vezes são apresentados na Tabela 5.

Tabela 5 Quantidade de vezes que um artigo foi utilizado como referência para trabalhos contidos no *corpus*

Quantidade	Referência
9	<i>Wilkinson MD, Links M. BioMOBY: an open source biological web services proposal. Brief Bioinform 2002;3 4):331–341.</i>
6	<i>Berners-Lee T, Hendler J, Lasilla O. The Semantic Web. Sci Am 2001;284(5):34–43.</i>
6	<i>Mons B, Ashburner M, Chichester C, et al. Calling on a million minds for community annotation in WikiProteins. Genome Biol 2008;9(5):R89.</i>
6	<i>Ruttenberg A, Clark T, Bug W, et al. Advancing translational research with the Semantic Web. BMC Bioinformatics 2007;8(Suppl 3):S2.</i>
5	<i>Cheung KH, Smith AK, Yip KYL et al. Semantic web approach to database integration in the life sciences. In: Baker CJO, Cheung KH (eds). Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences. New York: Springer, 2007, 11–30.</i>
5	<i>Dowell R, Jokerst R, Day A, Eddy S, Stein L. The distributed annotation system. BMC Bioinformatics 2001;2(1):7.</i>
5	<i>Hull D, Wolstencroft K, Stevens R, et al. Taverna: a tool for building and running workflows of services. NucleicAcids Res 2006;34:W729–32.</i>
4	<i>Belleau F, Nolin MA, Tourigny N, et al. Bio2RDF: towards a mashup to build bioinformatics knowledge systems. J Biomed Inform 2008; 41:706–16.</i>
4	<i>Good BM, Wilkinson MD. The life sciences semantic web is full of creeps! Brief Bioinform 2006; 7:275–286.</i>
4	<i>Hendler J (2003) Communication. Science and the semantic web. Science 299: 520–521.</i>
4	<i>Huss JW 3rd, Orozco C, Goodale J, et al. A gene wiki for community annotation of gene function. PLoS Biol 2008;6:e175.</i>
4	<i>Lee T, Pouliot Y, Wagner V, Gupta P, Stringer-Calvert D, Tenenbaum J, et al. BioWarehouse: a bioinformatics database warehouse toolkit. BMC Bioinformatics 2006;7(1):170.</i>
4	<i>Neerincx P, Leunissen J. Evolution of web services in bioinformatics. Brief Bioinformatics 2005;6(2):178–88.</i>

4	Post LJ, Roos M, Marshall MS, Driel RV, Breit TM (2007) A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. <i>Bioinformatics</i> .
4	Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. <i>Nat Biotechnol</i> 2007; 25:1251–5.
4	Stevens R, Robinson A, Goble C. myGrid: personalised bioinformatics on the information grid. <i>Bioinformatics</i> 2003;19(Suppl. 1):I302–4.
3	Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. <i>Nat Genet</i> 2000; 25:25–9.
3	Cheung KH, Yip KY, Townsend JP, et al. HCLS 2.0/3.0: health care and life sciences data mashup using Web 2.0/3.0. <i>J Biomed Inform</i> 2008; 41:694–705.
3	Ciccarese P, Wu E, Wong G, et al. The SWAN biomedical discourse ontology. <i>J Biomed Inform</i> 2008;41(5): 739–51.
3	Goble C, Stevens R. State of the nation in data integration for bioinformatics. <i>J Biomed Inform</i> 2008;41: 687–693.
3	Haas LM, Schwarz PM, Kodali P, et al. DiscoveryLink: a system for integrated access to life sciences data sources. <i>IBM Syst J</i> 2001; 40:489–511.
3	Kamel Boulos MN, Wheeler S (2007) The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. <i>Health Info Libr J</i> 24: 2–23.
3	Kawas E, Senger M, Wilkinson MD. BioMoby extensions to the Taverna workflow management and enactment software. <i>BMC Bioinformatics</i> 2006; 7:523.
3	Kinoshita, J. and Clark, T. (2007) <i>Alzforum, Methods in molecular biology</i> (Clifton, N.J, 401, 365-381.
3	Lord P, Bechhofer S, Wilkinson M, Schiltz G, Gessler D, Hull D, et al. Applying semantic web services to bioinformatics: experiences gained, lessons learnt. In: McIlraith SA, Plexousakis D, van Harmelen F, editors. <i>Third international semantic web conference (ISWC 2004)</i> , 2004. Hiroshima, Japan: Springer; 2004. p. 350–64.
3	Neumann EK, Quan D. Biodash: a semantic web dashboard for drug development. <i>Pac Symp Biocomput</i> 2006:176–87.
3	O'Reilly T. <i>What is Web 2.0: design patterns and business models for the next generation of software</i> . <a href="http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html">http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html</a> (1 February 2008, date last accessed).
3	Oinn T, Addis M, Ferris J, et al. Taverna: a tool for the composition and enactment of bioinformatics workflows. <i>Bioinformatics</i> 2004;20(17):3045–54.
3	Prlic A, Down TA, Kulesha E, et al. Integrating sequence and structural biology with DAS. <i>BMC Bioinformatics</i> 2007; 8:333.
3	Robu I, Robu V, Thirion B (2006) An introduction to the semantic web for health sciences librarians. <i>J Med Libr Assoc</i> 94: 198–205.
3	Salzberg S. Genome re-annotation: a wiki solution? <i>Genome Biol</i> 2007; 8:102.
3	Shadbolt N, Hall W, Berners-Lee T. The semantic web revisited, intelligent systems. <i>IEEE</i> 2006; 21:96–101.
3	Shah S, Huang Y, Xu T, Yuen M, Ling J, Ouellette BFF. Atlas—a data warehouse for integrative bioinformatics. <i>BMC Bioinformatics</i> 2005; 6(1):34.
3	Wang X, Gorlitsky R, Almeida JS (2005) From XML to RDF: how semantic web technologies will change the design of 'omic' standards. <i>Nat Biotechnol</i> 23: 1099–1103.

Fonte: elaborada pelo autor.

Optou-se, por motivo de brevidade, por não listar os artigos que foram utilizados como referência apenas 2 ou uma única vez.

#### 4.1.5 Referência a outros artigos do *corpus* da pesquisa

Após realizar mapeamento das referências bibliográficas dos 22 trabalhos, buscou-se mapear as referências cruzadas entre os trabalhos, ilustrando quais trabalhos selecionados para o *corpus* da pesquisa referenciaram outros trabalhos na mesma condição. No Quadro 4 é possível verificar quais trabalhos foram usados como referência para outros trabalhos contidos no *corpus*.

Quadro 4 Quantidade de referências a outros trabalhos contidos no *corpus*

Quantidade	Referência
3	Cheung KH, Yip KY, Townsend JP, <i>et al.</i> HCLS 2.0/3.0: health care and life sciences data mashup using <i>Web 2.0/3.0</i> . <i>J Biomed Inform</i> 2008;41: 694–705.
3	Goble C, Stevens R. State of the nation in data integration for bioinformatics. <i>J Biomed Inform</i> 2008;41: 687–693.
2	Zhang Z, Cheung KH, Townsend JP. Bringing <i>Web 2.0</i> to bioinformatics. <i>Brief Bioinform</i> 2008;10: 1–10.
1	Antezana, E., <i>et al.</i> (2009) Biological knowledge management: the emerging role of the Semantic <i>web</i> technologies, <i>Brief Bioinform</i> , 10, 392-407.
1	Bhagat, J., <i>et al.</i> (2010) BioCatalogue: a universal catalogue of <i>web</i> services for the life sciences, <i>Nucleic Acids Res</i> , 38, W689-694.
1	Giustini D. <i>Web 3.0</i> and medicine: make way for the semantic <i>web</i> . <i>BMJ</i> . 2007;335(7633):1273–74.

Fonte: elaborado pelo autor.

É possível constatar que, dentre os 22 trabalhos selecionados para o *corpus* da pesquisa, 6 foram usados como referência em outros trabalhos do *corpus*, conforme mostra o Quadro 4. Essas referências totalizaram 11 e, se comparado ao total de citações, representaram 1,2% do total.

O artigo dos autores Cheung *et al.* (2008), que foi citado 3 vezes dentre os 22 trabalhos do *corpus* evidencia a utilização da *Web 2.0* para obtenção do *mashup* de dados em prol dos cuidados de saúde, discutindo a integração dos dados, projetando tal estrutura para a utilização na *Web 3.0*, conceito que está interligado diretamente com a discussão dos outros artigos que abordam a *Web 2.0 / Web 3.0* na bioinformática, mostrando a relevância da temática.

Outro trabalho citado 3 vezes é o de Goble e Stevens (2008), que aborda diretamente a integração de dados na bioinformática. Os autores salientam o potencial de utilização da *Web 2.0* provendo o *mashup*, com foco na *web* semântica,

explorando os esforços e perspectivas sobre a temática da *Web 2.0* e/ou *Web 3.0* na bioinformática, o que se mostra relevante para a discussão por parte dos outros autores.

Zhang *et al.* (2008), o qual tem o ponto focal na proposição de uma comunidade colaborativa fazendo uso da *Web 2.0* para assunto ligados a bioinformática, foi citado 2 vezes, apresentando-se alinhado com a temática proposta para a presente pesquisa.

Os 3 artigos restantes, Antezana *et al.* (2009), Bhagat *et al.* (2010) e Giustini (2007), foram citados uma única vez em outro trabalho contido no *corpus* da RSL. Os 16 artigos restantes não foram citados nos artigos que compuseram o *corpus* da RSL, o que pode indicar certo isolamento dos autores ao discutirem acerca de uma temática que era do interesse de todos, ou o fato de muitos artigos estarem sendo publicados muito próximos uns dos outros, não dando a chance de que uns autores tivessem acesso ao trabalho dos outros antes de avançar com o seu próprio estudo.

#### 4.1.6 Categorização dos artigos

Foi possível identificar nos artigos selecionados 6 categorias em que se pôde agrupar as principais preocupações dos autores exporão analisar a forma como a *Web 2.0* interfere nas pesquisas de bioinformática. As 6 categorias foram as seguintes:

- *Integração de dados*: categoria em que os artigos têm como propósito o entendimento de como realizar a integração de diferentes bases de dados de bioinformática, expondo os benefícios e as dificuldades de assim proceder.
- *Pesquisa colaborativa*: propõe a ideia utilização das novas tecnologias para viabilizar a colaboração em pesquisas, seja pela participação de membros com anotações em pesquisas ou disponibilizando bases de dados e ferramentas *online* para que usuários possam executar diferentes análises em torno de bases de dados de uso comum.

- *Catálogo de dados*: agrupa artigos que reúnem listas contendo endereços (*links*) de repositórios de dados pertinentes à bioinformática. Diferentemente da categoria *integração de dados*, aqui não se busca o acesso à informação final, de forma unificada e integrada, mas a organização e acesso a informações que estão dispersas.
- *Serviços web*: os artigos contidos nessa categoria discutem a forma de desenvolver ferramentas *web* com intuito de integrar diferentes serviços em um *software* ou para a implantação de comunidades passíveis de moderação, em que possam ser discutidos assuntos pré-definidos.
- *Disciplina web*: essa categoria envolve artigos que tratam da importância da criação de uma disciplina com o propósito de estudar a aplicação das tecnologias de informação e comunicação de integração em rede na bioinformática.
- *Web semântica*: categoria que trata da discussão de uma *web* semântica e/ou *Web 3.0*, trazendo requisitos e projeções de utilização.

Como já mencionado, a divisão dos artigos em categorias se deu em categoria principal e categoria secundária.

A criação da categoria secundária se fez necessária uma vez que, na maioria dos artigos, havia uma temática secundária, que servia como suporte ou resultado da categoria principal, a partir da qual também ocorriam discussões interessantes que mereciam ser analisadas. Vale ressaltar que, similarmente à categoria principal, a categoria secundária também precisava estar ligada à relação entre *Web 2.0* e bioinformática.

O critério para classificação de uma categoria como principal ou secundária, dentro de um determinado artigo se deu com base na intensidade da argumentação. Foi definida como categoria principal sempre aquela com que o autor do trabalho se ocupou mais, ou seja, aquela que teve maior destaque em sua argumentação.

A classificação dos artigos nas categorias principal e secundária se deu conforme apresentado no Quadro 55, em que são fornecidos alguns trechos contidos em cada um dos artigos que levaram à sua classificação como principal ou secundária.

Quadro 5 Artigo x categorias principais e secundárias

Artigo		Categoria / Embasamento da categorização	
		Categoria principal	Categoria secundária
1	<i>Bringing Web 2.0 to bioinformatics</i>	<p>Integração de dados</p> <p>"Permitir a integração de dados hábeis a partir de fontes de dados numerosas, volumosas e heterogêneas é um grande desafio de bioinformática"</p>	<p>Pesquisa colaborativa</p> <p>"Ao estabelecer uma plataforma social, coletiva e colaborativa para criação, compartilhamento e integração de dados..."</p>
2	<i>HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0</i>	<p>Integração de dados</p> <p>"Descrevemos o potencial das tecnologias atuais da <i>Web 2.0</i> para obter o <i>mashup</i> de dados nos domínios de cuidados de saúde e ciências da vida (HCLS)"</p>	<p>Pesquisa colaborativa</p> <p>"Essas ferramentas podem ser usadas por pessoas sem experiência em programação para executar <i>mashup</i> de dados pela <i>web</i>."</p>
3	<i>State of the nation in data integration for bioinformatics</i>	<p>Integração de dados</p> <p>"Este artigo baseia-se na palestra inaugural de Carole Goble, proferida na Integração de Dados sobre Cuidados de Saúde e Ciências Biológicas"</p>	<p>Serviços <i>web</i></p> <p>"Os serviços da <i>web</i>, no entanto, tiveram uma maior penetração. Felizmente, a importância de interagir com os dados por meio de uma interface diferente de uma página da <i>web</i> 'apontar e clicar' agora é amplamente reconhecida, e é de se esperar que os dias de simulação de usuários e os resultados de captura de tela estejam contados."</p>
4	<i>Conducting research on the web: 2007 update for the bioinformatics links directory</i>	<p>Catálogo de dados</p> <p>"O <i>Bioinformatics Links Directory</i>, <a href="http://bioinformatics.ca/links_directory">http://bioinformatics.ca/links_directory</a>, é uma compilação de servidores mantida ativamente, publicada nesta e em edições anteriores das questões de pesquisa de ácidos nucleicos"</p>	<p>Integração de dados</p> <p>"Os cientistas já abraçam muitos desses conceitos da <i>Web 2.0</i>, como o valor de conjuntos de dados muito grandes e troca aberta de software, ideias e publicações. Existem também vários servidores de pesquisa que fornecem APIs para acesso programático aos seus recursos naturais."</p>
5	<i>A Dictyostelium bioinformatics resource update</i>	<p>Pesquisa colaborativa</p> <p>"o DictyBase descreve as anotações e recursos implementados desde 2006, incluindo melhor representação de <i>strain</i> e fenótipo, integração de elementos regulatórios transcricionais previstos, informações de domínio de proteína, vias bioquímicas, busca aprimorada em uma ferramenta <i>wiki</i> que permite aos membros da comunidade de pesquisa fornecer anotações."</p>	<p>Serviços <i>web</i></p> <p>"Continuaremos a fornecer novos conjuntos de dados e ferramentas para a comunidade de pesquisa. Outro objetivo é expandir o escopo de recursos externos em que genes e produtos gênicos de <i>Dictyostelium</i> são representados, que atualmente incluem o Gene Ontology, GenBank, ferramentas de análise UniProt e ortologia (InParanoid, OrthoMCL)."</p>

Artigo		Categoria / Embasamento da categorização	
		Categoria principal	Categoria secundária
6	<i>BioCatalogue: a universal catalogue of web services for the life sciences</i>	<p><b>Catálogo de dados</b></p> <p>"Um registro de serviços da <i>web</i> com informações sobre serviços disponíveis ajudará a reunir provedores de serviços e seus usuários."</p>	<p><b>Serviços <i>web</i></b></p> <p>"O uso de <i>web Services</i> para permitir o acesso programático à bioinformática <i>on-line</i> está se tornando cada vez mais importante nas ciências da vida."            "O BioCatalogue (<a href="http://www.biocatalogue.org/">http://www.biocatalogue.org/</a>) fornece uma interface comum para registrar, navegar e anotar os serviços da <i>web</i> para a comunidade de ciências da vida."</p>
7	<i>Web 3.0 and medicine</i>	<p><b>Integração de dados</b></p> <p>"No entanto, um recurso principal da <i>Web 3.0</i> é o uso de metadados - dados sobre dados. Isso transformará a <i>web</i> em um banco de dados gigante e a organizará de acordo com as linhas do PubMed, ou um de nossos confiáveis catálogos de bibliotecas médicas."</p>	<p><b>Web semântica</b></p> <p>"Consequentemente, os bibliotecários médicos acreditam que é necessário construir melhores mecanismos para a recuperação da informação."</p>
8	<i>Semantic web for Health Care and Life Sciences: a review of the state of the art</i>	<p><b>Integração de dados</b></p> <p>"A arquitetura consiste em uma ontologia de metadados, um repositório de metadados e uma interface que permite acesso ao repositório."</p>	<p><b>Pesquisa colaborativa</b></p> <p>"O artigo descreve como essa abordagem permite aos cientistas descobrir e vincular dados relevantes em diversas fontes de dados. Ele fornece uma plataforma para o desenvolvimento de aplicativos de informática integrativa."</p>
9	<i>At the intersection of public-health informatics and bioinformatics: using advanced web technologies for phylogeography.</i>	<p><b>Integração de dados</b></p> <p>"Este <i> mashup</i> de dados poderia permitir aos epidemiologistas compreender melhor a migração de doenças em hospedeiros animais, população nesses hospedeiros e o impacto da migração e da população viral no risco de infecção e doença em humanos. Descrevemos aqui o projeto preliminar e desenvolvimento de nosso sistema, ZooPhy, usando tecnologia avançada da <i>web</i>."</p>	<p><b>Pesquisa colaborativa</b></p> <p>"Essas estruturas permitem que recursos da <i>web</i> sejam anotados e definidos de uma maneira que facilite o compartilhamento de informações e a descoberta de conhecimento."</p>
10	<i>Data integration in bioinformatics: current efforts and challenges</i>	<p><b>Integração de dados</b></p> <p>"Apesar dos desafios, a promessa de integração de dados é alta: fontes de dados heterogêneas fornecem dados biológicos abrangendo uma ampla gama de campos de pesquisa. "</p>	<p><b>Pesquisa colaborativa</b></p> <p>" integração de dados em larga escala não requer apenas adoção de avanços em informática, mas também precisa de comunicações e colaborações entre pessoas em comunidades biológicas relacionadas para maximizar a abertura de dados via WS, padrões para dados biológicos, criar pipelines baseados em WS Semânticos e formar uma comunidade social. Essa comunidade aproveita a inteligência coletiva e a colaboração esforços para integração, análise e compartilhamento de dados."</p>
11	<i>Tools and collaborative environments for bioinformatics research</i>	<p><b>Pesquisa colaborativa</b></p> <p>"A pesquisa avançada requer interação intensiva entre uma infinidade de atores, frequentemente possuindo diferentes especialidades e, geralmente, trabalhando a distância uns dos outros. O campo da pesquisa colaborativa visa a estabelecer modelos e tecnologias adequados para apoiar adequadamente essas interações."</p>	<p><b>Serviços <i>web</i></b></p> <p>"Então, introduzimos alguns sistemas para criação de documentos colaborativos, incluindo sistemas <i>wiki</i> e ferramentas para desenvolvimento de ontologias, e revisamos alguns dos <i>wikis</i> biológicos mais interessantes."</p>

Artigo		Categoria / Embasamento da categorização	
		Categoria principal	Categoria secundária
12	<i>Biomashups: the new world of exploratory bioinformatics?</i>	<p><b>Integração de dados</b></p> <p>"Neste artigo investigamos a viabilidade de <i>mashups</i> como uma nova abordagem para a experimentação bioinformática, focando em um nicho exploratório entre o uso interativo da <i>web</i> e fluxos de trabalho robustos, e tentando identificar o intervalo de cálculos para o qual os <i>mashups</i> podem ser empregados."</p>	<p><b>Serviços <i>web</i></b></p> <p>"A bioinformática é dominada por bancos de dados <i>online</i> e ferramentas sofisticadas acessíveis pela <i>web</i>. Como tal, está idealmente posicionada para se beneficiar da combinação específica de serviços, de forma rápida e específica, via <i>mashups</i> na <i>web</i>."</p>
13	<i>Experience using web services for biological sequence analysis</i>	<p><b>Serviços <i>web</i></b></p> <p>"ferramentas através da <i>web</i> usando os chamados serviços da <i>web</i> tem um papel importante."</p> <p>"Neste artigo, discutimos as abordagens mais populares baseadas em SOAP / WS-I e REST e descrevemos nossa seção transversal da comunidade, experiências de fornecimento e uso de serviços da <i>web</i> no contexto da análise de sequências biológicas."</p>	<p><b>Integração de dados</b></p> <p>"Finalmente, problemas de integração de dados sintáticos e semânticos com vários serviços da <i>web</i> são discutidos."</p>
14	<i>Biological knowledge management: the emerging role of the semantic web technologies.</i>	<p><b>Integração de dados</b></p> <p>"A gestão adequada do conhecimento está, portanto, se tornando fundamentalmente importante para os cientistas da vida, especialmente se eles trabalham com abordagens que dependem completamente da integração do conhecimento, como a biologia de sistemas."</p>	<p><b>Catálogo de dados</b></p> <p>"Várias iniciativas para organizar as fontes de conhecimento biológico em um recurso prontamente explorável estão sendo realizadas atualmente. Ontologias e <i>web</i> semântica revolucionam esses esforços. Aqui, revemos os benefícios, tendências, possibilidades atuais e o potencial que isso representa para as biociências."</p>
15	<i>Building biomedical web communities using a semantically aware content management system</i>	<p><b>Serviços <i>web</i></b></p> <p>"O SCF apoia o discurso da comunidade no estilo 'Web 2.0' estruturado entre os pesquisadores, disponibiliza recursos de dados heterogêneos para o cientista colaborador, captura a semântica da relação entre os recursos e estrutura o discurso em torno dos recursos."</p>	<p><b>Pesquisa colaborativa</b></p> <p>"Desenvolvemos o SCF como uma plataforma reutilizável para colaboração <i>online</i>, estruturada e avançada, em pesquisa biomédica, que aproveita essas ontologias e recursos de RDF."</p>
16	<i>ASAP: a web-based platform for the analysis and interactive visualization of single-cell RNA-seq data</i>	<p><b>Pesquisa colaborativa</b></p> <p>"Este Pipeline Automatizado de Análise de Células Únicas – Automated Single-cell Analysis Pipeline (ASAP) combina uma ampla gama de algoritmos comumente usados com ferramentas de visualização sofisticadas. Em comparação com plataformas de análise scRNA-seq existentes, os pesquisadores (incluindo aqueles que não possuem experiência computacional) são capazes de interagir com os dados de maneira direta e em tempo real."</p>	<p><b>Serviços <i>web</i></b></p> <p>"Desenvolvemos uma plataforma totalmente integrada, baseada na <i>web</i>, visando à análise completa do alinhamento do genoma pós-genoma scRNA-seq: da análise, filtragem e normalização dos arquivos de dados de contagem de entrada à representação visual dos dados."</p>
17	<i>BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature</i>	<p><b>Integração de dados</b></p> <p>"Neste artigo, apresentamos o Bioinformatics Resource Inventory (BIRI), uma nova abordagem para descobrir e indexar automaticamente recursos públicos disponíveis de bioinformática usando informações extraídas da literatura científica."</p>	<p><b>Serviços <i>web</i></b></p> <p>"Desenvolvemos um serviço <i>web</i> para fornecer um conjunto de primitivos de consulta de alto nível para acessar o índice. O serviço <i>web</i> pode ser usado por serviços <i>web</i> de terceiros ou aplicativos baseados na <i>web</i>."</p>



Artigo		Categoria / Embasamento da categorização	
		Categoria principal	Categoria secundária
18	<i>The evolution of the web and implications for eResearch</i>	<p><b>Disciplina de web</b></p> <p>"Se quisermos entender e antecipar esses novos modos de colaboração, precisamos de uma disciplina que estude a web como um todo. Ciência da web é essa disciplina."</p>	<p><b>Pesquisa colaborativa</b></p> <p>"A web mudou a maneira como pesquisamos. No entanto, eScience e eResearch estão mudando a web. Essas atividades estão trazendo um enorme volume de dados, conteúdo da web, tornando-o reutilizável. Estão estabelecendo ferramentas e métodos de colaboração, que aprimoram o processo social da ciência."</p>
19	<i>Web 3.0 and health librarians: an introduction</i>	<p><b>Web semântica</b></p> <p>"Este artigo apresenta alguns dos principais conceitos e princípios da web 3.0 para bibliotecários de saúde. Ao fazê-lo, pretende explorar algumas das questões e terminologias associadas ao desenvolvimento projetado da web nos próximos 10 anos, e a um nível de generalidade que, esperamos, aumentará a conscientização e incentivará o debate."</p>	<p><b>Integração de dados</b></p> <p>"Os debates sobre a Web 3.0 ainda são um pouco teóricos, mas um tema comum é "desenvolver uma rede integrada de dados" com base em princípios sólidos de design de sistemas de informação. Alguns especialistas dizem que os princípios da biblioteconomia devem desempenhar um papel no aprimoramento da organização da web."</p>
20	<i>Neuropsychology 3.0: Evidence-based science and practice</i>	<p><b>Pesquisa colaborativa</b></p> <p>"(2) criação de bases de conhecimento neuropsicológicas colaborativas; e (3) projeto de métodos de avaliação baseados na web que permitem desenvolvimento livre, implementação de amostra grande e refinamento dos testes neuropsicológicos e os constructos que estes visam avaliar"</p>	<p><b>Integração de dados</b></p> <p>"Até agora não existem grandes repositórios para dados neuropsicológicos, apesar da consistência relativamente alta nos tipos de dados e homogeneidade substancial de variáveis específicas que são coletados. A evidência neuropsicológica compreende principalmente dados de grupos e dados de casos individuais. Os dados do grupo existem principalmente em publicações de pesquisa ou manuais proprietários de editores de teste"</p>
21	<i>A semantic web management model for integrative biomedical informatics</i>	<p><b>Web semântica</b></p> <p>"As tecnologias da web Semântica têm o potencial de atender à necessidade de distribuir e representações evolutivas que são críticas para os sistemas Biologia e pesquisa biomédica translacional."</p>	<p><b>Serviços web</b></p> <p>"Como esta tecnologia é incorporado ao desenvolvimento de aplicações, podemos esperar que tanto o software quanto o domínio de produtividade de propósito geral software específico instalado em nossos computadores pessoais será cada vez mais integrado com o controle remoto"</p>

Artigo		Categoria / Embasamento da categorização	
		Categoria principal	Categoria secundária
22	<i>Semantic web may be cancer information's next step forward</i>	<p><b>Web semântica</b></p> <p>“Tim Berners-Lee apelidou de “Web 3.0”, ou a Web Semântica - também pode ser o próximo marco na informação sobre o câncer. Já implantada no censo norte-americano, no catálogo da varejista eletrônica Best Buy e nas páginas do Facebook, a tecnologia da <i>web</i> Semântica poderia dar origem a um aplicativo matador que permite ao oncologista clínico acessar, integrar e analisar dados sobre drogas e genômica, registros médicos, e outras informações relacionadas ao câncer para melhorar o cuidado e a eficiência”</p>	<p><b>Integração de dados</b></p> <p>“O navegador pesquisa 10enas de bancos de dados de câncer em todo o mundo e, em segundos, retorna uma probabilidade que se aproxima ou afasta 100% de certeza, dependendo do tipo e quantidade de dados disponíveis. O mesmo navegador também pode fornecer as últimas descobertas de pesquisa em um caminho. Por exemplo, um geneticista que descobre um gene supressor de tumor entra nos achados em um banco de dados de genômica internacional. Um estatístico entra em um estudo populacional sobre o efeito do gene na sobrevivência do câncer em 5 anos em um banco de dados de epidemiologia. Uma startup de biotecnologia entra com informações sobre uma droga que ativa o gene em um banco de dados farmacêutico. Finalmente, um oncologista com um paciente que tem o gene carrega o arquivo do paciente em um navegador da <i>web</i> semântico. O navegador - que “rastrea” bancos de dados científicos para obter informações...”</p>

Fonte: elaborado pelo autor.

Feita a devida categorização, foi possível constatar a quantidade de artigos contidos em cada categoria, permitindo uma visão clara do volume de artigos com foco em cada uma delas.

A distribuição numérica dos artigos para as categorias principal e secundária é apresentada na Tabela 6.

Tabela 6 Categoria principal/secundária x quantidade de trabalhos

Categoria	Principal	Secundária	Total
Integração de dados	10 (45%)	5 (22%)	15 (35%)
Pesquisa colaborativa	4 (18%)	7 (30%)	11 (25%)
Serviços <i>web</i>	2 (9%)	8 (36%)	10 (22%)
Catálogo de dados	2 (9%)	1 (6%)	3 (7%)
Disciplina de <i>web</i>	1 (5%)	0 (0%)	1 (2%)
Web Semântica	3 (14%)	1 (6%)	4 (9%)
<b>Total</b>	<b>22</b>	<b>22</b>	

Fonte: elaborada pelo autor.

## 4.2 ANÁLISE DOS RESULTADOS

Uma vez mapeados e classificados os artigos, foi possível determinar o percentual que cada categoria ocupa dentro da totalidade dos artigos que compuseram a RSL. Os percentuais atribuídos para cada categoria sofreram arredondamento visando a uma melhor apresentação.

Tal classificação consiste no agrupamento entre as 6 categorias (*Integração de dados*, *Catálogo de dados*, *Pesquisa colaborativa*, *Serviços web*, *Disciplina web* e *Web semântica*), que foram aplicadas como principais e secundárias, permitindo assim uma “dupla classificação” de cada artigo.

### 4.2.1 Categorização principal dos trabalhos

Conforme citado anteriormente, após leituras dos trabalhos, foi proposta a categorização que está ilustrada no Quadro 5, que também contém trechos selecionados, que evidenciam o foco central de cada trabalho.

Os trabalhos do *corpus* da RSL que apresentaram como categoria principal “integração de dados” representam 45% do todo (10 trabalhos), tendo como foco principal a discussão em torno da integração de bases de dados de bioinformática, fazendo uso das tecnologias da *Web 2.0*.

Dentre os 10 trabalhos classificados como “integração de dados” na categoria principal, 7 apresentam uma discussão acerca de ferramentas e arquiteturas voltadas para a integração de dados, propondo a utilização de plataformas propriamente ditas ou *frameworks* que buscam a integração de dados, conforme visto nos trabalhos de Zhang *et al.* (2008), Cheung *et al.* (2008), Cheung *et al.* (2009), Scotch *et al.* (2010), Hogan *et al.* (2011), Zhang *et al.* (2011) e De la Calle *et al.* (2009).

Os 3 trabalhos restantes discutem dificuldades e perspectivas para a integração de dados, sendo apresentada como principal dificuldade a heterogeneidade no tipo dos dados e como perspectiva o uso de bases de dados integradas e confiáveis, buscando ganho de tempo nas pesquisas, conforme

destacam Goble e Stevens (2008), Giustini (2007) e Antezana, Kuiper e Mironov (2009).

Os trabalhos alocados na categoria principal “pesquisa colaborativa” representaram 18% do total dos trabalhos selecionados na RSL (quatro trabalhos), tendo como foco principal a discussão sobre a temática da pesquisa colaborativa no segmento da bioinformática fazendo uso da *Web 2.0*. Os 4 trabalhos contidos nesta categoria propõem a utilização de ferramentas do tipo *wiki* com o intuito de fomentar a colaboração entre os pares em pesquisas de bioinformática, criando bases de conhecimento e discutindo as características desse tipo de ferramenta (Fey *et al.*, 2008, Romano, Guigno e Pulvirenti, 2011, Vicent *et al.*, 2017 e Bilder, 2011).

No trabalho de Romano, Guigno e Pulvirenti (2011), a discussão principal gira em torno do uso de *wiki* na bioinformática, enfatizando alguns princípios e revisando as ferramentas mais relevantes das redes sociais, dando uma atenção as redes de apoio à colaboração científica.

Sobre ferramentas com foco na “pesquisa colaborativa”, Fey *et al.* (2008) e Vicent *et al.* (2017) são mais diretos ao proporem ferramentas específicas que contribuam para a colaboração entre pesquisadores. Fey *et al.* (2008) ilustram a utilização de um banco de dados específico, referente à *Dictyostelium discoideum*, abrigando a sequência genômica completa e sua respectiva literatura, propondo assim aos pares a participação com o fornecimento de anotações nas pesquisas científicas por meio de uma ferramenta com característica de *wiki*.

Vicent *et al.* (2017) apresentam uma plataforma *web* que tem como propósito fornecer a análise completa do alinhamento do genoma, permitindo aos membros da comunidade que não possuem o conhecimento específico realizar procedimentos de análises sofisticadas, com a colaboração podendo consistir no *upload* de bases de dados ou na utilização de análises realizadas por outros pesquisadores. Bilder (2011) trata da criação de bases de conhecimento neuropsicológicas colaborativas.

Os trabalhos classificados cuja categoria principal foi “catálogo de dados” representam 9% (2 trabalhos) do total dos artigos selecionados para a RSL. Diferente da “integração de dados” que visa a um agrupamento dos dados

propriamente ditos, os trabalhos contidos nesta categoria têm como propósito principal a proposição de diretórios que contenham o caminho para repositórios de dados ou listagem de *web services*. Logo, os repositórios de dados podem ser comparados a catálogos.

O primeiro dos 2 trabalhos é de autoria de Fox *et al.* (2007), que apresentam um diretório composto por *links* de servidores contendo pesquisas científicas ligadas a ácidos nucleicos, juntamente com ferramentas e outros recursos para pesquisas em ciências da vida. O segundo trabalho que compõe a categoria principal de “catálogo de dados” é de autoria de Bhagat *et al.* (2010), o qual trata de uma interface simples que apresenta vários serviços *web* voltados para as ciências da vida, além de prover a funcionalidade de registro de serviços *web* por parte dos membros da comunidade, denominado sistema *Biocatalogue*.

O sistema *Biocatalogue* tem como característica o monitoramento constante, permitindo a identificação de erros em serviços. A abordagem *Web 2.0* está presente, uma vez que todos os serviços disponíveis podem sofrer algum tipo de anotação por parte dos membros da comunidade científica (BHAGAT *et al.*, 2010).

A categoria principal “serviço *web*” representa 9% (2 trabalhos) do total dos trabalhos selecionados para a RSL, com foco principal na utilização de serviços *web* na bioinformática, ou na proposição de ferramentas embasadas na *Web 2.0* para funcionar como serviços, os quais podem ser considerados ferramentas *web* com o propósito de colaboração.

Ao abordar a utilização de *web services*, Stockinger *et al.* (2008) discutem as abordagens baseadas em aspectos técnicos e descrevem as experiências de fornecimento e uso de *web services* em prol das análises biológicas, trazendo à tona a discussão acerca dos problemas na utilização de *web services*. Buscando uma maior comunicação entre os membros da comunidade científica interessada em biotecnologia Das *et al.* (2008) propõem a utilização de um *framework* no formato de plataforma reutilizável para a colaboração *on-line* de forma estruturada, baseada na *Web 2.0* e buscando a estruturação em torno dos recursos de dados heterogêneos.

A categoria principal “disciplina *web*” teve um único trabalho alocado. Hall, De Roure e Shadbolt (2008) enfatizam a importância da *web* em suas diferentes vertentes (*Web 1.0*, *Web 2.0* e *Web 3.0*), propondo a criação de uma disciplina focada para estudá-las e compreender os seus impactos na biotecnologia. Os autores enfatizam o crescimento da colaboração na *web*, característica derivada da *Web 2.0*, reforçando a necessidade de buscar o embasamento para a geração da *web* semântica. Além disso, Hall, De Roure e Shadbolt (2008) reforçam que a *Web 2.0* está sendo usada com sucesso na bioinformática, devido à sua característica de colaboração e compartilhamento, proporcionando novas experiências.

A categoria “*Web semântica*” teve 3 trabalhos classificados como categoria principal, totalizando 14%. Cho e Giustini (2008), discutem a utilização da *web* semântica nas bibliotecas de saúde, externando como característica da *web* atual o volume de informações e a ideia do desenvolvimento de uma “rede integrada de dados” embasada em premissas sólidas contidas no desenho dos sistemas de informação. Deus *et al.* (2008) e Martin (2011) discutem a propriedade evolutiva da *web* semântica, com o intuito de se obter a análise automática do conjunto de dados.

#### 4.2.2 Categorização secundária dos trabalhos

Além da classificação nas categorias principais, foi identificado nos trabalhos um segundo tópico de interesse dos autores, envolvendo uma abordagem que pode ser considerada complementar ou resultante do foco principal. Logo, foi realizada uma classificação dentre as categorias estabelecidas (Integração de dados, Catálogo de dados, Pesquisa colaborativa, Serviços *web*, Disciplina *web* e *Web Semântica*).

Ainda na Tabela 6, é possível observar a frequência com que cada uma das categorias emergentes do estudo apareceu nos trabalhos de forma secundária. Da mesma forma que foi realizado para a categoria principal, foram identificados trechos nos trabalhos que permitiram e justificaram a categoria secundária definida, conforme ilustrado no Quadro 5.

Por se tratar de uma classificação secundária, os trabalhos tendem a estar relacionados à categoria secundária de forma menos intensa do que à categoria

principal, devido à associação se dar em função de uma premissa adotada ou um resultado obtido.

Dentre as 6 categorias emergentes, não houve nenhum trabalho que apresentasse “*Disciplina de web*” como classificação secundária. Isto decorreu de essa categoria ter sido criada em função de um único artigo que percebia a necessidade de a área de bioinformática estudar a evolução da *web* e a forma como isso afetaria os estudos futuros da área. Tratou-se de um trabalho atípico preocupado com a proposição da criação de uma disciplina nova.

Feita tal explanação, dentre as 5 categorias restantes (Integração de dados, Catálogo de dados, Pesquisa colaborativa, Serviços *web* e *Web* semântica), “*serviços web*” contou com 8 trabalhos que a trataram como um segundo aspecto importante em sua discussão, representando 36% dos trabalhos. Os 8 trabalhos que estão contidos em tal categoria secundária compreendem, de forma similar, que para atender seus diferentes objetivos, serviços *web* se mostram como um requisito fundamental, permitindo conexões entre diferentes bancos de dados, juntamente com diferentes membros da comunidade, reunidos em torno de um assunto em comum. Os artigos que compartilharam essa visão foram: Goble e Stevens (2008), Fey *et al.* (2008), Bhagat *et al.* (2010), Giustini (2007), Romano, Guigno e Pulvirenti (2011), Hogan *et al.* (2011), Vicent *et al.* (2017) e De la Calle *et al.* (2009).

A segunda categoria secundária mais frequente foi “*pesquisa colaborativa*”, a qual apareceu em 7 trabalhos nessa condição de categoria secundária, representando 30% dos trabalhos que fizeram parte do *corpus* da RSL. Esses trabalhos tratam do compartilhamento de informação ou da inteligência coletiva como fator importante da construção do conhecimento e integração entre os membros da comunidade. São eles: Zhang *et al.* (2008), Cheung *et al.* (2008), Cheung *et al.* (2009), Scotch *et al.* (2010), Zhang *et al.* (2011), Das *et al.* (2008) e Hall, De Roure e Shadbolt (2008).

A categoria “*integração de dados*” teve 3 trabalhos classificados como categoria secundária: Fox *et al.* (2007), Stockinger *et al.* (2008) e Cho e Giustini (2008). Ao abordar o tema, Stockinger *et al.* (2008) reforçam a importância da

discussão acerca da integração de dados sintáticos e semânticos. Mesmo que de forma indireta, Fox *et al.* (2007) salientam a utilização de conceitos da *Web 2.0* por parte dos cientistas, os quais valorizam os grandes conjuntos de dados e os meios tecnológicos disponíveis para sua integração. Cho e Giustini (2008) trazem a discussão para o âmbito de bibliotecas voltadas para a saúde e a integração dos dados para um alicerce voltado à *web* semântica.

Fechando a classificação, a categoria secundária “*Catálogo de dados*” ocorreu em um único artigo. Antezana, Kuiper e Mironov (2009) enfatizam existirem várias iniciativas para organizar as fontes de conhecimento da área biológica, onde se tem a participação expressiva das tecnologias ontológicas e de *web* semântica para auxiliar nesta tarefa, apresentando potencial a ser explorado nas biociências.

#### 4.2.3 Categorização agregada dos trabalhos

Feita a devida classificação dos trabalhos em categoria principal e categoria secundária, foi possível constatar, de forma numérica, a proporção que cada categoria ocupa dentro o todo. Buscou-se, aqui, unificar as classificações (principal e secundária) a fim de quantificar em sua totalidade a participação dos trabalhos e seus respectivos percentuais.

Uma vez que cada trabalho foi classificado em uma categoria principal e uma categoria secundária, no agregado o mesmo trabalho foi computado duas vezes, resultando em 44 categorizações (22 categorizações como principal e 22 categorizações como secundária).

A categoria agregada que mais recebeu trabalhos foi “*integração de dados*” respondendo por 35% das categorizações. A categoria “*pesquisa colaborativa*” vem em segundo lugar, detendo 25% das categorizações agregadas, e “*serviços web*” detêm 22% dos trabalhos classificados no agregado. Somadas as 3 respondem por 82% das categorizações dos trabalhos.

Nas 3 últimas categorias no agregado “*web semântica*” detêm 9%, “*catálogos de dados*” detêm 7% e fechando o agregado, “*disciplina de dados*” detêm 2%. No agregado, as 3 últimas categorias somaram 18% do total.



### 4.3 DISCUSSÃO DOS RESULTADOS

Este capítulo tem como propósito a discussão das principais ideias discutidas nos 22 trabalhos que compuseram a RSL, a fim de entender a relação entre *Web 2.0* e bioinformática. Para essa discussão, optou-se por abordar as categorias de forma unificada, uma vez que a classificação em principal ou secundária se deu internamente a cada um dos trabalhos, com base na ênfase com que o tópico de uma categoria era tratado no estudo e foi feito o uso da mesma taxonomia para classificação entre principal e secundária, com a distinção no enfoque.

Conforme visto no item 4.2, foram definidas 6 categorias adotadas na classificação dos trabalhos (*Integração de dados, Catálogo de dados, Pesquisa colaborativa, Serviços web, Disciplina web e Web semântica*).

A integração de dados se refere a disponibilizar, sem necessidade de esforço integrativo adicional, dados contidos em bases de várias origens, conforme enfatizam Zhang *et al.* (2008), ideia que vai ao encontro do que propõem Handsfield *et al.* (2009), que consideram a coletividade uma propriedade da *Web 2.0*. Tal propriedade viabiliza a inteligência distribuída e coordenada preconizada por Lévy (2003) e permite o compartilhamento de bases de várias origens e distribuídas (ZHANG *et al.*, 2008).

Um fator determinante na integração de dados são as ferramentas utilizadas. Ferramentas *Web* e *frameworks* que possuem arquitetura embasada nos conceitos da *Web 2.0* são apresentadas nos trabalhos de Zhang, Cheung, Townsend e Jeffrey (2008), Cheung *et al.* (2008), Cheung, *et al.* (2009), Scotch *et al.* (2010), Hogan *et al.* (2011) e De la Calle *et al.* (2009).

Solomon e Schrum (2007) reforçam a ideia de que a arquitetura das ferramentas deve dispor do propósito explícito de colaboração e manipulação das informações de forma coletiva, reforçando essa importante característica da *Web 2.0*.

Cormode e Krishnamurthy (2008) comparam a *Web 1.0*, que tem como característica o fato de os criadores de conteúdo serem limitados, com a *Web 2.0*, que tem como filosofia a descentralização e colaboração na criação de conteúdo,

característica essa defendida por Zhang *et al.* (2008) ao abordar as bases de diferentes origens.

Ainda entre os trabalhos categorizados como preocupados com *integração de dados*, Giustini (2007) enfatiza a aplicação da *Web 2.0* no acesso a publicações científicas, propondo uma grande base de conhecimento de assuntos ligados a saúde, permitindo buscas em dados mais heterogêneos e refinados, ponto de vista este que vai ao encontro do que é proposto por Mount (2001), que sugere a necessidade do acesso e processamento de grandes volumes de dados biológicos, fazendo uso de ferramentas computacionais.

Ao focar em *serviços web*, possíveis usos da informática na bioinformática são destacados, conforme indicado por Muniz (2003) e Gibas e Jambeck (2001), propondo a geração de grandes bases de dados a partir da biologia molecular e a análise destas informações, por meio dos algoritmos computacionais, conceitos que são explorados por Stockinger *et al.* (2008) e Das *et al.* (2008) ao abordarem o uso de ferramentas e *serviços web* a serem utilizados para análises biológicas.

Para Stockinger *et al.* (2008) e Das *et al.* (2008), nas arquiteturas *web* é possível identificar o conceito proposto por O'Reilly (2007), em que a *web* funciona como plataforma e os dados como diferencial, com enfoque do software acima da plataforma.

Ainda na linha de *serviços web*, Das *et al.* (2008) propõem a utilização das ferramentas voltadas para as comunidades biomédicas como um canal de compartilhamento de informações entre pesquisadores, reforçando a necessidade de uma estrutura focada na produtividade e na promoção das pesquisas científicas na área das ciências naturais. É possível identificar, na argumentação desses autores, a relação entre a biologia e informática com enfoque nos recursos computacionais, o que já havia sido destacado por Lesk (2002), que atribui a evolução nos estudos da área de biologia aos avanços computacionais ocorridos ao longo das últimas décadas.

Vicent *et al.* (2017) também ressaltam que a plataforma baseada na *web*, com o propósito de analisar dados ligados ao alinhamento do genoma, prioriza a

utilização de recursos computacionais e apresenta ganho de eficiência, se comparada a outras abordagens. Para esses autores, é possível identificar uma preocupação com a utilização da plataforma por parte dos pesquisadores com diferentes níveis de experiência computacional pertencentes a comunidade, trazendo à tona a pesquisa colaborativa. Tal preocupação pode ser identificada no posicionamento de Levy (2009) ao abordar a gestão do conhecimento, focando-se no conhecimento produzido fazendo uso de ferramentas *Web 2.0* e, portanto, na participação ativa dos usuários.

Acerca da *pesquisa colaborativa*, Romano, Guigno e Pulvirenti (2011) enfatizam que a pesquisa em nível avançado tem como premissa a interação intensiva por parte de uma infinidade de atores, se fazendo necessário o estabelecimento de propostas e tecnologias adequadas com o intuito de suportar tais interações.

Com intuito de aprimoramento de um segmento específico da biotecnologia, Fey *et al.* (2008) descrevem uma ferramenta contendo a sequência completa do genoma do *Dictyostelium*, contendo informações precisas e funcionais, permitindo aos pesquisadores interagir e contribuir por meio de um *wiki*, construindo conhecimento de forma colaborativa com os demais membros da comunidade científica.

Nos artigos de Romano, Guigno e Pulvirenti (2011) e Fey *et al.* (2008) é possível identificar o posicionamento no conceito defendido por Michalsky (2012), ao propor a relação entre *Web 2.0* e inteligência, o qual considera que a colaboração entre membros de um determinado agrupamento de usuários propicia resultados e, conseqüentemente, demanda por ferramentas com o propósito de suportar a participação dos indivíduos, estimulando a inteligência coletiva.

Os termos acima são características claras da inteligência coletiva descrita por Lévy (2003), para quem a participação coletiva é valorizada. Além deste, Alag (2009) afirma que, quando existe colaboração ou competição entre indivíduos, existe a tendência do surgimento da inteligência coletiva, característica que fica clara na

forma como são descritos os objetivos de colaboração nos artigos inseridos na categoria *pesquisa colaborativa*.

Em Cho e Giustini (2008), Deus *et al.* (2008) e Martin (2011) é possível identificar a atenção para com o crescimento das bases de dados e uma preocupação com a sua correta formatação, preocupação essa que tem como propósito a criação de um alicerce para a *web semântica*, tecnologia que tem em sua essência o processamento e entendimento dos dados por parte dos *softwares*, conforme expõem Berners-Lee *et al.* (2001), o que só pode ocorrer a partir da estruturação desses dados, procedimento que, segundo Berners-Lee *et al.* (2001), está em andamento.

Giustini (2007) expõe uma discussão junto a médicos sobre os métodos de publicações científicas, destacando a característica embrionária da *Web 1.0* e *Web 2.0* em contraponto à tendência da *Web 3.0* ser mais madura, devido à capacidade de processamento cognitivo e recuperação da informação, o que é corroborado por Gualtieri e Coccia (2010) ao abordarem a indexação de documentos e a ação de agentes não-humanos processando as informações, ideia similar à proposta por Berners-Lee *et al.* (2001) ao alocarem a responsabilidade de tal ação a *softwares*.

Acerca do *catálogo de dados*, mostrou-se a ideia de organização de fontes de conhecimento na área biológica com o propósito de se ter um catálogo de endereços referentes a repositórios, conforme explanado por Antezana, Kuiper e Mironov (2009). Diferente da *integração de dados* que busca repositórios centralizados, o *catálogo de dados* é algo similar a um indexador.

A respeito da *web semântica*, o resultado da pesquisa mostrou a aplicação das ferramentas com características da *Web 2.0* como pré-requisito para o funcionamento da *web semântica*. Cho e Giustini (2008) propuseram a utilização da *web semântica* em bibliotecas de saúde. Deus *et al.* (2008) e Martin (2011) propuseram a evolução da *web semântica*, com o intuito de se obter uma automação na análise do conjunto de dados.

Fechando a discussão em torno dos resultados, a categoria *disciplina web*, por meio do trabalho dos autores Hall, De Roure e Shadbolt (2008), enfatizou a

importância de criação de uma disciplina com o propósito de discutir as diferentes vertentes da *web*, neste caso com o enfoque na aplicação na biotecnologia. Os autores ainda reforçaram a aplicação da *Web 2.0* à colaboração, sendo esta considerada como base para a aplicação da *web* semântica.

## 5 CONCLUSÃO E RECOMENDAÇÕES

Esta pesquisa buscou, por meio de uma revisão sistemática da literatura, compreender como a inteligência coletiva potencializada por ferramentas da *Web 2.0* pode ser utilizada para aumentar a eficácia dos estudos e projetos na área de bioinformática. O estudo foi realizado analisando-se a percepção dos próprios pesquisadores da área, evidenciada em suas publicações a respeito do assunto, que foram mapeadas para possibilitar a pesquisa.

Assim, a fim de atender o primeiro objetivo da pesquisa, foi realizada a seleção de publicações que tratavam do uso da inteligência coletiva por meio das ferramentas da *Web 2.0* em projetos de bioinformática. Para a seleção dos trabalhos que fizeram parte da RSL, foi usado como referência o protocolo adotado por Padilha e Graeml (2015) e Vianna, Graeml e Peinado (2017), inspirados no protocolo originalmente proposto por Kitchenham (2004), cuja aplicação resultou na seleção de 22 trabalhos para comporem o *corpus* da pesquisa, sendo um capítulo de livro e vinte e um artigos.

Em um primeiro momento, foi feita a busca dos trabalhos envolvendo as expressões “*bioinformatics*” e “*Web 2.0*”. Durante a leitura dos trabalhos, contudo, a discussão projetava uma conexão com a *Web 3.0*, ou *web* semântica, como alguns autores a intitulam. Em virtude disso, foi realizada uma nova busca, fazendo uso das expressões “*bioinformatics*” e “*Web 3.0*”, a fim de se obter abrangência maior da pesquisa.

Após a leitura de todos os trabalhos selecionados, em sua integralidade, foi possível constatar que distintos autores trouxeram diferentes ideias e discussões em torno das formas de aplicação da inteligência coletiva e colaboração entre os pares, por meio do uso da *Web 2.0* no campo da bioinformática, que podem ser consideradas complementares entre si, ou realizando discussões para o uso da *Web 3.0* no campo da bioinformática.

Após a leitura dos trabalhos, foi realizado um mapeamento voltado a identificar algumas características dos trabalhos, sendo elas a temporalidade das

publicações, os países em que estavam localizados os periódicos em que os trabalhos foram publicados e as universidades e departamentos a que os pesquisadores estavam afiliados.

No que tange ao entendimento temporal, buscou-se mapear o período em que se deu a discussão da aplicação da *Web 2.0* na bioinformática. Os trabalhos selecionados para a composição do *corpus* da RSL foram publicados no período entre 2007 e 2017, sendo que, depois de 2011, a discussão do tema arrefeceu. Entre 2007 e 2011 foram publicados vinte e um dos 22 trabalhos que compuseram a RSL.

A partir da análise da linha do tempo das publicações, acredita-se que o fenômeno da *Web 2.0* e seus efeitos/contribuição para a bioinformática tenham sido considerados relativamente bem compreendidos depois daqueles primeiros anos de intensa publicação a respeito do assunto. É possível que, a partir do momento que a *web* semântica se materialize, de forma que seus impactos na pesquisa possam ser mais nitidamente percebidos, ocorra uma nova onda de trabalhos discutindo o assunto. Berners-Lee *et al.* (2001) imaginavam que isso já estaria acontecendo a partir de 2010, mas, embora haja expectativa a respeito, a transição da *Web 2.0* para uma *Web 3.0* ainda parece insipiente.

No que diz respeito à localização dos periódicos onde os trabalhos foram publicados, constatou-se que, dentre os 22 trabalhos, 17 foram publicados em periódicos sediados no Reino Unido, totalizando aproximadamente 77% dos trabalhos. Dentre os cinco trabalhos restantes, quatro foram publicados em meios dos Estados Unidos e um no Canadá, totalizando os 23% restantes. A partir de tais números é possível constatar que os autores priorizaram periódicos ingleses para tais publicações.

Contudo, como em um mundo globalizado a publicação tende a ocorrer em periódicos que garantam mais visibilidade à obra e prestígio aos autores, a identificação da localização das editoras dos periódicos talvez traga informação menos valiosa do que a das instituições a que estão afiliados os autores. Por isso, buscou-se mapear as universidades e/ou centros de pesquisas e os departamentos

em que atuam os autores dos trabalhos que compuseram o *corpus* da RSL. A partir dessa análise, foi possível criar um possível cenário de quais áreas de estudo estão relacionadas com o uso da inteligência coletiva em projetos de bioinformática, atendendo o segundo objetivo específico desta pesquisa. Como resultado deste objetivo, foram identificadas cinco áreas de pesquisas preponderantes, relacionadas aos departamentos em que atuam os autores. São elas: “*Computação*”, “*Bioinformática*”, “*Medicina*”, “*Biologia*”, “*Psicologia*” e “*Ciências Sociais*”.

O maior número de pesquisadores faz parte de departamentos ligados à Computação, seguida pela área de estudos em Bioinformática. Essas duas áreas de estudos juntas totalizam 57% dos autores. Na sequência, as áreas de Medicina e Biologia ocupam o terceiro e quarto lugar, totalizando 30% do total dos autores. Psicologia e Ciências Sociais são responsáveis por menos de 2% do total. Tais indicadores evidenciaram a multidisciplinaridade, algo que tem relação direta com a inteligência coletiva ao envolver a multidão e ao se referir à bioinformática, envolvendo pesquisadores de diferentes departamentos em prol da mesma temática.

A fim de entender a colaboração de pesquisadores associados a diferentes universidades e centros de pesquisa, constatou-se que, em alguns trabalhos, houve a participação de pesquisadores de duas ou mais instituições, chegando a ocorrer um artigo com participação de autores de sete instituições diferentes, localizadas em cinco países, de três continentes. Por si só, essa participação de autores de lugares tão distantes já é uma evidência da capacidade da *Web 2.0* proporcionar o fomento da inteligência coletiva, na construção de conhecimento, de forma distribuída.

Além disso, constatou-se que, das 36 instituições diferentes em que estão alocados os pesquisadores, 16 estão localizadas na América do Norte, especificamente nos EUA, 14 estão localizadas na Europa e 6 estão localizadas em outros continentes. Tais números evidenciam uma predominância na discussão por parte de pesquisadores afiliados a universidades e centros de pesquisas situados nos EUA.

Adentrando no conteúdo dos trabalhos, a fim de mapear as ideias neles discutidas, criou-se uma categorização das temáticas prevalentes, identificando-se



quais delas apareciam de forma principal ou secundária nos trabalhos analisados. A categoria principal ilustrava a ideia central do autor. A segunda temática prevalente no trabalho era então considerada a temática secundária.

Atendendo o terceiro e último objetivo específico desta pesquisa, foram identificadas 6 categorias emergentes, utilizadas para rotular os artigos, quanto à sua temática principal e secundária: “*Integração de dados*”, “*Catálogo de dados*”, “*Pesquisa colaborativa*”, “*Serviços web*”, “*Disciplina web*” e “*Web semântica*”. As categorias emergiram dos trabalhos conforme enfoque dado pelos autores e terminologia utilizada. Alguns trechos dos trabalhos, que levaram à sua categorização, foram destacados, na seção de análise, para que o leitor pudesse compreender claramente como a categorização se deu.

“*Integração de dados*” e “*Pesquisa colaborativa*” apareceram em 63% dos trabalhos classificados como categoria principal, seguidas de “*Catálogo de dados*”, “*Serviços web*” e “*Web semântica*” que apareceram com destaque em 32% dos trabalhos. A última categoria “*Disciplina web*” apareceu em apenas um artigo. Vistos tais indicadores, fica evidente a percepção dos autores de que integração de dados e colaboração em pesquisa são dois aspectos em que a *Web 2.0* traz grande contribuição à pesquisa na área de bioinformática.

Agregando-se as temáticas que apareceram como temática principal dos trabalhos com aquelas que ocorreram para suportá-las, ou em consequência delas (temática secundária), “*Integração de dados*”, “*Pesquisa colaborativa*” e “*Serviços web*” estiveram presentes em 82% dos trabalhos.

A integração de dados mereceu destaque em virtude da possibilidade de criação e/ou integração de bases de dados distribuídas, que possibilitem dados estruturados e centralizados, viabilizando o acesso dos pesquisadores a grandes massas de informação. Tais bases são derivadas da coletividade, uma vez que se tem pesquisadores trabalhando geograficamente separados e em projetos distintos, que, ainda assim, podem se beneficiar de bases centralizadas, de uso comum.

A pesquisa colaborativa emergiu como a segunda temática de maior atenção, com destaque para a discussão do uso de ferramentas e ambientes onde

pesquisadores com propósitos similares possam trabalhar, mesmo que geograficamente separados, de forma orquestrada, a fim de construir conhecimento em conjunto.

A terceira temática mais discutida foi relacionada a *serviços web*, que têm como finalidade a operacionalização da coletividade, para que as ferramentas permitam aos pesquisadores a construção do conhecimento coletivamente, fazendo uso da *Web 2.0* como plataforma, com característica de facilitadora da atividade de pesquisa.

As temáticas *catálogo de dados* e *web semântica* tiveram uma participação menor nas discussões, mesmo com a aplicação da inteligência coletiva. A temática *disciplina web* foi discutida somente em um trabalho, trazendo uma proposição bem específica do *core* da discussão desta pesquisa com a proposição de uma disciplina que possa discutir sobre a *web* e suas possibilidades de resultado para a bioinformática.

Esta pesquisa permitiu mapear quais as temáticas propostas pelos autores para a aplicação da inteligência coletiva com base na utilização de ferramentas da *Web 2.0*, enaltecendo a participação da coletividade em projetos de bioinformática. Além disto, foi possível identificar tal característica na estrutura de confecção dos trabalhos utilizados nesta RSL, onde evidenciou-se a coletividade dos pesquisadores trabalhando de forma coletiva e organizada, característica que foi citada no *Handbook of Collective Intelligence* (2011), que apresenta a definição de inteligência coletiva como se tratando de indivíduos organizados em grupos nos quais realizam atividades de maneira coletiva, aparentando ser inteligentes.

O mapeamento realizado nesta pesquisa permitiu criar uma classificação que pode ser aplicada em outros trabalhos envolvendo inteligência coletiva, para além da área da bioinformática. O agrupamento permitiu compreender como a inteligência coletiva gerada por meio da aplicação das ferramentas da *Web 2.0* pode ser aplicada na geração de valor nos estudos de bioinformática.

Evidenciou-se, durante a leitura dos trabalhos, uma transição, considerando que a *web semântica* (também conhecida como *Web 3.0*) é a “evolução” da *Web 2.0*.

Recomenda-se um estudo mais aprofundando desta transição e a verificação do seu real acontecimento, uma vez que, no “auge” das publicações acerca da temática, a era da *web* semântica ainda não tinha sido iniciada, mas apenas projetada. Na verdade, na percepção dos autores que discutem a temática de bioinformática, *web* semântica parece ainda ser mais uma promessa do que algo que já esteja se concretizando nas pesquisas em bioinformática. Assim que resultados promissores começarem a ocorrer nessa nova frente, seguramente haverá uma nova leva de artigos tratando do assunto, assim como ocorreu entre 2007 e 2011 para explicar a *Web 2.0*, que estava se materializando.

## REFERÊNCIAS

ABRAM, S. *Web 2.0, Library 2.0, and Librarian 2.0: Preparing for the 2.0 World*. In: **Library and Information Services in Astronomy V**. 2007. p. 161, 2007.

ALAG, S. **Collective intelligence in action**. New York: Manning, 2009.

ANTEZANA, E.; KUIPER, M.; MIRONOV, V. Biological knowledge management: the emerging role of the semantic *web* technologies. **Briefings in bioinformatics**, v. 10, n. 4, p. 392-407, 2009.

ATTWOOD, T. K. *et al.* Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective. In: **Bioinformatics-trends and methodologies**. IntechOpen, 2011.

BERNERS-LEE, T. *et al.* The semantic *web*. **Scientific american**, v. 284, n. 5, p. 28-37, 2001.

BHAGAT, J. *et al.* BioCatalogue: a universal catalogue of *web* services for the life sciences. **Nucleic acids research**, v. 38, n. suppl\_2, p. W689-W694, 2010.

BOCCATO, V. R. C. Metodologia da pesquisa bibliográfica na área odontológica e o artigo científico como forma de comunicação. **Rev. Odontol. Univ. Cidade São Paulo**, São Paulo, v. 18, n. 3, p. 265-274, 2006.

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. O método da revisão integrativa nos estudos organizacionais. **Gestão e Sociedade**, v. 5, n. 11, p. 121-136, 2011.

BOULOS, M.; WHEELER, S. "The emerging *web 2.0* social software: an enabling suit of sociable technologies in health and health care education", **Health Information and Libraries Journal**, Vol. 24 No. 1, pp. 2-23, 2007.

CHEUNG, K. *et al.* HCLS 2.0/3.0: Health care and life sciences data mashup using *Web 2.0/3.0*. **Journal of biomedical informatics**, v. 41, n. 5, p. 694-705, 2008.

CHEUNG, k.; PRUD'HOMMEAUX, E.; WANG, Y.; STEPHENS, S., Semantic *web* for Health Care and Life Sciences: a review of the state of the art, **Briefings in Bioinformatics**, Volume 10, Issue 2, March 2009.

CHO, A.; GIUSTINI, D. *Web 3.0* and health librarians: an introduction. **Journal of the Canadian Health Libraries Association**, v. 29, n. 1, p. 13-18, 2008.

CHOUDHURY, N. World wide *web* and its journey from *Web 1.0* to *Web 4.0*. **International Journal of Computer Science and Information Technologies**, v. 5, n. 6, p. 8096-8100, 2014.

CORMODE, G.; KRISHNAMURTHY, B. Key differences between *Web 1.0* and *Web 2.0*. **First Monday**, v. 13, n. 6, 2008.

CUNHA, M. B. da. Para saber mais: fontes de informação em ciência e tecnologia. 2001.

CUNHA, C.; MELO, M. A confiança nos relacionamentos interorganizacionais: o campo da biotecnologia em análise. **RAE – Eletrônica**, v. 5, n.2, p.1-26, 2006.

DE ARAÚJO, N. D. *et al.* A era da bioinformática: seu potencial e suas implicações para as ciências da saúde. **Estudos de biologia**, v. 30, n. 70/72, 2008.

DE LA CALLE, G.*et al.* BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature. **BMC bioinformatics**, v. 10, n. 1, p. 320, 2009.

DOSZKOCS, T. Semantic search engines mean well. **Online**, v. 34, n. 4, p. 36-42, 2010.

FALEIRO, F. G.; ANDRADE, SRM de. Biotecnologia: uma visão geral. Biotecnologia: estado de arte e aplicações na agropecuária. **Planaltina, Embrapa**, p. 13-29, 2011.

FERRO, E S.; Biotecnologia translacional: hemopressina e outros peptídeos intracelulares. **Estudos avançados**, v. 24, n. 70, p. 109-121, 2010.

FEY, P. *et al.* dictyBase—a Dictyostelium bioinformatics resource update. **Nucleic acids research**, v. 37, n. suppl\_1, p. D515-D519, 2008.

FIGUEIREDO, L.; PENTEADO, M. I.; MEDEIROS, P. Patentes em biotecnologia – patenteamento em biotecnologia agropecuária: cenário brasileiro. **Biotecnologia Ciência e Desenvolvimento**, v. 9, n. 36, 2006.

FOX, J. A.; MCMILLAN, S.; OUELLETTE, BF F. Conducting research on the *web*: 2007 update for the bioinformatics links directory. **Nucleic acids research**, v. 35, n. suppl\_2, p. W3-W5, 2007.

GIBAS, C.; JAMBECK, P.; FENTON, J. M. Developing bioinformatics computer skills. " **O'Reilly Media, Inc.**", 2001.

GIUSTINI, D. "Web 3.0 and medicine." **BMJ (Clinical research ed.)** vol. 335,7633, 2007.

GOBLE, C.; STEVENS, R. State of the nation in data integration for bioinformatics. **Journal of biomedical informatics**, v. 41, n. 5, p. 687-693, 2008.

HAGEN, J. B. The origins of bioinformatics. **Nature Reviews Genetics**, v. 1, n. 3, p. 231, 2000.

HALL, W.; DE ROURE, D.; SHADBOLT, N. The evolution of the *web* and implications for eResearch. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 367, n. 1890, p. 991-1001, 2008.

HANDBOOK of collective intelligence (2011) Disponível em: <<http://scripts.mit.edu/~cci/HCI/index.php>>. Acesso em: abril 2019.

HANDSFIELD, L. J.; DEAN, T. R.; CIELOCHA, K. M. Becoming critical consumers and producers of text: Teaching literacy with *Web 1.0* and *Web 2.0*. **The Reading Teacher**, v. 63, n. 1, p. 40-50, 2009.

HAZLETT, S.A.; MCADAM, R.; GALLAGHER, S. "Theory building in knowledge management: in search of paradigms", **Journal of Management Inquiry**, Vol. 14 No. 1, pp. 31-42, 2005.

HEDBERG, J.G.; BRUDVIK, O.C. Supporting dialogic literacy through mashing and modding of places and spaces. **Theory Into Practice**, 47(2), 138–149, 2008.

HENDLER, J." Science and the semantic *web*." **Science**, 299, 520-521, 2003.

HERRERA M. C. L. Componentes para interoperabilidade entre redes sociais na *Web 2.0*. **Tese de Doutorado**. Universidade de São Paulo, 2013.

HESPER, B.; HOGEWEG, P. Bioinformatica: een werkconcept. **Kameleon**, v. 1, n. 6, p. 28-29, 1970.

HOGAN, J. M. *et al.* Biomashups: the new world of exploratory bioinformatics? **Concurrency and Computation: Practice and Experience**, v. 23, n. 11, p. 1169-1178, 2011.

HOGEWEG, P. Simulating the growth of cellular forms. **Simulation**, v. 31, n. 3, p. 90-96, 1978.

HULTQUIST, S.; HARRISON, R.; YANG, Y. Patenting bioinformatic inventions emerging trends in the united states. **Nature Biotechnology**, v. 20, n. 7, p. 743 4, 2002.

JACSÓ, P. GoogleScholar: the pros and the cons. **Online Information Review**, v. 29, n. 2, p. 208-214, abr. 2005.

JITHESH, P. V. *et al.* Genegrid: grid based solution for bioinformatics application integration and experiment execution. **18th IEEE Symposium on Computer-Based Medical Systems**, p. 523-528, june 2005.

KITCHENHAM, B. **Procedures for undertaking systematic review**. Joint Technical Report, Computer Science Department, Keele University and ICT National Australia, 2004.

LEIMEISTER, Jan Marco. Collective intelligence. **Business & Information Systems Engineering**, v. 2, n. 4, p. 245-248, 2010.

LEVY, M. *WEB 2.0* implications on knowledge management. **Journal of knowledge management**, v. 13, n. 1, p. 120-134, 2009.

LÉVY, P. A inteligência coletiva: por uma antropologia do ciberespaço. São Paulo: Loyola, 2003. **\_. Cibercultura. Tradução de Carlos Irineu da Costa.–São Paulo: Ed**, v. 34, 1999.

MALEEWONG, K.; ANUTARIYA, C.; Wuwongse, V. (2008). A collective intelligence approach to collaborative knowledge creation. In: **Fourth International Conference on Semantics, Knowledge and Grid**, p. 64- 70, IEEE.

MALONE, T. W.; CROWSTON, K. The interdisciplinary study of coordination. **ACM Computing Surveys (CSUR)**, v. 26, n. 1, p. 87-119, 1994

MALONE, T. W.; LAUBACHER, R.; DELLAROCAS, C. Harnessing Crowds: Mapping the Genome of Collective Intelligence. **MIT Sloan Research Paper** No. 4732-09, 2009.

MCKENNA, R. Real-time marketing. **Harvard Business Review**, v. 73, n. 4, p. 87-95, 1995.

MIRANDA, G. F.; GUALTIERI, F.; COCCIA, P. How the new *web* generations are changing library and information services. **Medical Reference Services Quarterly**, v. 29, n. 2, p. 132-145, 2010.

MOUNT, D. M. Bioinformatics - Sequence and Genome Analysis. 1. ed. [S.l.]: **Cold Spring Harbor Laboratory Press**, 2001.

MUNIZ, J. R. Aplicação da bioinformática no estudo dos genes e enzimas envolvidos na síntese da goma fastidiana produzida pela *xylella fastidiosa*. Dissertação (Mestrado) | **Instituto de Física de São Carlos-Universidade de São Paulo**, São Carlos, 2003.

NAMBISAN, S.; NAMBISAN, P. How to profit from a better'virtual customer environment'. **MIT Sloan management review**, v. 49, n. 3, p. 53, 2008.

OECD, A. A framework for biotechnology statistics. **Organisation for Economic Co Operation and Development, Paris**, 2005.

O'REILLY, T. What is *Web 2.0*: Design patterns and business models for the next generation of software. **Communications & strategies**, n. 1, p. 17, 2007.

PADILHA, M.; GRAEML, A. Inteligência coletiva e gestão do conhecimento: quem é meio e quem é fim. In: AMCIS 2015–Americas Conference on Information Systems. 2015.

PAROUTIS, S.; AL SALEH, A. Determinants of knowledge sharing using *Web 2.0* technologies. **Journal of knowledge management**, v. 13, n. 4, p. 52-63, 2009.

PIZZANI, L. *et al.* A arte da pesquisa bibliográfica na busca do conhecimento. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 10, n. 2, p. 53-66, 2012.

PREECE, J; SHNEIDERMAN, B. The reader-to-leader framework: Motivating technology-mediated social participation. **AIS transactions on human-computer interaction**, v. 1, n. 1, p. 13-32, 2009.

ROMANO, P.; GIUGNO, R.; PULVIRENTI, A. Tools and collaborative environments for bioinformatics research. **Briefings in bioinformatics**, v. 12, n. 6, p. 549-561, 2011.

DA ROSA, A. LM *et al.* Development of a collaborative environment applied to pediatric oncology. In: **Proceedings of the 2006 ACM symposium on Applied computing**. ACM, 2006. p. 256-257.

ROSEN, D.; NELSON, C. *Web 2.0: A new generation of learners and education*. **Computers in the Schools**, v. 25, n. 3-4, p. 211-225, 2008.

SALOMON, D. V. *Como fazer uma monografia*. 11a ed. São Paulo: Martins Fontes; 2004.

SCOTCH, M. *et al.* At the intersection of public-health informatics and bioinformatics: using advanced *web* technologies for phylogeography. **Epidemiology (Cambridge, Mass.)**, v. 21, n. 6, p. 764, 2010.

SERAGAN, T. **Programming collective intelligence: building smart Web 2.0 applications**. Toby Segaran., Beijing, 2007.

SEVERINO, A. J. *Metodologia do trabalho científico*. 22<sup>a</sup> ed rev ampl São Paulo: **Cortez**; 2002.

SOLOMON, G.; SCHRUM, L. *Web 2.0: New tools, new schools*. **Eugene, OR: International Society for Technology in Education**, 2007.

STOCKINGER, Heinz *et al.* Experience using *web* services for biological sequence analysis. **Briefings in bioinformatics**, v. 9, n. 6, p. 493-505, 2008.

SUDESHNA, D. *et al.* Building biomedical *web* communities using a semantically aware content management system, **Briefings in Bioinformatics**, Volume 10, Issue 2, March 2009, Pages 129–138, 2009.

SZUBA, T. M. *Computational collective intelligence*. **John Wiley & Sons, Inc.**, 2001.

VIANNA, F. R. P. M.; GRAEML, A.; PEINADO, J. As motivações que levam usuários a participar de plataformas de crowdsourcing online. In: **Simpósio de Administração da Produção, Logística e Operações Internacionais, São Paulo, agosto, 2017. Anais...:XX SIMPOI**, São Paulo, 2017.

VICENT, G. *et al.* ASAP: a *web*-based platform for the analysis and interactive visualization of single-cell RNA-seq data, **Bioinformatics**, Volume 33, Issue 19, 01 October 2017, Pages 3123–3125, 2017.

VOLPATO, E. de S. N. *Pesquisa bibliográfica em ciências biomédicas*. **Jornal de Pneumologia**, v. 26, n. 2, p. 77-80, 2000.



Yuan, W.; Chen, Y.; Wang, R.; Du Z. (2008). Collective intelligence in knowledge management. In: **The International Federation for Information Processing**, v. 254, p. 651-655, Springer.

ZHANG, Z.; CHEUNG, K.; TOWNSEND, J. P. Bringing *web 2.0* to bioinformatics. **Briefings in bioinformatics**, v. 10, n. 1, p. 1-10, 2008.

WECHSLER, D. Die Messung der Intelligenz Erwachsener Huber. **Bern Stuttgart Wien**, 1964.