



UNIVERSIDADE CATÓLICA PORTUGUESA
CENTRO REGIONAL DAS BEIRAS
DEPARTAMENTO DE CIÊNCIAS DE SAÚDE

*FATORES ASSOCIADOS A CANCRO:
- PROPOSTA DE ONTOLOGIA*

Dissertação apresentada à Universidade Católica Portuguesa
para obtenção do grau de Mestre em Medicina Dentária

Por
Ana Margarida Marques

Viseu, 2013



UNIVERSIDADE CATÓLICA PORTUGUESA
CENTRO REGIONAL DAS BEIRAS
DEPARTAMENTO DE CIÊNCIAS DE SAÚDE

*FATORES ASSOCIADOS A CANCRO:
- PROPOSTA DE ONTOLOGIA*

Dissertação apresentada à Universidade Católica Portuguesa
para obtenção do grau de Mestre em Medicina Dentária

Por: Ana Margarida Marques

Sob orientação de:

Professora Doutora Marlene Maria Tourais de Barros

Professor Doutor Joel Perdiz Arrais

Viseu, 2013

“As origens desta doença confundem-se com as da própria vida”

Georges Beau

Ao meu pai,
por todo apoio transatlântico, e por ser quem é.

À minha avó Zélia,
pelo apoio incondicional e com certeza pelo orgulho de me ver chegar a esta etapa.

À Ester
pela orientação maternal, o que permitiu realizar este percurso.

Aos meus tios Rosa e Amilcar Marques
pelo carinho e apoio incondicional.

Ao meu irmão Fernando Soares,
sem o teu apoio nunca teria sido possível.

À minha Mãe,
estejas onde estiveres, vês eu consegui.

Resumo

O cancro surge como consequência de múltiplos eventos moleculares, resultando numa desregulação celular tendo como causa última a perda de homeostase celular. Alguns pacientes são susceptíveis a estes eventos de desregulação devido a características hereditárias, outros, por contacto com fatores exógenos que provocam a desregulação celular. Depois das doenças cardiovasculares o cancro é a patologia que mais mata em todo mundo a seguir, no entanto, se detetado em fases precoces torna-se menos debilitante e com uma superior taxa de sobrevivência.

Apesar do extenso volume de informação publicada e dedicada ao efeito de fatores com impacto na patologia carcinogénica, a terminologia utilizada não se encontra uniformizada nem organizada, dificultando a sua interpretação. Os sistemas de Informação que utilizam a tecnologia Web Semântica permitem a uniformização de terminologias dispersas.

O trabalho realizado no âmbito da presente dissertação visou a criação de um sistema de ontologias descritoras dos fatores associados a cancro que foram reunidos na ferramenta OntoXcancer. O OntoXcancer é uma ferramenta bioinformática que permite a gestão e análise de fatores associados a cancro e permitiu a construção da estrutura de uma base de dados - FactorXcancer. No âmbito desta dissertação, os fatores associados a cancro foram catalogados de acordo com as ontologias geradas pelo OntoXcancer e a informação bibliográfica referente a estes fatores permitiu o estabelecimento de uma estratégia metodológica para a consulta da informação disponível no FactorXcancer.

Palavras-chave: Ontologias; Web semântica; Cancro.

Abstract

Cancer appears as a consequence of multiple molecular events resulting in cell disruption having as ultimate cause the loss of cellular homeostasis. Some patients are susceptible to these events of deregulation due to hereditary characteristics, others, by contact with external factors that caused cell disruption. Cancer has the second highest mortality rate all over the world, only behind cardiovascular diseases. However, if detected at an early stage, it becomes less debilitating and the survival rate increases.

Despite the extensive amount of published data and dedicated investigations to best understand the factors that impact on carcinogenic pathology, the terminology is not standardized or organized, and this information remains difficult to interpret. Information Systems that use Semantic Web technology allow the standardization of dispersed terminologies, which will help in this matter.

The work performed under this dissertation aimed to create a system of ontologies that could describe the factors associated with cancer. They were then gathered in OntoXcancer tool. The OntoXcancer is a bioinformatics tool that allows the management and analysis of factors associated with cancer and allowed the construction of a structured database - FactorXcancer. Within this dissertation, the cancer-associated factors were cataloged according to the ontologies generated by OntoXcancer and bibliographic information concerning these factors allowed the establishment of a methodology for consulting of the information available in FactorXcancer.

Keywords: Ontologies, SemanticWeb; Cancer.

Agradecimentos

À Professora Dra. Marlene Barros,
por todo o incentivo, motivação e grande ajuda durante a elaboração deste trabalho e,
em especial por me ter ajudado a ter uma nova perspectiva sobre a vida.

Ao Professor Dr. Joel Arrais,
pela ajuda e motivação dadas ao longo do desenvolvimento deste trabalho.

Ao Prof. Nuno Rosa
pela ajuda extra ao longo do desenvolvimento deste trabalho.

À minha especial amiga, Maria Isabel Nunes,
pelo carinho e apoio incondicional.

Aos meus primos Rita, Luís, Tomás e Rui Marques,
por serem a minha fonte de alegria.

Ao meu binómio Tiago paiva, pelo crescimento e aprendizagem mútua.

Aos grandes amigos, Diogo Calçada, Gonçalo Monteiro, Manuel Ribeiro, Catarina
Fernandes e tantos outros que ao longo destes anos fizeram parte de momentos
memoráveis e inesquecíveis.

Índice Geral

Resumo	XVIII
Abstract	XX
Agradecimentos	XXII
Índice Geral.....	XXIV
Lista de Abreviaturas	XXVII
1. INTRODUÇÃO	33
A.1 Motivação para o desenvolvimento de uma ontologia de fatores associados a cancro.....	33
A.2 Requisitos para a construção de ontologias relacionadas com fatores associados a cancro	38
A.2.1 Web Semântica	38
A.2.1.1 Arquitetura da Web Semântica	39
A.2.3 Ontologias	40
A.2.4 Bio-ontologias.....	44
A.2.5 Linguagens de desenvolvimento de uma ontologia	46
A.2.6 Aplicações Semânticas na área da Biomedicina	47
B.1 Sistema de classificação de neoplasias.....	49
B.2 Fatores etiológicos em cancro	50
3. OBJETIVOS	54
5. MATERIAIS E MÉTODOS	58
4.1 Construção da base de dados - <i>FactorXcancer</i>	58
4.1.2 Metodologia aplicada na pesquisa da bibliografia referente a fatores associados a cancro	58
4.2 Base de dados baseada em ontologias – <i>FactorXcancer</i>	63
4.3 Anotação da base de dados sobre fatores de neoplasias	66
4.4 Desenvolvimento da <i>OntoXcancer</i> uma ontologia dedicada a fatores associados a cancro	68

4.4.1 Programa informático utilizado para a construção do <i>OntoXcancer</i> - <i>Protégé</i>	68
4.4.2 Construção da biblioteca ontológica – <i>OntoXcancer</i>	70
4.4.3 Determinação da metodologia para análise estatística da base de dados - <i>FactorXcancer</i>	103
5. RESULTADOS E DISCUSSÃO.....	107
5.1 Exemplos de redes semânticas do <i>OntoXcancer</i>	109
5.2 Exemplos de como a informação disponível no – <i>FactorXcancer</i> –pode ser consultada	111
5.3 Resultados obtidos pelo tratamento estatístico	122
6. CONCLUSÃO	128
7. DESENVOLVIMENTOS FUTUROS	132
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	136
9. ANEXOS	158
10. Índice de Figuras.....	165
11. Índice de Gráficos	169
12. Índice de Tabelas	171
13. Índice de Quadros	173

Lista de Abreviaturas

ADA- American Dental Association
ADH- Antidiuretic hormone
ALDH- Aldehyde dehydrogenases
BKV - BK polyomavirus
CaBIG- The Cancer Biomedical Informatics Grid
CID- International Classification of Diseases
CMV - Cytomegalovirus
DAML - DARPA Agent Markup Language
DNA- Deoxyribonucleic acid
EBV- 1 - Epstein-Barr virus Type 1
EBV- 2 - Epstein-Barr virus Type 2
EGFR- Epidermal growth factor receptor
GEO - Gene Expression Omnibus
GO - Gene Ontology
GSMTM1- Glutathione S-transferase M1
GST- Glutathione-S-transferase
HBC - Hepatitis C virus
HBV - Hepatitis B virus
HGNC – HUGO Gene Nomenclature Committee
HHV- 8 - Human herpes virus 8
HIV - Acquired Immune Deficiency Syndrome Virus
HIV- 1 - Human Immunodeficiency Virus Type 1
HPV - Human papillomavirus
HPV- 11 - Human papillomavirus type 11
HPV- 16 - Human papillomavirus type 16
HPV- 18 - Human papillomavirus type 18
HPV- 31 - Human papillomavirus type 31
HPV- 32 - Human papillomavirus type 32
HPV- 33 - Human papillomavirus type 33

HPV- 35 - Human papillomavirus type 35
HPV- 39 - Human papillomavirus type 39
HPV- 45 - Human papillomavirus type 45
HPV- 51 - Human papillomavirus type 51
HPV- 52 - Human papillomavirus type 52
HPV- 56 - Human papillomavirus type 56
HPV- 58 - Human papillomavirus type 58
HPV- 59 - Human papillomavirus type 59
HPV- 6 - Human papillomavirus type 6
HPV- 68 - Human papillomavirus type 68
HPV- 73 - Human papillomavirus type 73
HPV- 82 - Human papillomavirus type 82
HSV- 1 - Herpes Simplex Virus Type 1
HSV- 2 - Herpes Simplex Virus Type 2
HTLV- 1 - Human T-lymphotropic virus 1
IARC- Internacional Agency for Research on cancer
ICD - International Classification of Diseases
IGF-I- Insuline growth factor I
IMC- Índice de massa corporal
ISO - International Standards Organization
JCV –HumanPolyomavirus JC
KSHV- Kaposi Sarcoma-Associated Herpes virus 8
LOP- Líquen Plano Oral
MDM-2 - E3 ubiquitin-protein ligase
MeSH - Medical Subject Headings
NACA –Cândida albicans não albicans
NCBI - National Center for Biotechnology Information
NCBO - National Center for Biomedical Ontology
NCI Thesaurus -National Cancer Institute’s Thesaurus
NCICB - NCI Center for Bioinformatics
NIH - National Institutes of Health
NLM - U.S. National Library of Medicine

NNN- N-nitrosornicotine

OBI - Ontology for Biomedical Investigations

OBO - The Open Biological and Biomedical Ontologies

OC- Oral cancer

OIL - Ontology Interchange Language

OMS - Organização Mundial de Saúde

OWL - Web Ontology Language

p53 - Cellular tumor antigen p53

PAH- Polycyclic aromatic hydrocarbons

RDF - Resource Description Framework

RL –Radicais livres

SCCHN- Squamous Cell Carcinoma of the Head and Neck

SHOE -Simple HTML Ontology Extension

SI - Sistemas de Informação

SV40 -Simianvirus 40

TI - Tecnologias de Informação

TSGs- Genes supressores tumorais

UMLS - Unified Medical Language System

VEGF- Vascular endothelial growth factor

W3C - World Wide Web Consortium

WHO- World Health Organization

WS - Web Semântica

WWW - World Wide Web

XML - Extensible Markup Language

XP- Xeroderma Pigmentososo

Capítulo I

Introdução

1. INTRODUÇÃO

A.1 Motivação para o desenvolvimento de uma ontologia de fatores associados a cancro.

Na atualidade, o conhecimento científico cria quantidades extraordinárias de informação acessível a qualquer pessoa, em qualquer parte do mundo, permitindo uma mudança no paradigma de investigação nos domínios da biologia e da medicina. No entanto, o fato da informação não estar uniformizada, torna inexistente a sua total compreensão por parte de toda a comunidade científica, é nesta linha que surge a necessidade da existência de técnicas de organização da informação⁽¹⁾.

A organização de informação científica tem por base dois pressupostos importantes: armazenamento de forma metódica, facilmente acessível e reutilização. Como tal, toda a informação terá de ser armazenada de uma forma a que os resultados de experiências, publicações científicas, entradas estruturadas em bases de dados catalogadas, entre outros, possam estar acessíveis e serem partilhados. No entanto, a forma como este processo é constituído não garante a reutilização nem a acessibilidade da informação a toda a comunidade científica. Além disso, os dados recuperados de uma base de dados, por usufruírem de uma linguagem própria, tem que ser interpretados para que possam fornecer suporte para o conhecimento científico⁽²⁾.

Ao associarmos o aumento do volume de informação à recente necessidade de organização e acessibilidade de conteúdos científicos, a comunidade científica sentiu a necessidade de criar a Web Semântica (WS), com base na utilização de ontologias. A Web Semântica (WS) é uma extensão da web que nos permite a integração manual dos dados, para que sejam interpretáveis pelos sistemas de informação. A WS permite também a interligação entre os dados de modo a atribuir significado, (conceito/definição) aos conteúdos publicados para que sejam perceptíveis e uniformizados quer pelos sistemas de informação quer pela comunidade científica^(3,12).

As ontologias fornecem uma língua franca que permite o processamento de dados entre os sistemas informáticos e a integração de recursos *web* de forma

inteligente, possibilitando pesquisas rápidas e precisas, facilitando a comunicação entre dispositivos heterogêneos acessíveis *online*⁽³⁾.

Com intuito de simplificar a partilha e reutilizar a informação, as ontologias descrevem conceitos, relações, restrições, axiomas de um domínio usando uma organização taxonómica especializada⁽⁴⁾. Especialistas das áreas da Engenharia Informática defendem a criação de ontologias genéricas. No entanto, de acordo com Leite⁽⁵⁾, a *web* do futuro deve ser constituída por várias ontologias pequenas e altamente contextualizadas, desenvolvidas localmente e não por especialistas. Assim sendo, o desenvolvimento de ontologias ou a reutilização de dados ontológicos pré-existentes deve ser simples. Permitindo que qualquer indivíduo, sem conhecimentos da área possa criá-las, permitindo deste modo a disponibilidade do maior número de dados possível. As ontologias constituem a base que permite uma comparação e reutilização de conhecimentos, pois estabelecem terminologias e interpretações semânticas comuns. As ontologias são criadas com o objetivo de definir critérios de regulação de termos e relações⁽¹⁵⁾, ou seja critérios que regulam os termos e relações entre si, também chamados de redes semânticas. Por exemplo, quem possuir o Gene ERCC5, vai desenvolver a patologia Xeroderma Pigmentoso, isto é, um termo (Gene ERCC5), tem relação com o outro (Xeroderma Pigmentoso). As ontologias podem ser utilizadas como apoio em diversas áreas de investigação, como a representação e gestão do conhecimento, processamento de linguagem natural (ou seja comum), bases de dados, integração de dados, bibliotecas digitais ou recuperação de dados isolados por vários sistemas⁽¹⁶⁾.

O crescente volume de informação cada vez mais complexa e heterogênea, cria necessidade de uma integração automática da mesma. As ontologias desempenham um papel importante nesta integração de informação, tornando possível aproveitar os dados biomédicos provenientes de fontes divergentes, com a certeza de que os mesmos conceitos terão o mesmo significado, mesmo quando utilizados em contextos díspares⁽⁶⁾. De acordo com o bioinformático Jim Golden, "*The key to bioinformatics is integration, integration, integration*". Russ Altman da Universidade de Stanford acrescenta "*To answer most interesting biological problems, you need to combine data from many data sources (...) However, creating seamless access to multiple data sources is extremely difficult*"⁽⁷⁾.

No entanto, vários estudos recentes demonstram a existência de lacunas nas terminologias existentes na área da biomedicina, não se encontrando em conformidade com os princípios terminológicos e ontológicos^(6,8-11). Na comunidade científica utilizam-se várias terminologias para o mesmo conceito. Atendendo à nomenclatura de patologias, verifica-se que “Ataxia Telangiectasia” equivale a “Síndrome de Louis- Bar”, e ambos os termos são igualmente utilizados. Mais exemplos são o conceito de “ cancro do estomago” e “cancro gastrico”, e a relação entre “Doença de Cowden” e “Síndrome de múltiplos hamartomas”.

O objetivo da WS passa pelo desenvolvimento de um modelo tecnológico que permite a partilha global do conhecimento, assistido pelos benefícios dos Sistemas de Informação (SI). Para que isto seja possível, integram-se linguagens específicas que serão descritas mais adiante. A utilização de sistemas de informação que utilizem a tecnologia WS é, atualmente, uma das soluções propostas e mais promissoras para integração de dados e conhecimento proveniente de várias fontes heterogéneas⁽¹³⁾. Com o surgimento desta metodologia surgem novas necessidades e com elas, estratégias para que os dados semânticos possam ser interpretáveis pelos sistemas de informação. Para resolver este problema surgem as ontologias⁽¹⁴⁾.

O aumento do volume de dados científicos tem-se notado especialmente na área da saúde, onde cada vez existem mais estudos e onde o acesso a este tipo de conteúdos, por parte dos profissionais é cada vez mais complicado. Um dos ramos da saúde em que esta dificuldade é mais notória é o das doenças oncológicas.

Em termos oncológicos, o cancro é uma neoplasia resultante da proliferação descontrolada de células constituintes dos vários tecidos do organismo humano. Esta proliferação descontrolada leva a uma acumulação progressiva das células alteradas, conduzindo à formação de um nódulo ou massa tumoral que pode ser visível ou palpável, quando superficial, ou apenas detetável por métodos imagiológicos. O cancro é uma patologia com origem multifatorial associada principalmente ao envelhecimento e desregulação celular decorrente da acumulação de alterações genéticas ou da expressão de alelos específicos herdados⁽¹⁷⁾. A neoplasia maligna tem um desenvolvimento estimulado pela interação entre fatores exógenos e fatores do hospedeiro ou endógenos (próprios de cada indivíduo). Por exemplo, a hereditariedade genética, o sexo, a idade e fatores exógenos como a agressão por agentes físicos, biológicos e químicos. Embora estes fatores já se encontrem bastante

documentados como tendo um papel ativo na predisposição à gênese tumoral, os seus papéis específicos ainda não foram completamente esclarecidos⁽¹⁷⁻²¹⁾.

A investigação oncológica, clínica e epidemiológica tem permitido enormes progressos na prevenção, diagnóstico e tratamento do cancro⁽²²⁾.

O diagnóstico precoce desempenha um papel essencial no, tratamento precoce, na progressão e no melhoramento da qualidade de vida dos doentes oncológicos, permitindo melhorar os resultados do tratamento. Apesar dos avanços significativos no tratamento e no diagnóstico precoce, a taxa de sobrevivência destes pacientes, não tem sofrido alterações significativas^(23,24). Aos fatores associados está reservado um papel importante no diagnóstico e prevenção de cancro. No entanto, caracterizar a influência dos fatores de risco é um desafio, uma vez que se tem vindo a verificar uma mudança de comportamento humano tanto a nível social como alimentar, que levam a alterações microbiológicas e genéticas. Assim sendo, é pertinente que os estudos científicos continuem a ser desenvolvidos abrindo novos caminhos para o combate a esta epidemia: o cancro⁽²⁵⁾.

Devido a todos os avanços no diagnóstico e terapêutica ao longo das últimas décadas, a complexidade de informação nesta área tem aumentado e o tratamento e gestão da mesma revela-se um processo extremamente moroso e complexo⁽²⁶⁾.

As ontologias podem ser utilizadas no diagnóstico de cancro, Alfonso Marco (2013), utiliza a integração de ontologias num sistema informático que ajuda no apoio à decisão e detecção de cancro da mama⁽¹⁵⁾. A construção de um sistema com base ontológica é um processo semelhante ao da construção de um programa orientado a objetos, muito embora, as classes e objetos em ontologias sejam sobre um domínio. Uma boa forma de desenvolver um processo de construção de ontologia é produzir um documento específico de ontologias escritas em linguagem natural, com informações do tipo: objetivo, públicos-alvo, relações possíveis e grau de formalidade usado para codificar a ontologia e a sua abrangência. Uma ontologia específica deve ser concisa, incluir apenas termos relevantes, significados dos termos e as suas relações que fazem sentido no domínio⁽¹⁶⁾. Atualmente inúmeras metodologias, linguagens e ferramentas são utilizadas para apoiar todo o processo de construção de uma ontologia⁽¹⁶⁾.

Para a construção de uma ontologia, em primeiro investiga-se e junta-se toda a informação referente à ontologia, depois definem-se as classes, que são formas de

organização e domínios e são os conceitos característicos mais adequados que devem ser integrados no sistema desenvolvido. A Figura 1 retrata uma metodologia geral

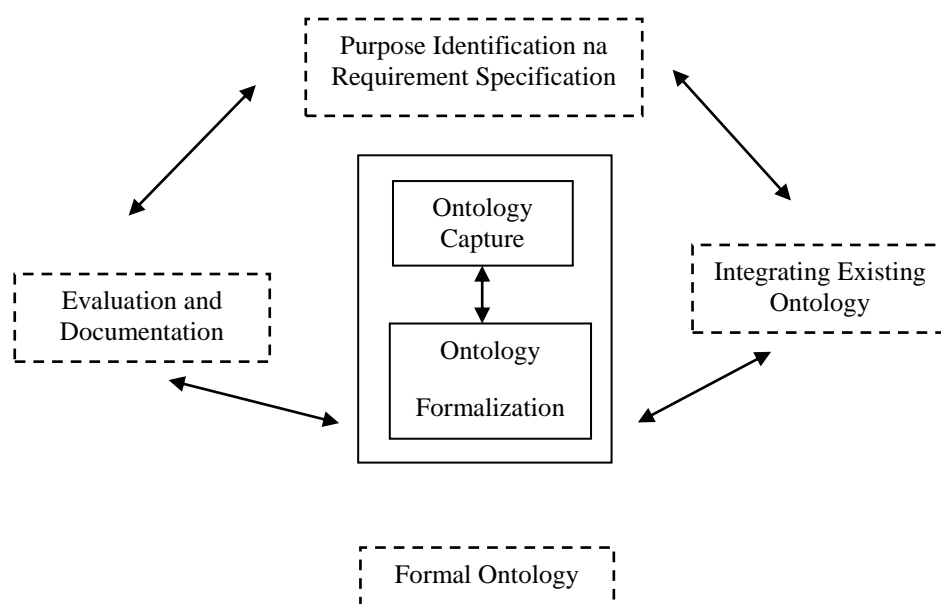


Figura 1 - Metodologia geral para construir uma ontologia. Adaptado de Alfonse *et al*⁽¹⁶⁾.

Sendo uma patologia multifatorial complexa, que atinge uma fração significativa da população mundial, o cancro é o principal alvo dos esforços multidisciplinares utilizando estes novos e poderosos meios tecnológicos, as ontologias⁽²⁷⁾. No entanto, a falta de uma infraestrutura de informação aberta e compartilhada com toda a comunidade científica impede as instituições de pesquisa científica/clínica de serem capazes de reutilizar dados das diferentes fontes de dados⁽²⁷⁾. O isolamento de várias especialidades envolvidas na compreensão e tratamento do cancro torna-se assim numa das principais dificuldades no que diz respeito à utilização de ontologias na pesquisa biomédica⁽²⁸⁾.

Em suma, a presente dissertação tem como objetivo a implementação de uma metodologia e o desenvolvimento de uma base de dados com princípios ontológicos capaz de catalogar, organizar, inferir e correlacionar, dados heterogéneos obtidos pelo conhecimento científico a partir da literatura disponível, para análise e gestão de informação, nomeadamente relativa a fatores associados a cancro. O objetivo passa por desenvolver uma forma de integração e reutilização de informação entre a comunidade científica, através do desenvolvimento de uma estrutura ontológica que permita a catalogação de múltiplos dados sobre fenómenos complexos, tais como,

fatores associados a doenças, modificações genéticas associadas a cânceros, fatores de risco extrínsecos presentes na rotina social que se associem a doenças/cânceros, entre outros. Com a revisão de literatura desenvolveu-se uma estrutura de base de dados referente a fatores associados a cancro, onde ficou evidente a grande dispersão de informação e ambiguidade de conceitos existentes. Surge assim a necessidade de utilização de princípios ontológicos no projeto desenvolvido, de modo a preencher esta lacuna.

A.2 Requisitos para a construção de ontologias relacionadas com fatores associados a cancro

Com o desenvolvimento da *web*, o processo de divulgação da informação científica foi substancialmente alterado. Na atualidade, as publicações científicas ficam disponíveis em todo o mundo em segundos e o aparecimento de sistemas de bases de dados e de redes que permitem o armazenamento e reutilização de praticamente todos os dados obtidos em ensaios, alteraram por completo o paradigma científico, permitindo o acesso a materiais mais diversificados e em muito maior quantidade. Os recentes avanços na biologia molecular, genética, bioquímica e outras áreas têm sido estimulados em larga escala por esta disponibilidade de dados, assim como pelas tecnologias mais recentes cada vez mais automatizadas. Ainda que, há algumas décadas atrás a maior parte do trabalho experimental era realizado de forma manual, hoje em dia, este pode ser executado a partir de dispositivos computadorizados, tornando todo o processo mais rápido. Todas estas novas abordagens levaram a uma diminuição do tempo necessário para o desenvolvimento científico, aumentando as publicações científicas e a partilha de conhecimento⁽²⁾.

Neste capítulo serão abordados quais os processos que levaram a esta disponibilidade, acessibilidade e reutilização de informação entre toda a comunidade científica.

A.2.1 Web Semântica

A pesquisa biomédica evoluiu para uma ciência geradora intensiva de dados, produzindo quantidades extraordinárias de informação a qual pode ser integrada a partir de recursos distintos⁽²⁹⁾. A análise dos dados e posterior avaliação crítica completa o processo de aquisição do conhecimento. Em domínios como a medicina translacional, onde vários tipos de dados, provenientes de diferentes fontes e em diferentes formatos, estão envolvidos, a integração, reutilização e interoperabilidade de dados são requisitos fundamentais para uma análise eficiente⁽³⁰⁻³²⁾.

Em 1995, Davidson *et al.* questionou a viabilidade da integração de dados, já que a estrutura dos mesmos deveria acompanhar a evolução científica^(33,34). Louie *et al.*, em 2007, identificou alguns desafios para a integração de dados em genómica.

A falta de fontes de dados clínicos e problemas com a privacidade associados aos mesmos, são alguns dos desafios que mostram claramente que a integração de dados permanece uma área de pesquisa em aberto, aumentando de complexidade à medida que aumenta o número de domínios heterogéneos a serem integrados⁽³⁵⁾.

Tim Berners-Lee *et al.* refere que a WS é uma extensão da Web atual que permitirá uma cooperação entre os sistemas de informação (SI) e a comunidade científica. Segundo este autor, a WS interliga significados de palavras e tem como finalidade atribuir um significado aos conteúdos publicados pela comunidade científica na Web, de modo a torná-las perceptíveis⁽³⁾. Com a WS os SI terão uma maior capacidade para compreender a informação depositada na Web, facilitando a procura, filtragem e categorização da informação, conduzindo a sistemas capazes de produzir novo e melhor conhecimento.

A.2.1.1 Arquitetura da Web Semântica

A web é um extensíssimo repositório de dados e conhecimento, e as tecnologias da web semântica suportam a interação da interoperabilidade entre as máquinas. O World Wide Web Consortium (W3C), criado em 1994, é a principal organização internacional que conduz à evolução da World Wide Web (WWW) até ao seu máximo potencial, a partir do desenvolvimento de protocolos que promovam esta evolução e assegurem a sua interoperabilidade. Os seus principais objetivos são:

- Construção de uma Web para todos: o valor social da web é permitir a comunicação humana e a oportunidade de partilhar conhecimento, superando barreiras de língua, culturas, *softwares*, *hardwares*, localização geográfica entre outros.
- Web em tudo: acesso à web a partir de qualquer tipo de dispositivo. Tão simples, fácil e conveniente como o acesso à web a partir de um *desktop*.
- Web como base de conhecimento: permitindo a resolução de problemas que de outra forma seriam muito complexos de se resolverem.
- Web como meio útil para transações sociais. Novas tecnologias (W3C) devem permitir transações seguras, onde a responsabilização, confiança e confidencialidade, sejam um ponto primordial⁽³⁶⁾.

É neste contexto que surge o conceito da arquitetura da Web Semântica, ilustrada pela Figura 2.

A WS é uma extensão da web atual e apresenta uma arquitetura em camadas, utilizando padrões e tecnologias de forma hierárquica. Esta arquitetura por camadas define as linguagens e tecnologias usadas para um perfeito funcionamento da Web. Este tipo de abordagem assume que os níveis superiores utilizam as capacidades dos níveis inferiores, permitindo que os níveis inferiores possam ser aplicados de modo estável, mesmo que os níveis superiores ainda estejam a ser formulados⁽³⁷⁾.

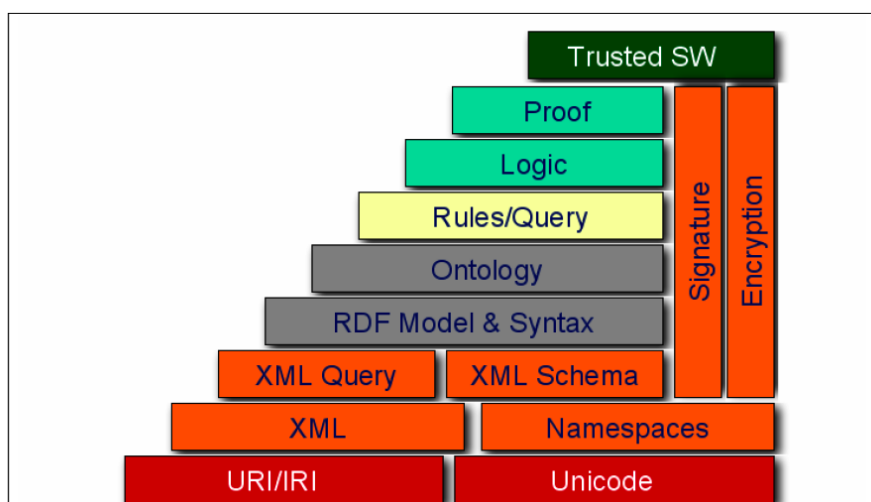


Figura 2 - Arquitetura da Web Semântica. Adaptado de Sigmore⁽³⁷⁾.

A ideia de organizar o conhecimento de forma estruturada é pelo menos tão antiga quanto o próprio conceito do conhecimento. Aristóteles, focando-se nesta ideia, procurou uma forma sistemática para o problema prático de representar a estrutura da realidade e, embora a filosofia tenha acumulado ferramentas analíticas para os problemas ontológicos, ainda assenta na base aristotélica. Muitas das ideias e termos ontológicos, como as noções de categoria e hierarquia, tem origem na filosofia de Aristóteles⁽³⁸⁾, como ilustrado pela Figura 3.

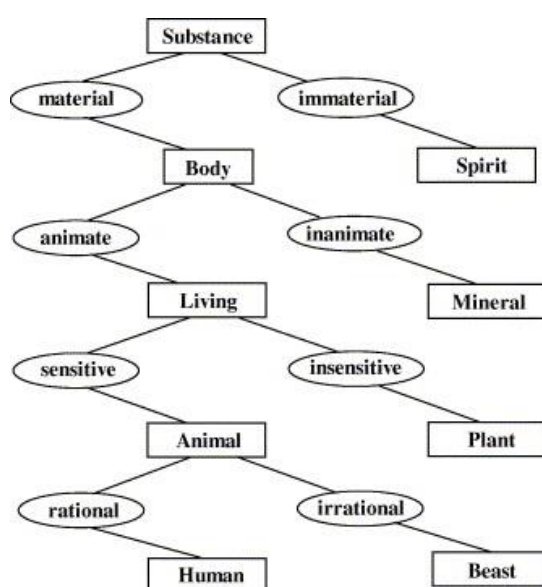


Figura 3 - Árvore de Porfírio, com as categorias de Aristóteles (em retângulos). As linhas representam relações entre categorias. Adaptado de (AlexanderYu 2006)⁽³⁶⁾.

O termo ontologia deriva do grego “*ontos*”, ser, e “*logos*”, palavra, e define estruturas organizadas a partir de conceitos e da interligação entre os mesmos. As ontologias melhoram a organização da informação através da sua uniformização e padronização, permitindo fazer representações lógicas que possibilitam a recolha de citações⁽³⁹⁾.

Na literatura encontram-se inúmeras definições para o termo ontologias. Estas definições variam consoante o contexto, as propostas de aplicação em diferentes áreas de conhecimento e a maneira de construção da ontologia, quer ao nível das metodologias, quer em termos de ferramentas e linguagens. Tal diversidade tem dificultado a escolha e a utilização das técnicas disponíveis para a manipulação de ontologias na organização de informação⁽¹⁾.

Nos SI não há qualquer acordo quanto ao significado de uma ontologia. Estas são geralmente consideradas artefactos de representação, não existindo clareza ou consenso quanto às condições necessárias para a aplicação do termo. Os significados atribuídos ao termo ontologia são muitas vezes conflituosos⁽²⁾.

A definição de ontologia encontrada mais frequentemente na literatura da WS é a proposta por Gruber (1993), “*A specification of a representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and the objects — is called na ontology. An ontology is an explicit specification of a conceptualization*”⁽⁴⁰⁾.

As ontologias definem os termos e as relações básicas que compõem uma terminologia de uma área específica, bem como regras de termos e relações para definir extensões dessa terminologia. Elas concedem a base para a interoperabilidade entre sistemas. São também utilizadas como uma base para a integração de fontes de informação e como um modelo para a consulta de fontes de informação, auxiliando a comunidade científica no desenvolvimento e manutenção de ontologias⁽⁴¹⁾.

Por outro lado, segundo Noy N.F., McGuinness D. L., as ontologias são igualmente usadas para construir bases de conhecimento. Uma base de conhecimento é formada por uma ontologia e todo um conjunto de instâncias individuais das suas classes. As bases de conhecimento podem ser consultadas pela comunidade científica de forma a enriquecerem, reutilizarem e guardarem todos os dados⁽⁴⁾.

Outra das definições de ontologia caracteriza-a como sendo um sistema que utiliza uma terminologia específica para descrever entidades, classes, propriedades e funções relacionadas com um determinado assunto⁽⁴²⁾, que no caso da presente investigação será referente a fatores associados a cancro.

Para a construção de uma ontologia, além das principais funções descritas, é necessário ponderar as razões que levam à sua construção:⁽⁴⁾

- Partilha de dados comuns em estruturas de informações entre as pessoas ou agentes de *software*;
- Reutilização do conhecimento de domínio;
- Verificação de hipóteses de domínio explícito;
- Separação do conhecimento de domínio do conhecimento operacional;
- Análise do conhecimento de domínio;

Segundo Noy N.F., McGuinness D. L. (2001), o desenvolvimento de uma ontologia inclui fases como a elaboração e definição de classes, a organização numa estrutura hierárquica, a especificação de *slots* e a atribuição de valores válidos⁽⁴⁾.

Gómez (1998) cita outra abordagem para o desenvolvimento de uma ontologia defendendo o seguinte conjunto de princípios^(43,44):

- Clareza e objetividade: uma ontologia deve fornecer definições precisas e objetivas em forma de linguagem natural;
- Integralidade: As definições devem ser expressas por condições necessárias e suficientes;
- Coerência: Uma ontologia deve permitir fazer inferências consistentes com as definições;
- Inclusão de novos termos gerais ou específicos devem ser incluídos na ontologia sem revisão das definições existentes;
- Mínimo compromisso ontológico: Poucos axiomas quanto possível sobre o mundo modelado;
- Diferenciação do princípio ontológico: Classes com diferentes critérios de identidade devem ser separadas;

Alguns estudos têm em consideração diferentes critérios no desenvolvimento da ontologia, como por exemplo a formalidade da língua ou o nível de dependência de uma tarefa. Guarino (1998) considera o último e identifica o último e identifica os quatro tipos básicos de ontologias^(43,45):

- Ontologias gerais ou de nível superior: descrevem conceitos gerais como espaço, tempo, substância, objeto, ocorrência ou ação, que não dependem de um problema ou domínio particular. No entanto, o desenvolvimento deste tipo de ontologias ainda não foi realizado;
- Ontologias de domínio: descrevem um vocabulário próprio de um dado domínio (como a biologia ou medicina);

- Ontologias de tarefa: fornecem um vocabulário sistematizado de termos relacionados à execução de uma tarefa específica, independente do domínio em questão;
- Ontologias de aplicação: contêm as definições necessárias à aplicação de uma tarefa num dado domínio;

Para iniciar todo o processo de construção de uma ontologia, é crucial identificar o tipo de ontologia de acordo com uma classificação particular⁽⁴³⁾.

A.2.4 Bio-ontologias

Atualmente a aplicação de tecnologias e *softwares* direcionados para grandes volumes de dados genéticos, fisiológicos e patológicos, resulta num crescimento exponencial de informações. Lidar corretamente com esta informação é crucial no entendimento dos mecanismos moleculares de doenças multifatoriais, como o cancro. Só assim é possível conseguir avanços significativos em áreas como a fisiopatologia clínica, farmacoterapia, abordagens diagnósticas e mecanismos de prevenção⁽⁴⁶⁾.

As ontologias do domínio da biologia e da biomedicina designam-se por Bio-ontologias e são formadas por conceitos biológicos e relações entre eles. A sua aplicação mais comum é a anotação baseada em ontologia de dados básicos, que consiste nos dados da associação de elementos da ontologia (isto é, conceitos e relações), para que as ontologias possam ser utilizadas por toda a comunidade científica, a partir de sistemas de informação e onde se possa pesquisar e navegar⁽⁴⁷⁾.

As bio-ontologias transmitem o significado biomédico de estudos científicos, num formato acessível aos sistemas de informação, e permitem a integração de dados e conhecimento provenientes de várias fontes. A conjectura atual mudou drasticamente a forma como a biomedicina constrói as suas ontologias. Em primeiro lugar, como as ontologias se tornaram uma corrente comum dentro de biomedicina, desenvolvem-se de forma colaborativa entre cada vez mais grupos da comunidade científica. Em segundo lugar, as ontologias estão a tornar-se de tal forma extensas (por exemplo, os conceitos *Thesaurus NCI* que já abrange 103670 Classes), que não permitem a

organização eficiente sem recorrer a organizações como o Centro de NCI de Bioinformática (NCICB)⁽³⁸⁾.

As Bio-ontologias são cada vez mais utilizadas em várias aplicações de bioinformática, que vão desde a pesquisa, à anotação semântica e análise de dados em larga escala. Para representar este tipo de ontologia, a seguinte lista contém apenas uma pequena amostra de ontologias biomédicas notáveis que utilizam algum processo colaborativo estruturado para o seu contínuo desenvolvimento⁽⁴⁸⁾:

- O *Gene Ontology (GO)* oferece uma terminologia para a unificação da representação de genes e seus produtos em diferentes espécies numa base de dados⁽⁴⁹⁾;
- *National Cancer Institute's Thesaurus (NCI Thesaurus)* é uma ontologia de referência biomédica, que abrange as áreas de biologia básica do cancro, da ciência translacional e oncologia clínica desenvolvida pelo *NCI Center for Bioinformatics (NCICB)*⁽⁵⁰⁾;
- *Ontology for Biomedical Investigations (OBI)* é uma ontologia que descreve as experiências biológicas e médicas: (<http://obi-ontology.org>);
- *International Classification of Diseases (ICD)* é uma ferramenta de diagnóstico padrão da Organização Mundial de Saúde (OMS) em epidemiologia, gestão da saúde e para fins clínicos. Esta inclui a análise da situação geral de saúde de grupos populacionais. Ela é usada para monitorizar a incidência e prevalência de doenças e outros problemas de saúde: (<http://www.who.int/classifications/icd/en/>);
- *Medical Subject Headings (MeSH)* é uma terminologia criada pela *National Library of Medicine* para a indexação da literatura⁽⁵¹⁾. *MeSH* faculta um conjunto padrão de termos que as bibliotecas médicas utilizam para descrever os principais temas abordados na literatura. Originalmente concebido para ajudar a comunidade científica a realizar uma pesquisa bibliográfica, os terminologia padrão fornecida pelo *MeSH* provaram ser extremamente úteis para aumentar os métodos de processamento de linguagem natural para o processamento de texto, extração e classificação⁽⁵²⁻⁵⁴⁾;

A.2.5 Linguagens de desenvolvimento de uma ontologia

Recentemente, membros de diferentes comunidades realizaram experiências para o protótipo da WS com diferentes origens e objetivos em mente. Criaram uma infinidade de linguagens para representar as ontologias e dados de instância na Web como o *Hyper Text Markup Language (HTML)*, *Resource Description Framework (RDF)* e *Ontology Interchange Language* como representado pela Figura 4. Normalmente, uma linguagem Web Semântica serve para descrever ontologias e dados de instância (que é o preenchimento dos conceitos necessários para a criação de ontologias) que contém uma descrição hierárquica de conceitos importantes num domínio, intituladas de classes⁽⁵⁵⁾.

ACRONYM	LANGUAGE	CHARACTERISTICS
HTML	Hyper Text Markup Language	Simplicity.
XML	Extensible	Extensions for arbitrary domains and specific tasks.
SHOE	Simple HTML Ontology Extensions	It is a XML compatible knowledge representation language for the web. It allows page authors to annotate their web documents. It is not actively maintained.
RDF	Resource Description Framework	Syntactic conventions and simple data models to represent semantics. It supports interoperability aspects with object-attribute-value relationships.
RDFS	Resource Description Framework Schema	Primitives to model basic ontologies with RDF.
OIL	Ontology Inference Layer/ Ontology Interchange Language	Primitives to model ontologies from frame-based languages, formal semantics and reasoning support based on descriptive logic, a proposal for syntactic interchange of annotations. It is compatible with RDF Schema [Fenselet al. 2000].
DAML	DARPA Agent Markup Language	It is formed by DAML-ONT, (a language of ontologies) and DAML-Logic (a language able to express axioms and rules). It inherits many characteristics of OIL, however, it is less compatible with RDF than OIL.
XSL	Extensible Stylesheet Language	It provides a standard to describe mappings between different terminologies, (a translation mechanism between XML documents).
XOL	Ontology Exchange Language	Simplicity, a generic approach to define ontologies. It has two syntactical variants based on XML and RDF Schema.

Figura 4 - Algumas linguagens de apoio à gestão de ontologias. Adaptado de María Neto⁽⁴¹⁾.

A partilha de informação com o mesmo significado entre sistemas de informação e a comunidade científica requer a existência de uma linguagem semântica padrão. Ao longo da última década, a comunidade web semântica, e em

particular a *World Wide Web Consortium* (W3C), têm vindo a desenvolver um conjunto de linguagens base de forma a atingir uma terminologia comum. Algumas dessas tecnologias tornaram-se padrões para a web semântica. A W3C promoveu o aparecimento de novas linguagens como XML, XML *Schema*, *Resource Description Framework* (RDF), *RDF Schema*, SPARQL, SKOS (*Simple Knowledge Organization System*) e OWL, que vêm permitir a interoperabilidade e cooperação entre os sistemas informáticos e a comunidade científica⁽³⁴⁾.

Segundo Staab, *et al.* (2009), estas linguagens ontológicas permitem uma escrita clara, de conceitos e terminologias e têm como principais requisitos uma sintaxe e semântica bem definidas^(37,56). Segundo Custódio (2012), “A OWL é a linguagem mais recente e completa, pois captura os diferentes conceitos e relações que ocorrem nas ontologias. As ontologias criadas com a OWL permitem representar explicitamente a semântica exata de classes, das suas instâncias e propriedades, dentro de um mesmo domínio. A OWL é uma extensão do RDF/RDFS e tem um poder expressivo semântico maior do que o RDF e RDFS”⁽³⁷⁾.

A.2.6 Aplicações Semânticas na área da Biomedicina

Com o aumento exponencial do número de ontologias biomédicas, o mesmo acontece com o número de bibliotecas que as organizam. Algumas bibliotecas pesquisam pela web para recolha de ontologias, como o *Swoogle*, o *Watson* e o *OntoSelect*. Outros solicitam as suas próprias ontologias como a biblioteca DAML e *SchemaWeb*, estando limitados ao armazenamento de ontologias com base em linguagem de representação ontológicas^(57,58).

O *Bioportal* é uma das aplicações mais conhecidas no domínio das ontologias. Permite o acesso às ontologias biomédicas mais comuns, disponibiliza ferramentas para trabalhar com elas, permite navegar na biblioteca de ontologias, pesquisar um termo em várias ontologias, navegar mapeamentos entre termos de diferentes ontologias, anotar texto com termos de ontologias, pesquisar recursos biomédicos para um dado termo, entre outros⁽⁵⁹⁾.

O *Bioportal*, desenvolvido pelo *National Center for Biomedical Ontology* (NCBO) Figura 5, também permite a pesquisa integrada de recursos de dados

biomédicos, como o *Gene Expression Omnibus* (GEO), *ClinicalTrials.gov* e *ArrayExpress*, por meio de anotação e indexação desses recursos como ontologias. Assim, este recurso, não só fornece à comunidade científica o acesso a ontologias biomédicas, mas também fornece suporte para integrar dados de outros recursos.

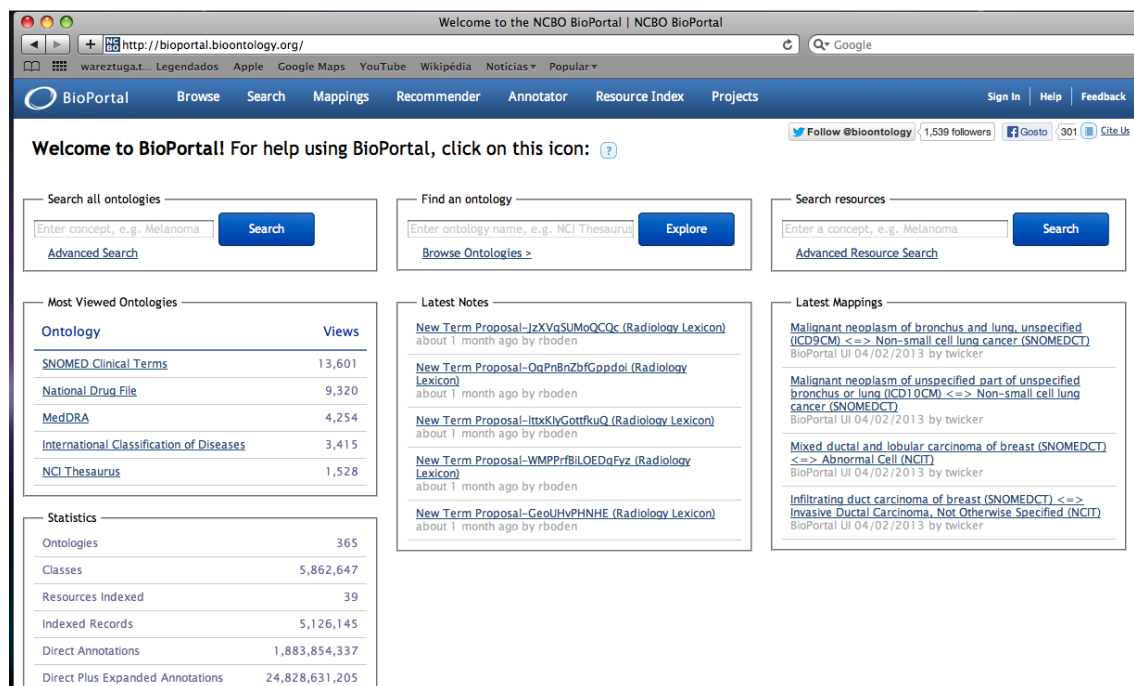


Figura 5 - Página web da biblioteca Bioportal.

O *Bioportal* é uma biblioteca aberta de ontologias biomédicas que armazena ontologias desenvolvidas em vários formatos. Prevê atualizações automáticas por submissão de novas versões pelo utilizador, e que permite o acesso através de navegadores da Web e através de *Web Services*. No *Bioportal* os utilizadores inscritos podem navegar e procurar as ontologias, apresentar novas versões de ontologias no seu repositório, comentar sobre qualquer ontologia (ou parte dela), adicionar uma revisão da ontologia, descrever a sua experiência na utilização de uma ontologia ou fazer sugestões para o desenvolvimento de novas ontologias. O fato de permitir que os membros da comunidade possam contribuir ativamente para o conteúdo, do *Bioportal*, aumentando assim o valor do conteúdo distingue-se de uma forma positiva de outros repositórios de ontologias⁽⁵⁹⁾.

Em 2008, *Bioportal* (<http://Bioportal.bioontology.org/>) continha 72 ontologias (300 000 classes) e tem crescido significativamente, já que em 2011 continha 260 ontologias (4,8 milhões de classes). Vários grupos publicam as suas ontologias no

Bioportal, incluindo *caBIG*, *Clinical and Translational Science Awards*, *Consultative Group on International Agricultural Research*, *OBO library*, a *Proteomics Standards Initiative*, o Sistema *Unified Medical Language System (UMLS)* e a *World Health Organization (WHO) Family of International Classifications*⁽⁶⁰⁾

As ontologias biomédicas proporcionam um conhecimento de domínio essencial para a integração de dados, recuperação, reutilização e processamento de informação e que fornecem um instrumento de apoio à decisão. No entanto, falta-nos a interface entre estes conceitos e a sua integração via serviços Web. Estes conceitos/ontologias foram transcritos para uma linguagem (OWL e RDF, formato OBO) para que fossem reconhecidos na web (como referenciado anteriormente) utilizando a ferramenta *Protégé*⁽³⁷⁾.

B.1 Sistema de classificação de neoplasias.

O termo “cabeça e pescoço” na classificação anátomo-topográfica de cancros, é utilizado maioritariamente para descrever os tumores malignos do trato aerodigestivo superior, envolvendo as regiões anatómicas da cavidade oral, faringe e laringe. O subgrupo major de carcinomas da “cabeça e pescoço” situa-se na cavidade oral, abrangendo as membranas da mucosa oral como os lábios, a língua em toda a sua extensão, as gengivas, o soalho da boca, o palato duro e mole, assim como a faringe, que compreende a orofaringe, hipofaringe e nasofaringe. Outros tumores que ocorrem nesta área, como os do cérebro, tiroide e melanomas não são convencionalmente incluídos quando aplicamos o termo “cancro da cabeça e pescoço”. No entanto, esta discrepância de termos leva a alguma discórdia na comunidade científica⁽⁶¹⁾. Na presente dissertação considera-se o termo “cancro da cabeça e pescoço” útil de acordo com a semelhança histológica e fisiológica (a nível de continuidade) no complexo “cabeça e pescoço” e será este o termo utilizado.

O Sistema de Classificação das patologias neoplásicas está organizado num sistema codificado determinado pela *World Health Organization (WHO)* e utiliza a maior base de dados internacional de classificação de patologias, a *Internacional Classification of Diseases (CID)*. Este sistema consiste numa ferramenta de diagnóstico padrão para a epidemiologia e gestão da saúde, sendo também útil para

fins clínicos. Esta ferramenta inclui a análise da situação geral de saúde de grupos populacionais específicos e é utilizada para monitorizar a incidência e prevalência de doenças assim como outros fatores associados à saúde. A codificação determina que a CID-O é uma classificação de eixo duplo, um sistema comum de codificação tanto para topografia como para a morfologia. O código topográfico utilizado para a maioria das neoplasias é o mesmo utilizado para neoplasia maligna CID-10 e dentro dele a codificação das neoplasias vai de C00 a D48. A título de exemplo, os tumores da cavidade oral e orofaringe incluem as zonas anatómicas como o lábio, língua, cavidade oral [CID-10: C00-06], e orofaringe [CID-10: C09-C10], mas exclui os tumores das glândulas salivares [C07-08] e outros locais faríngeos de [C11-13]^(20,62).

B.2 Fatores etiológicos em cancro

Em epidemiologia, fator de risco refere-se a um agente com a capacidade de aumentar a probabilidade de um indivíduo saudável de desenvolver determinada patologia⁽⁶³⁾.

Os fatores de risco associados ao cancro oral ainda não se encontram claramente estabelecidos, contudo, é possível apontar fatores exógenos e endógenos com capacidade para promover o desenvolvimento de patologia cancerígena. Existem inúmeras citações que referem uma elevada correlação entre o cancro e o fumo do tabaco (carcinogénico completo por ser simultaneamente indutor e promotor da carcinogénese). No entanto, são muitos os fatores que podem agir de uma forma sinérgica na iniciação e promoção da carcinogénese. Fatores químicos como o álcool, fatores biológicos como o vírus do papiloma humano (HPV), fatores oro-dentários, deficiências alimentares, candidíase crónica e infeção viral foram também identificados como fatores que contribuem para o desenvolvimento do cancro oral^(21,64-66). Os hábitos tabágicos e alcoólicos estão envolvidos em 75% de todos os carcinomas das células escamosas da cabeça e pescoço, (SCCHN) e possuem um efeito sinérgico quando combinado⁽⁶⁷⁾. Estes fatores são igualmente partilhados com outros tumores⁽⁶⁵⁾, verificando-se que este processo carcinogénico pode decorrer num período de 10 a 20 anos desde o início do contacto com o agente, até que as lesões atinjam um tamanho que possibilite a sua identificação clínica⁽¹⁹⁾.

Capítulo III

Objetivos

3. OBJETIVOS

Uma vez que se verificou que não existia ainda uma base de dados com a catalogação dos fatores com implicação em cancro, com a presente dissertação pretende-se desenvolver a estrutura de uma base de dados referente a fatores associados a cancro.

O trabalho delineado para a presente dissertação tem como objetivos:

1. Revisão bibliográfica como precursora da criação da Base de Dados.
2. Criação de ontologias descritoras para fatores associados a cancro;
3. Desenvolvimento de uma base de dados referente a fatores associados a cancro;
4. Estabelecimento de uma metodologia para consulta da base de dados criada.

Materiais e Métodos

4. MATERIAIS E MÉTODOS

4.1 Construção da base de dados - *FactorXcancer*

Visando a estruturação de uma base de dados que permitisse a subsequente catalogação dos fatores associados a cancro, foi desenhada uma pesquisa bibliográfica na base de dados *online* Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed>- Figura 6), de forma a identificar o maior número de descritores aplicados em todos os tipos de cancro.

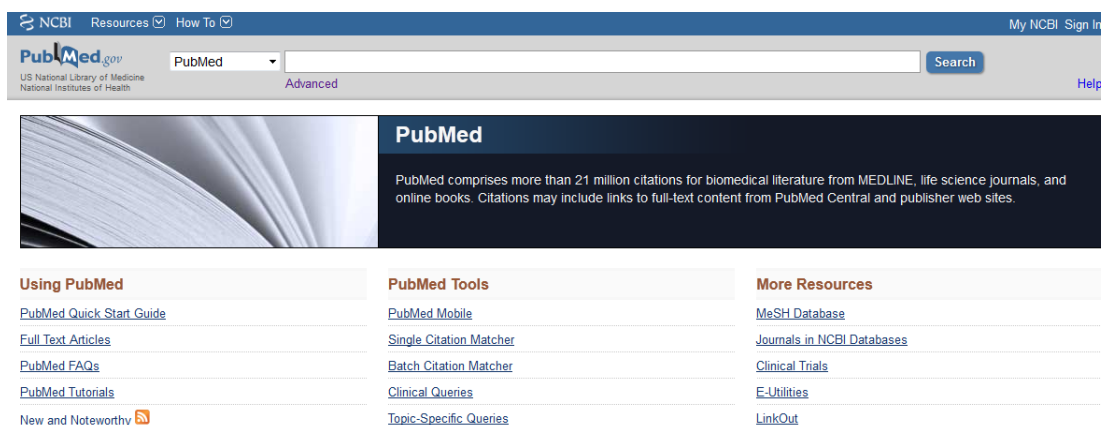


Figura 6 - Página web da base de dados Pubmed.

4.1.2 Metodologia aplicada na pesquisa da bibliografia referente a fatores associados a cancro

A pesquisa bibliográfica foi conduzida utilizando de forma inclusiva os seguintes termos: “*Risk factors of oral carcinoma*”, “*oral cavity cancer risk*”; “*oropharyngeal cancer risk*”; “*oral squamous cell cancer risk*”; “*oral cancer risk*”; “*oral carcinogenesis*”; “*head and neck cancer*”; “*head and neck neoplasms*”; “*risk factors of cancer*”; “*polymorphism and cancer*”; “*genetic in oral cancer*”; “*genetic susceptibility of cancer*”; “*dietary factors in oral cancer*”; “*etiology of oral cancer*”; “*human papillomavirus and oral cancer*”; “*infections and cancer*”; “*risk factors of cervical cancer*”; “*etiology of cancer*”.

Como critérios de filtragem foram definidos: artigos publicados entre Janeiro de 2000 até Janeiro de 2012. Foram selecionados os últimos 12 artigos publicados em cada ano, como mostram as Tabelas I e II, página 42. A Tabela I representa o número de artigos consultados por ano de publicação e o tipo de cancro identificado. A Tabela II representa o número de artigos consultados por tipo de cancro identificado. Pretendeu-se também apreciar a evolução da terminologia aplicada para descrever fatores associados a cancro e para isso foi estabelecida comparação entre a bibliografia mais e menos recente. Neste sentido, procedeu-se à leitura de bibliográfica mais antiga, referente ao intervalo de tempo entre 1975 – 1999. É de salientar que sempre que foi necessário informação adicional sobre determinado fator, a pesquisa foi alargada para além do intervalo temporal previamente estipulado.

A Tabela I representa o número de artigos consultados por ano de publicação, com indicação do tipo de cancro identificado em cada ano.

TABELA I		
Ano de Publicação	Nº artigos	Tipo de cancro
1975 - 1999	26	Oral ⁽⁷⁰⁻⁸⁸⁾ ; Faríngeo ^(71,73,79,80,83-87) ; Cabeça e pescoço ^(74,78,82,88) ; Cavidade Nasal ⁽⁸⁴⁾ ; Pancreático ^(88,89) ; Cólon ^(73,78,90,91) ; Reto ^(73,78,90,91) ; Laríngeo ^(77-79,83) ; Pulmão ^(75,78) ; Bexiga ^(75,77,78) ; Gástrico ^(77,78) ; Mama ^(76,92) ; Orofaríngeo ⁽⁷³⁾ ; Pénis ⁽⁸⁰⁾ ; Pele ⁽⁷⁸⁾ ; Esofágico ⁽⁸³⁾ ; Melanoma ⁽⁷⁸⁾ ; Cervical ^(73,76,80,93-96) ; Papilomatose respiratória recorrente ⁽⁷⁸⁾ .
2000	11	Oral ⁽⁹⁷⁻⁹⁹⁾ ; Faríngeo ^(97,98) ; Mama ^(100,101) ; Pele ⁽¹⁰²⁾ ; Linfoma Non-Hodgkin ⁽¹⁰³⁾ ; Linfoma Hodgkin ^(103,104) ; Linfoma Burkitt ⁽¹⁰³⁾ ; Linfoma B-ceel ⁽¹⁰³⁾ ; Hiperplasia do gânglio linfático gigante ^(103,104) ; Leucemia das células T adultas ^(103,104) ; Leiomyosarcoma ⁽¹⁰³⁾ ; Ânus ⁽¹⁰³⁾ ; Fígado ^(101,103,104) ; Gástrico ⁽¹⁰³⁾ ; Cólon ^(101,105) ; Reto ^(101,105) ; Nasofaríngeo ^(103,104) ; Ovários ⁽¹⁰¹⁾ ; Mesotelioma ⁽¹⁰⁴⁾ ; Cérebro ⁽¹⁰⁴⁾ ; Osteosarcoma ⁽¹⁰⁴⁾ ; Orofaríngeo ^(102,106) ; Pulmão ⁽⁹⁹⁾ ; Renal ⁽¹⁰¹⁾ ; Cabeça e pescoço ^(103,106,107) ; Cervical ^(102,103) ; Sarcoma de Kaposi ⁽¹⁰³⁾ ; Vulvar ^(102,103) ; Pénis ^(102,103,106) ; Ano-genital ^(102,106) ; Pancreático ⁽¹⁰¹⁾ ; Próstata ^(101,108) ; Endométrio ⁽¹⁰¹⁾ ; Vesícula Biliar ⁽¹⁰¹⁾ .
2001	11	Oral ⁽¹⁰⁹⁻¹¹⁵⁾ ; Orofaríngeo ^(110,114,115) ; Laríngeo ⁽¹¹²⁾ ; Faríngeo ⁽¹¹²⁾ ; Esofágico ^(113,116-118) ; Gástrico ⁽¹¹⁶⁾ ; Cervical ^(110,115) ; Ano-genital ^(110,115) ; Cabeça e pescoço ^(112,119,120) .
2002	10	Oral ⁽¹²¹⁻¹²⁵⁾ ; Cólon ^(126,127) ; Reto ^(126,127) ; Fígado ^(126,127) ; Bexiga ^(126,127) ; Gástrico ^(126,128) ; Pele ⁽¹²⁷⁾ ; Pulmão ⁽¹²⁷⁾ ; Mesotelioma ⁽¹²⁷⁾ ; Orofaríngeo ⁽¹²¹⁾ ; Hipofaríngeo ⁽¹²¹⁾ ; Mama ⁽¹²²⁾ ; Esofágico ^(121,128,129) ; Faríngeo ⁽¹²²⁾ ; Laríngeo ⁽¹²⁵⁾ ; Cervical ^(130,131) ; Génito-Urinário ⁽¹²⁷⁾ .
2003	12	Oral ⁽¹³²⁻¹³⁶⁾ ; Cólon ⁽¹³²⁾ ; Reto ⁽¹³²⁾ ; Mama ^(132,137) ; Pulmão ^(132,134) ; Bexiga ^(132,134) ; Gástrico ^(132,134) ; Esofágico ^(132,134,137-140) ; Pancreático ^(134,141) ; Orofaríngeo ^(133,136,142) ; Laríngeo ⁽¹³⁴⁾ ; Cavidade nasal ⁽¹³⁴⁾ ; Renal ⁽¹³⁴⁾ ; Fígado ^(134,138) ; Pénis ⁽¹³³⁾ ; Cabeça e pescoço ^(136,138,142) ; Leucemia ⁽¹³⁸⁾ ; Faríngeo ⁽¹³⁵⁾ ; Hipofaríngeo ⁽¹³⁶⁾ ; Cervical ^(133,137,138,143) .
2004	13	Oral ⁽¹⁴⁴⁻¹⁵⁰⁾ ; Cólon ^(145,147,150,151) ; Reto ^(145,147,150,151) ; Faríngeo ⁽¹⁴⁵⁾ ; Orofaríngeo ^(145,150) ; Laríngeo ^(145,150) ; Pancreático ^(145,150,152) ; Pele ⁽¹⁵³⁾ ; Linfoma Non-Hodgkin ^(145,152) ; Leucemia das células T adultas ^(145,152,153) ; Ânus ⁽¹⁴⁵⁾ ; Mama ^(145,147) ; Pénis ^(145,152) ; Pulmão ^(145,150) ; Fígado ^(145,152,153) ; Bexiga ^(145,150) ; Gástrico ^(145,150,152) ; Vesícula Biliar ⁽¹⁴⁵⁾ ; Leucemia ^(145,150) ; Nasofaríngeo ^(150,153) ; Linfoma de Hodgkin ^(152,153) ; Linfoma Burkitt ^(152,153) ; Linfoma B-ceel ⁽¹⁵²⁾ ; Hiperplasia do gânglio linfático gigante ^(152,153) ; Renal ^(145,150) ; Esofágico ^(147,150,154,155) ; Linfoma Imunoblástico ⁽¹⁵³⁾ ; Cérebro ⁽¹⁵²⁾ ; Cervical ^(145,146,150,152,153) ; Ano-genital ⁽¹⁵³⁾ ; Mesotelioma ⁽¹⁵²⁾ ; Sarcoma de Kaposi ^(152,153) ; Hipofaríngeo ⁽¹⁵⁰⁾ ; Cavidade Nasal ⁽¹⁵⁰⁾ ; Génito-urinário ⁽¹⁵²⁾ ; Cabeça e pescoço ⁽¹⁵⁶⁾ .
2005	13	Oral ⁽¹⁵⁷⁻¹⁶²⁾ ; Faríngeo ^(157,158) ; Cabeça e pescoço ^(163,164) ; Orofaríngeo ^(159,160,165) ; Laríngeo ^(159,160) ; Esofágico ^(161,166-169) ; Cavidade Nasal ⁽¹⁶¹⁾ ; Pulmão ⁽¹⁶¹⁾ .

2006	12	Oral ⁽¹⁷⁰⁻¹⁷⁶⁾ ; Faríngeo ^(172,174) ; Laríngeo ⁽¹⁷²⁾ ; Cólon ⁽¹⁷²⁾ ; Reto ⁽¹⁷²⁾ ; Pulmão ⁽¹⁷²⁾ ; Mesotelioma ⁽¹⁷²⁾ ; Bexiga ^(172,177,178) ; Esofágico ^(172,179,180) ; Gástrico ^(172,177) ; Nasofaríngeo ⁽¹⁷⁷⁾ ; Linfoma Burkitt ⁽¹⁷⁷⁾ ; Fígado ⁽¹⁷⁷⁾ ; Orofaríngeo ⁽¹⁷³⁾ ; Hipofaríngeo ⁽¹⁸¹⁾ ; Cervical ⁽¹⁷⁷⁾ ; Cabeça e pescoço ⁽¹⁷⁴⁾ .
2007	14	Oral ⁽¹⁸²⁻¹⁸⁹⁾ ; Nasofaríngeo ^(185,189) ; Laríngeo ^(185,190) ; Esofágico ^(186,189-192) ; Orofaríngeo ^(187,193) ; Mieloma ⁽¹⁹⁰⁾ ; Pancreático ^(190,192,194) ; Bexiga ^(190,195) ; Cólon ^(190,192) ; Retal ^(190,192) ; Melanoma ⁽¹⁹⁰⁾ ; Pulmão ^(189,190,192) ; Fígado ^(190,192) ; Gástrico ⁽¹⁸⁹⁾ ; Leucemia ⁽¹⁹⁰⁾ ; Pele ⁽¹⁸⁹⁾ ; Cervical ^(189,193) ; Endométrio ⁽¹⁹⁰⁾ ; Mieloma ⁽¹⁹⁰⁾ ; Leiomyosarcoma ⁽¹⁸⁹⁾ .
2008	12	Oral ⁽¹⁹⁶⁻²⁰⁰⁾ ; Faríngeo ^(196-198,200) ; Nasofaríngeo ⁽¹⁹⁶⁾ ; Orofaríngeo ⁽¹⁹⁶⁾ ; Hipofaríngeo ⁽¹⁹⁶⁾ ; Laríngeo ^(196,197,200) ; Pancreático ^(198,200,201) ; Pulmão ^(198,200,201) ; Esofágico ^(198,200,202) ; Gástrico ^(198,200) ; Cavidade Nasal ⁽¹⁹⁷⁾ ; Renal ⁽²⁰¹⁾ ; Linfoma Non-Hodgkin ^(201,203) ; Linfoma de Hodgkin ⁽²⁰¹⁾ ; Mama ⁽²⁰⁰⁾ ; Fígado ⁽²⁰⁰⁾ ; Próstata ⁽²⁰⁰⁾ ; Cólon ⁽²⁰⁰⁾ ; Reto ⁽²⁰⁰⁾ ; Leucemia ⁽²⁰⁰⁾ ; Bexiga ⁽²⁰⁰⁾ ; Cervical ⁽²⁰⁰⁾ ; Cabeça e pescoço ^(196,197,204) .
2009	16	Oral ⁽²⁰⁴⁻²¹³⁾ ; Faríngeo ^(205,210) ; Orofaríngeo ⁽²¹⁴⁾ ; Laríngeo ^(209,210,214) ; Esofágico ^(209,215,216) ; Renal ^(209,216) ; Gástrico ⁽²¹⁷⁾ ; Linfoma Non-Hodgkin ⁽²¹⁶⁾ ; Cólon ^(216,218) ; Reto ^(216,218) ; Pulmão ⁽²¹⁶⁾ ; Bexiga ⁽²¹⁶⁾ ; Próstata ⁽²¹⁶⁾ ; Cervical ⁽²¹⁶⁾ ; Cabeça e pescoço ⁽²¹⁹⁾ .
2010	11	Oral ⁽²²⁰⁻²²⁵⁾ ; Faríngeo ^(221,225) ; Orofaríngeo ^(221,226-228) ; Pulmão ⁽²²¹⁾ ; Bexiga ⁽²²¹⁾ ; Mama ⁽²²²⁾ ; Laríngeo ⁽²²⁵⁾ ; Esofágico ^(224,229) ; Gástrico ^(224,229) ; Cabeça e pescoço ^(228,230) .
2011	14	Oral ⁽²³¹⁻²⁴¹⁾ ; Orofaríngeo ^(234-236,240,242) ; Laríngeo ^(234,236) ; Cervical ^(234-236,243,244) ; Ânus ^(236,244) ; Pénis ⁽²³⁶⁾ ; Pulmão ⁽²³⁶⁾ ; Linfoma Non-Hodgkin ⁽²⁴⁴⁾ ; Cólon ⁽²⁴⁴⁾ ; Reto ⁽²⁴⁴⁾ .
2012	16	Oral ⁽²⁴⁵⁻²⁵⁷⁾ ; Cabeça e pescoço ^(245,258) ; Faríngeo ^(247,248) ; Orofaríngeo ^(245,249) ; Ânus ⁽²⁴⁹⁾ ; Esofágico ⁽²⁵⁹⁾ ; Cervical ^(248,249,259,260) ; Gástrico ⁽²⁶¹⁾ ; Pulmão ⁽²⁵⁴⁾ ; Laríngeo ⁽²⁴⁵⁾ ; Ano-genital ⁽²⁵⁹⁾ .

Tabela I - Representação da bibliografia consultada referente a fatores associados a cancro.

A Tabela II, página 61, representa o número de artigos consultados para cada tipo de cancro. Os tipos de cancro com * são pertencentes ao termo “cancro da cabeça e pescoço”.

TABELA II			
Nº de artigos	Tipo de Cancro	Nº de artigos	Tipo de Cancro
98	Cancro Oral ^(70-88,97-99,109-115,121-125,132-136,144-150,157-162,170-176,182-189,196,197,199,200,204-213,220-225,231-233,235-241,245-257)	5	Ânus ^(103,145,236,244,249)
22	Faríngeo ^{*(71,73,79,83-85,87,97,98,112,122,135,145,157,158,172,174,205,210,221,225,247,248)}	30	Cólon ^(73,78,90,91,101,105,126,127,132,145,147,150,151,172,190,192,200,216,218,244)
8	Nasofaríngeo ^{*(103,104,150,153,177,185,189,196)}	30	Reto ^(73,78,90,91,101,105,126,127,132,145,147,150,151,172,190,192,200,216,218,244)
30	Orofaringeo ^{*(73,102,106,110,114,115,121,133,134,136,142,145,150,159,160,165,173,187,193,196,214,221,226-228,234-236,240,242,245,249)}	11	Mama ^(76,92,100,101,122,132,137,145,147,200,222)
5	Hipofaríngeo ^{*(121,136,150,181,196)}	32	Cervical ^(73,80,93-95,102,103,110,115,130,131,133,134,138,143,145,146,150,152,153,177,189,193,200,216,234-236,243,244,248,249,259,260)
22	Laríngeo ^{*(77-79,82,112,134,145,150,159,160,172,185,190,196,197,200,209,210,214,225,234,236,245)}	2	Vulvar ^(102,103)
5	Cavidade Nasal ^{*(84,134,150,161,197)}	2	Endométrio ^(101,190)
38	Esofágico ^{*(83,113,116-118,121,128,129,132,134,137,138,140,147,150,154,155,161,166-169,172,179,180,186,189-192,198,200,202,209,215,216,224,225,249,259,262)}	24	Cabeça e Pescoço ^(74,78,88,103,106,107,112,119,120,136,138,142,156,163,164,174,196,197,219,224,230,245,258,263)
14	Pancreático ^(88,89,101,134,141,145,150,152,190,192,194,198,200,201)	6	Ano-genital ^(102,106,110,115,153,259)
7	Renal ^(101,134,145,150,201,209,216)	1	Ovários ⁽¹⁰¹⁾
5	Pele ^(78,102,127,153,189)	20	Pulmão ^(75,78,99,127,132,134,145,150,161,172,189,190,192,198,200,201,216,221,236,254)
3	Sarcoma de Kaposi ^(103,152,153)	1	Papilomatose respiratória recorrente ⁽⁷⁸⁾
2	Melanoma ^(88,190)	4	Mesotelioma ^(104,127,152,172)
7	Linfoma Non-Hodgkin ^(103,145,152,201,203,216,244)	14	Fígado ^(101,103,104,126,127,134,138,145,152,153,177,190,192,200)
2	Linfoma B-Ceel ^(103,152)	14	Bexiga ^(75,77,78,126,127,132,134,145,150,190,195,200,216,221)
5	Linfoma Hodgkin ^(103,104,152,153,201)	18	Gástrico ^(77,78,103,126,128,132,134,145,152,172,177,189,198,200,217,224,229,261)
3	Linfoma Burkitt ^(103,152,153)	2	Vesícula Biliar ^(101,145)
4	Hiperplasia do gânglio linfático gigante ^(103,104,152,153)	6	Próstata ^(101,106,200,216)
4	Leucemia das células T adultas ^(103,104,145,153)	4	Cérebro ⁽¹⁵²⁾
5	Leucemia ^(138,145,150,190,200)	2	Génito-Urinário ⁽¹⁵²⁾
2	Leiomyosarcoma ^(103,189)	1	Linfoma Imunoblástico ⁽¹⁵³⁾
1	Osteosarcoma ⁽¹⁰⁴⁾	7	Pénis ^(80,102,103,106,145,152,236)
1	Mieloma ⁽¹⁹⁰⁾		

Tabela II - Representação do tipo cancerígeno identificado em cada artigo consultada pelo presente estudo.

4.2 Base de dados baseada em ontologias – *FactorXcancer*

A base de dados dedicada a factores associados a cancro foi designada por *FactorXcancer* e a sua estrutura foi realizada na aplicação *Excel* da empresa *Microsoft*TM.

A *FactorXcancer* começou por integrar a diversa terminologia utilizada pelos múltiplos autores. Com a conclusão da catalogação dos factores associados a cancro verificou-se que existia uma discrepância imensa entre as terminologias utilizadas pelos diferentes autores, para descrever um vastíssimo número de factores com implicação nos diversos tipos de cancro.

No sentido de a base de dados *FactorXcancer* reflectir uma estrutura com uma terminologia universal foi necessário construir/reutilizar um sistema de ontologias, o *OntoXcancer*, adequado para a catalogação dos factores com implicação em cancro.

Após a construção da *OntoXcancer*, foi necessário converter a *FactorXcancer* de modo a integrar o sistema de ontologias desenvolvido neste trabalho. Por uma questão de melhor compreensão do processo de desenvolvimento da base de dados começaremos por descrever a base de dados em si, a *FactorXcancer* e só depois iremos descrever o procedimento para o desenvolvimento das ontologias- a *OntoXcancer*.

Dada a extensão da base de dados, esta é apresentada como anexo em suporte digital que acompanha a presente dissertação. Passaremos no entanto a descrever a sua organização.

A base de dados foi organizada de acordo com o tipo de cancro. Para cada cancro, foram catalogadas ontologias específicas, que descrevem os factores específicos atribuídos pelas referências bibliográficas. Após a designação de cada cancro a organização hierárquica da base de dados é estabelecida de igual modo como as das ontologias, com Superclasses, Classes e Subclasses.

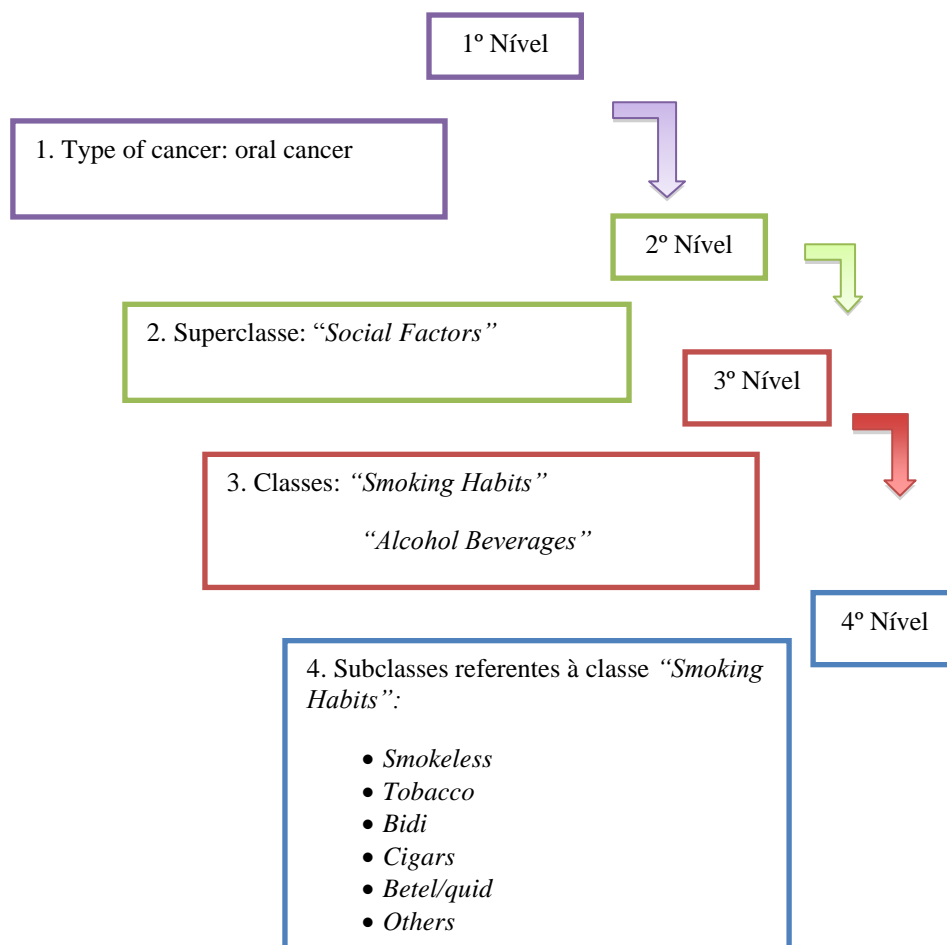
Desta forma, cada coluna do cabeçalho da estrutura da base de dados consiste numa ontologia que funciona como um descritor de uma estrutura hierárquica a ser apresentada no capítulo referente aos Resultados e Discussão (pág. 107-124).

Visando a futura disponibilização da base de dados a toda a comunidade científica as ontologias foram criadas em língua inglesa.

A anotação da base de dados com a informação referente a cada um dos artigos consultados seguiu os seguintes critérios:

1. Código *Pubmed*, do artigo;
2. Ano de publicação;
3. Identificação do cancro em questão (num total identificaram-se 37 tipos de neoplasia maligna);
4. Organização dos fatores de risco em superclasses, classes e subclasses consoante as ontologias criadas (*OntoXcancer*):
 - a. Na Superclasse “*Social Factors*” encontram-se identificados os fatores associados referentes aos comportamentos sociais humanos, tais como os hábitos tabágicos e alcoólicos;
 - b. Na Superclasse “*DietFactors*” são identificados os fatores referentes à dieta alimentar e à forma como estes são confeccionados;
 - c. Na Superclasse “*ClinicalFactors*” catalogaram-se fatores associados a patologias sistémicas, devidamente organizadas por hierarquias reutilizadas da organização *MeSH*;
 - d. Na Superclasse “*EnvironmentFactors*” inserem-se os fatores associados ambientais e ocupacionais;
 - e. Segue-se a Superclasse “*Genetic*”, dedicada aos genes, polimorfismos e doenças genéticas identificadas ;
 - f. “*Therapy*” refere-se a terapias;
 - g. “*Pharmacotherapy*” organiza os fármacos pelo local de acção, (organização *MeSH*);

Dentro de cada Superclasse encontram-se inseridas as Classes e subclasses que constituem as ontologias a que correspondem os fatores associados. O esquema seguinte exemplifica a organização das ontologias na base de dados pelos diferentes níveis, referentes a cancro oral e para Superclasse –“*Social Factors*”.



4.3 Anotação da base de dados sobre fatores de neoplasias

A anotação da informação relativa a cada artigo foi realizada de forma sequencial. Numa primeira fase o preenchimento dos dados identificativos do artigo utilizando o código *Pubmed*, ano de publicação e o tipo de cancro estudado, utilizando o carácter *x*, para identificar a célula da tabela correspondente ao tipo de cancro, como ilustra a Figura 7.

	PMID	Ano	Oral cancer	
1				
2	18569329	2008	x	>
3	18569329	2008	x	>
4	18569329	2008	x	>
5	18569329	2008	x	>
6	20381409	2010	x	

Figura 7 - Representação da base de dados, com exemplo de preenchimento da 1 fase: na coluna 2 insere o código *pubmed*, o ano, e o tipo de cancro.

De seguida anotaram-se as características ontológicas dos fatores. É atribuído o valor *1* se o artigo descrever uma ontologia correspondente ao fator como constituindo risco; *-1* se o artigo descrever uma ontologia considerada como um fator protetor e *0* quando o artigo elimina a ontologia/fator do estudo, ou seja, quando é retirado da amostra (fator de exclusão). Normalmente este tipo de método de estudo é realizado quando se quer obter dados referentes a relações de sinergias/antagonismos entre fatores.

Resumindo, a base de dados deve ser preenchida na horizontal, sendo que uma linha descreve, a identificação do artigo, seguindo-se pela caracterização do tipo de cancro, e termina pela catalogação dos fatores associados descritos pela literatura referente a fatores associados a cancro. Caso o artigo descreva informação referente a diferentes fatores, estes são identificados numa nova linha, exceto quando existem citações de relações, sendo neste caso identificado na mesma linha.

Para que se entenda esta relação/ conversão entre ontologias (*OntoXcancer*) e a base de dados (*FactorXcancer*), apresenta-se uma secção ilustrativa da base de dados e

da ontologia correspondente à mesma secção organizacional (Figuras 8 e 9), onde os fatores catalogados são: “*Social Factors*”, “*Smoking Habits*”, *smokeless (heavy or moderate)*; *Tobacco (Heavy or moderate)*; “*Betel Quid*; *Cigars*; *Opium*”).

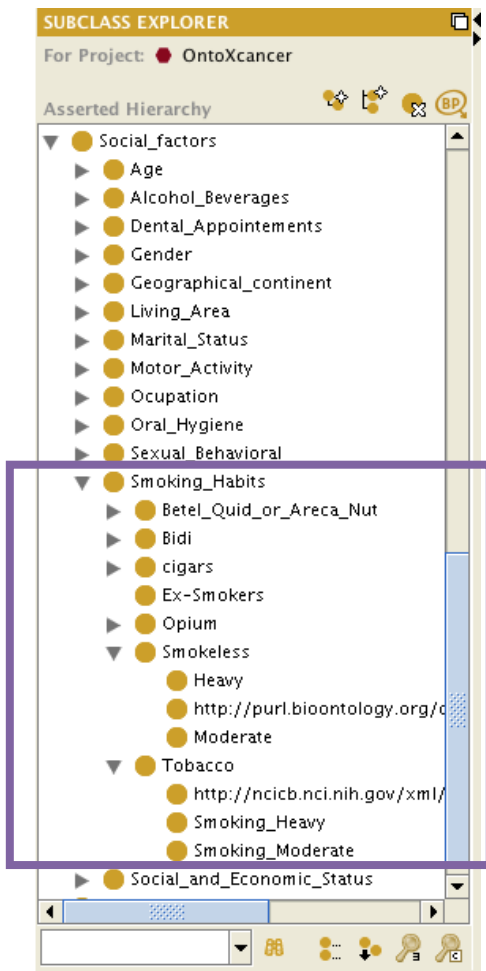


Figura 9 - Representação de uma secção da ontologia, com a descrição da classe “*Smoking Habits*”.

Social Factors	Smoking Habits	Smokeless	Snuff	Chewing	heavy smokeless tobacco	moderate smokeless tobacco	Smoking tobacco	heavy Smoking tobacco	moderate Smoking tobacco
1	1	1	1	1	1	1	1	1	1
1	1						1		

Figura 9 - Representação de uma secção da base de dados, com a descrição da classe “*Smoking Habits*”.

4.4 Desenvolvimento da *OntoXcancer* uma ontologia dedicada a fatores associados a cancro

OntoXcancer é uma ferramenta bioinformática com um conjunto de ontologias criadas de raiz e ontologias pré-existentes de outras bibliotecas ontológicas. O *OntoXcancer* integra classes e conceitos de fatores, servindo de plataforma de interação entre a base de dados inicial e o *FactorXcancer*.

4.4.1 Programa informático utilizado para a construção do *OntoXcancer* - *Protégé*

Neste projeto foi utilizado o editor de ontologias *Protégé* (<http://protege.stanford.edu/>) versão 3.5 para Mac OS, como representado pela Figura 10. É um *software* gratuito, de código aberto, desenvolvido em cooperação entre a Universidade de Stanford e a Universidade de Manchester. Este editor permitiu a conversão da estrutura da base de dados inicial para uma linguagem baseada em ontologias interpretável tanto pela web como pelo utilizador.

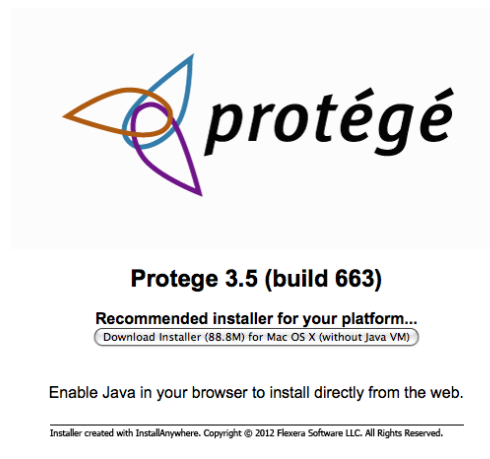


Figura 10 - Página web representativa do *Protégé* 3.5.

Protégé é altamente personalizável e relativamente intuitivo. Para que isto seja possível este *software* emprega as seguintes características:

- Um modelo de conhecimento extensível, onde podemos redefinir as primitivas de representação que o sistema utiliza;
- Um formato de arquivo de saída personalizável, onde se pode implementar componentes que traduzem a partir da representação interna do *Protégé* para uma representação de texto em qualquer linguagem formal;
- A interface de utilizador é personalizável, permitindo substituir componentes de interface para exibir e adquirir dados com novos componentes que se adaptam melhor na nova linguagem;
- Uma arquitetura extensível que permite a integração com outras aplicações, ou seja, é possível ligar o *Protégé* diretamente a outras bibliotecas semânticas externas⁽⁵⁵⁾, como o *Bioportal*⁽⁵⁹⁾;

Segundo Custódio (2012), “Esta ferramenta disponibiliza uma interface intuitiva para o desenvolvimento de ontologias, através da manipulação dos seus diversos painéis, possibilitando o desenho hierárquico de classes, a descrição de propriedades, a construção de restrições e regras, a definição de conceitos e comentários, entre outros”.⁽³⁷⁾

Com a ferramenta *Protégé* conseguimos ter uma interface gráfica que permitiu o desenvolvimento de ontologias adequadas ao que pretendíamos. Esta ferramenta oferece de igual modo uma interface gráfica que permite o desenvolvimento de uma ontologia sem a necessidade de entender a sintaxe de uma linguagem de saída, como RDFS, suporta o desenvolvimento de ontologias em várias linguagens, incluindo a RDF e a OWL e permite a interoperabilidade com outros sistemas de representação de informação⁽⁴³⁾.

Uma ontologia juntamente com um conjunto de instâncias individuais das classes constitui uma base de informação. Na realidade, existe uma linha muito tênue, entre onde a ontologia (especificação e representação de uma terminologia) termina e começa a base de informação (por exemplo o *OntoXcancer*).

Em termos práticos, o desenvolvimento de uma ontologia inclui⁽⁴⁾:

- Classes que definem a ontologia;
- Organização taxonómica hierárquica dessas classes em superclasse, classe e subclasse;
- Definição de *slots*, ou seja valores, da ontologia e descrever esses valores;
- Preenchimento dos *slots* para cada caso, (preenchimento de instância);

4.4.2 Construção da biblioteca ontológica – *OntoXcancer*

Para a criação do *OntoXcancer*, em primeira instância, foi necessário agrupar toda a informação catalogada a partir da bibliografia consultada referente a fatores associados a cancro e que tinha sido depositada no *FactorXcancer*.

A nível da organização hierárquica, uma vez que as ontologias referentes a cancro são vastas, muitas delas foram reutilizadas, a partir de outras bibliotecas ontológicas dado que a ontologia já existia, e integradas no *OntoXcancer*. No entanto, estas ontologias que foram reutilizadas encontravam-se englobadas em diferentes bibliotecas com organizações hierárquicas distintas. Apesar da extensa pesquisa feita não encontramos uma organização hierárquica que cumprisse os objetivos propostos para a nossa base de dados sobre factores em cancro. Optámos portanto por criar uma organização hierárquica adequada.

Definimos então a organização hierárquica do *OntoXcancer* que comporta 7 Superclasses: “*Clinical Factors*”, “*Diet Factors*”, “*Environment Factors*”, “*Genetic*”; “*Pharmacotherapy*”; “*Social Factors*” e “*Therapy*”, como ilustrado pela Figura 11, página 71.

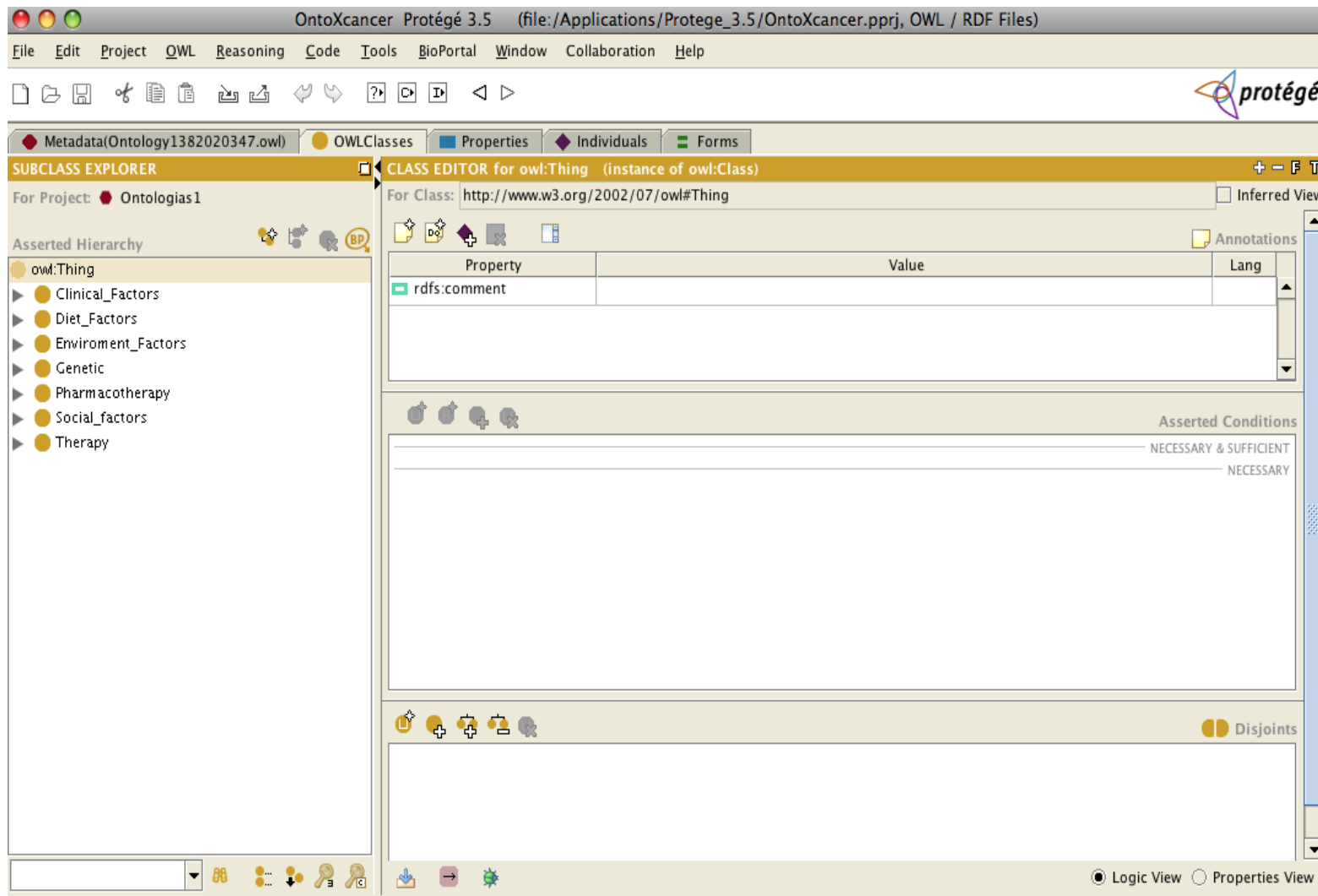


Figura 11 - Representação das Superclasses hierárquicas do *OntoXcancer* no programa *protégé*.

Quisemos no entanto, utilizar termos genericamente aceites pela comunidade médica, para isso recorremos à organização proposta pela *Medical Subject Headings (Mesh)*, o que também nos facilitou a organização hierárquica de algumas Superclasses, nomeadamente “*Clinical Factors*” e “*Pharmacotherapy*”.

Mesh é uma biblioteca de termos controlada pela *National Library of Medicine (NLM)* de terminologias utilizadas manualmente em índices de artigos para *MEDLINE/PubMed*. Abrange uma ampla gama de temas biológicos e termos médicos tais como patologias, anatomia, química, fármacos, e fenómenos e processos. Atualmente possui 26.142 termos organizados numa estrutura hierárquica⁽⁶⁸⁾. A Figura 12 representa a página Web representativa da *Medical Subjects Headings (MESH)*.

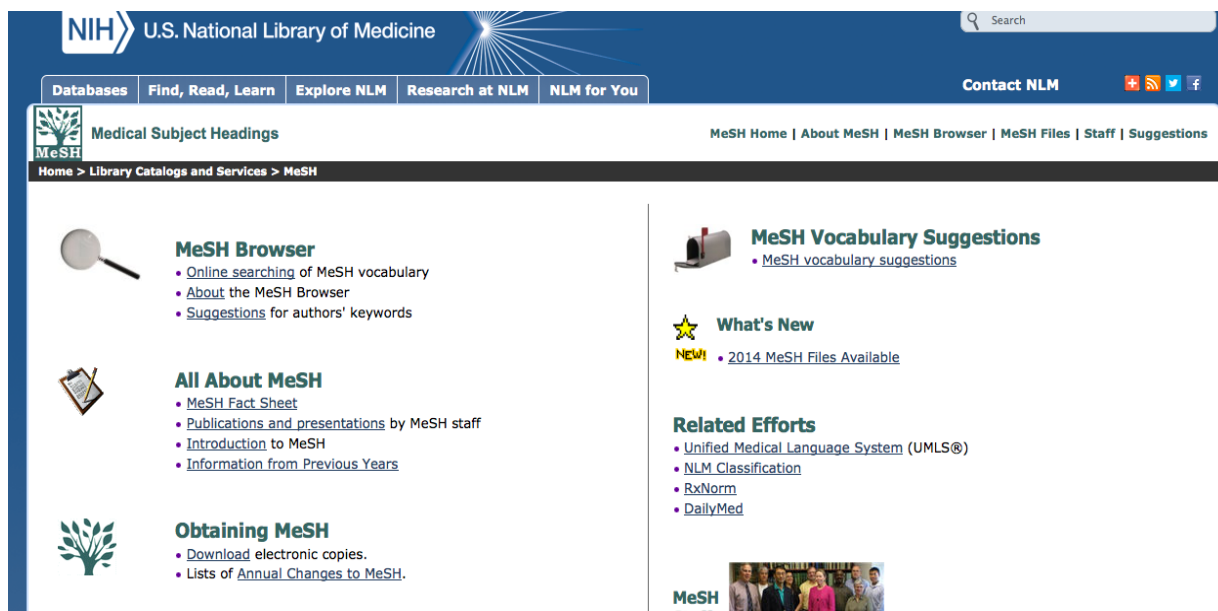
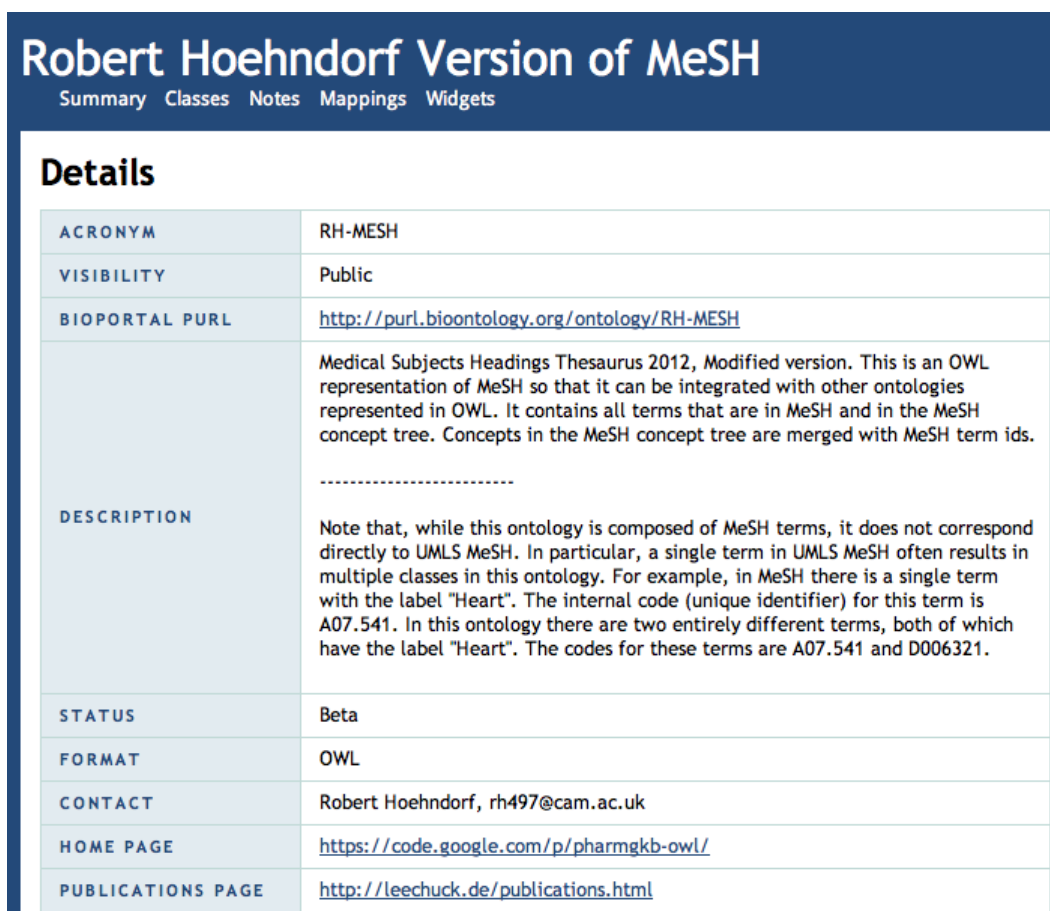


Figura 12 - Página web da *Medical SubjectHeadings(MESH)*.

Além da biblioteca *Mesh*, utilizámos conjuntamente a organização hierárquica de “*Robert Hoehndorf version of Mesh*”, que consiste numa versão modificada da biblioteca *Mesh*.

Ambas as bibliotecas são públicas e encontram-se também disponíveis a partir *Bioportal*. Estas possuem uma representação da linguagem Web OWL e RDF de modo a que possam ser integradas noutras ontologias representadas com a mesma linguagem pelo programa *Protégé*. A biblioteca “*Robert Hoehndorf version of Mesh*” contém todos os termos e conceitos *MeSH*, mas com uma organização hierárquica que se adaptara melhor à organização hierárquica pré-estabelecida para a *OntoXcancer* (Figura

11, página 71). A Figura 13, representa a biblioteca ontológica “*Robert Hoehndorf version of Mesh*” a partir do *Bioportal*.



Robert Hoehndorf Version of MeSH	
Summary Classes Notes Mappings Widgets	
Details	
ACRONYM	RH-MESH
VISIBILITY	Public
BIOPORTAL PURL	http://purl.bioontology.org/ontology/RH-MESH
DESCRIPTION	<p>Medical Subjects Headings Thesaurus 2012, Modified version. This is an OWL representation of MeSH so that it can be integrated with other ontologies represented in OWL. It contains all terms that are in MeSH and in the MeSH concept tree. Concepts in the MeSH concept tree are merged with MeSH term ids.</p> <p>-----</p> <p>Note that, while this ontology is composed of MeSH terms, it does not correspond directly to UMLS MeSH. In particular, a single term in UMLS MeSH often results in multiple classes in this ontology. For example, in MeSH there is a single term with the label "Heart". The internal code (unique identifier) for this term is A07.541. In this ontology there are two entirely different terms, both of which have the label "Heart". The codes for these terms are A07.541 and D006321.</p>
STATUS	Beta
FORMAT	OWL
CONTACT	Robert Hoehndorf, rh497@cam.ac.uk
HOME PAGE	https://code.google.com/p/pharmgkb-owl/
PUBLICATIONS PAGE	http://leechuck.de/publications.html

Figura 13 - Representação da biblioteca de ontologias *Robert Hoehndorf version of Mesh*, a partir da página Web do *Bioportal*.

Recuando até à organização hierárquica do *OntoXcancer*, (Figura 11, página 71) dentro de cada uma destas superclasses, a informação ontológica referente aos fatores associados a cancro são organizados em classes e subclasses, que correspondem a organizações hierárquicas inferiores das Superclasses. Por exemplo, na Superclasse “*Clinical Factors*” (que é a primeira superclasse da hierarquia) encontram-se inseridos todos os descritores ontológicos referentes aos fatores clínicos implicados em cancro, catalogados a partir da literatura consultada referente a fatores associados a cancro. A Figura 14, página 78, retrata toda a organização hierárquica da Superclasse “*Clinical Factors*”, onde o seu preenchimento com ontologias dependeu da informação descrita pela literatura consultada.

A Superclasse “*Clinical Factors*” possui vinte e uma Classes, que representam grupos de patologias sistêmicas:

1. “*Cardiovascular Diseases*”: integra todas as ontologias referentes às patologias relacionadas com o sistema cardiovascular, organizando-as em subclasses, como: “*Cardiovascular Abnormalities*”, *Cardiovascular Infections*; *Heart Diseases*; “*Pregnancy Complications Cardiovascular*” e “*Vascular Diseases*”.
2. “*Digestive System*”: reúne todas as ontologias representativas das patologias do sistema digestivo, organizando-as em subclasses como: “*Biliary Tract Diseases*”; “*Digestive System Abnormalies*”; “*Digestive System Fistula*”; “*Gastric Neoplasm*” (também pertencente à classe “*Neoplasm*” introduzindo assim o termo rede semântica, que será abordado futuramente); “*Gastrointestinal Diseases*”; “*Liver Diseases*”; “*Pancreatic Diseases*” e “*Peritoneal Diseases*”.
3. “*Endocrine System Diseases*”: abrange as ontologias que retratam as doenças endócrinas, organizando-se em subclasses, como: “*Adrenal Gland Diseases*”; “*Diabetes Mellitus*”; “*Dwarfism*”; “*Endocrine Gland Neoplasm*” (também pertencente à classe “*Neoplasm*”); “*Gonadal Disorders*”; “*Parathyroid Diseases*”; “*Thyroid Diseases*” e “*Tuberculosis Endocrine*”.
4. “*Eye Diseases*”: engloba as ontologias alusivas às doenças oculares, organizando-as em subdivisões encontrando-se até ao momento anotada uma subcategoria: “*Eye Infections Parasitic*”. Esta subcategoria também se enquadra na classe “*Parasitic Diseases*”.
5. “*Family History*”: que ainda não engloba nenhuma subcategoria.
6. “*Female Urogenital Diseases and Pregnancy Complications*”: abrange as ontologias referentes às patologias do sistema urogenital feminino organizando-as em duas subclasses: “*Female Urogenital Diseases*” e “*Pregnancy Complications*”.
7. “*Genetic Diseases*”: retrata as ontologias referentes às patologias de origem genética, organizando-as em subclasses, como: “*Anemia de Fanconi*”; “*Ataxia-telangiectasia*”; “*Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy*”; “*Blooms Syndrome*”; “*Li Fraumeni Syndrome*”; “*Multiple Hamartoma Syndrome*”; “*Syndrome Lynch*” (I e II) e “*Xeroderma Pigmentosum*”; (esta classe pertence de igual forma à Superclasse “*Genetic*”).

8. *“Hemicand Lymphatic Diseases”*: que organiza as ontologias alusivas às patologias do sistema linfático e sanguíneo, organizando-as em subclasses, como: *“Hematologic Diseases”* e *“LymphaticDiseases”*.
9. *“Immune System Diseases”*: reúne as ontologias pertencentes às patologias do sistema imune, organizando-as em subclasses: *“Autoimmune Diseases”* (*“Lupus Erythematosus Systemic”*); *“Blood Group Incompatibility”*; *“Glomerulo nephritis Membrano proloferative”*; *“Graft vs Host Disease”*; *“Hypersensitivity”*; *“Immune Reconstittution Inflammatory Syndrome”*; *“Immunologic Deficiency Syndrome”*; *“Immuno proliferative Disorders”* e *“Purpura Thrombocytopenic”*.
10. *“Infections”*: abrange todas as ontologias relativas às infecções bacterianas, virais e eucariotas catalogadas pela base de dados.
11. *“Male Urogenital Diseases”*: descreve as ontologias pertencentes às patologias do sistema Urogenital. As subclasses organizadas nesta classe são: *“Genital Diseases Male”*; *“Pelvic Floor Disorders”*; *“Tuberculosis Urogenital”*; *“Urogenital Abnormalities”*; *“Urogenital Neoplasm”* e *“Urologic Diseases”*.
12. *“Musculoskeletal Diseases”*: reúne todas as ontologias alusivas às doenças músculo-esqueléticas, organizando-as em subclasses como: *“BoneDiseases”*; *“Cartilage Diseases”*; *“Fasciitis”*; *“Foot Deformities”*; *“Foot Diseases”*; *“HandDeformities”*; *“Jaw Diseases”*; *“Joint Diseases”*; *“Muscular Diseases”*; *“Musculoskeletal Abnormalies”*; *“Rheumatic Diseases”* e *“Tennis Elbow”*.
13. *“Neoplasm”*: enquadraas ontologias referentes às 37 patologias cancerígenas catalogadas pela base de dados, descritas pela tabela II, página 44, acrescentando algumas que surgiram pelo preenchimento do *OntoXcancer*, que foram organizadas da seguinte forma: *“Otorhinolaryngologic Neoplasm”*; *“Respiratory Tract Neoplasm”*; *“Salivary Gland Neoplasm”*; *“Parotid Neoplasm”*; *“Pharyngeal Neoplasm”* e *“Endocrine Gland Neoplasm”*.
14. *“Nervous System Diseases”*: acomoda as ontologias referentes às patologias do sistema nervoso, organizando-as em subclasses como: *“Autoimmune Diseases of the Nervous System”*; *“Autonomic Nervous System Diseases”*; *“Central Nervous System Diseases”*; *“Central Nervous System Parasitic Infections”*; *“Chronobiology Disorders”*; *“Cranial Nerve Diseases”*; *“Sleep Disorders”*; *“Demyelinating Diseases”* e *“Trauma Nervous System”*.

15. “*Nutricional and Metabolic Diseases*”: onde se encontram anotadas as ontologias alusivas às patologias metabólicas e nutricionais, encontrando-se organizadas em subclasses, tais como: “*Metabolic Diseases*” e “*Nutrition Diseases*”.
16. “*Occupational Diseases*”: onde ainda não se encontra descrita nenhuma ontologia com uma organização hierárquica inferior.
17. “*Otorhinolaryngologic Diseases*”: enquadra as ontologias relativas às doenças otorrinolaringologistas, sendo organizadas em subclasses como: “*Ciliary Motility Disorders*”; “*Ear Diseases*”; “*Laryngeal Diseases*”; “*Nose Diseases*”; “*Otorhinolaryngologic Neoplasm*” (que também pertence à classe “*Neoplasm*”); “*Pharyngeal Diseases*”.
18. “*Parasitic Diseases*”: abrange as ontologias alusivas às doenças derivadas de infecção por parasitas, que se encontram organizadas em subclasses como: “*Central Nervous System Parasitic Infections*”; “*Coinfection*”; “*Eye Infections Parasitic*”; “*Helminthiasis*”; “*Intestinal Diseases Parasitic*”; “*Liver Diseases Parasitic*”; “*Lung Diseases Parasitic*”; “*Mesomycetozoea Infections*”; “*Opportunistic Infections*”; “*Parasitemia*”; “*Parasitic Diseases Animal*”; “*Pregnancy Complications Parasitic*”; “*Protozoan Infections*”; “*Skin Diseases Parasitic*” e “*Zoonoses*”.
19. “*Respiratory Tract Diseases*”: engloba as ontologias relativas às patologias do trato respiratório, organizado em subclasses como: “*Bronchial Diseases*”; “*Ciliary Motility Disorders*”; “*Granuloma Respiratory Tract*”; “*Laryngeal Diseases*” (que também pertence à classe *Otorhinolaryngologic Diseases*); “*Lung Diseases*”; “*Nose Diseases*”; “*Pleural Diseases*”; “*Respiration Disorders*”; “*Respiration Hypersensitivity*”; “*Respiratory System Abnormalities*”; “*Respiratory Tract Fistula*”; “*Respiratory Tract Infections*”; “*Respiratory Tract Neoplasm*” (que também pertence à classe “*Neoplasm*”); “*Thoracic Diseases*” e “*Tracheal Diseases*”.
20. “*Skin and Connective Tissue Diseases*”: onde estão organizadas ontologias referentes às doenças de pele em subclasses, tais como: “*Connective Tissue Diseases*”; “*Skin Diseases*”; “*Skin Diseases Parasitic*” (que pertence igualmente à classe “*Parasitic Diseases*”).

21. "*Stomatognathic Diseases*": abrange todas as ontologias relativas às patologias do sistema estomatognático, sendo este organizado em subclasses como: "*Dental Prothesis*"; "*Jaw Diseases*"; "*Mouth Diseases*", "*Pharyngeal Diseases*" (que pertence igualmente à classe "*Otorhinolaryngologic*") e "*Tooth Diseases*".

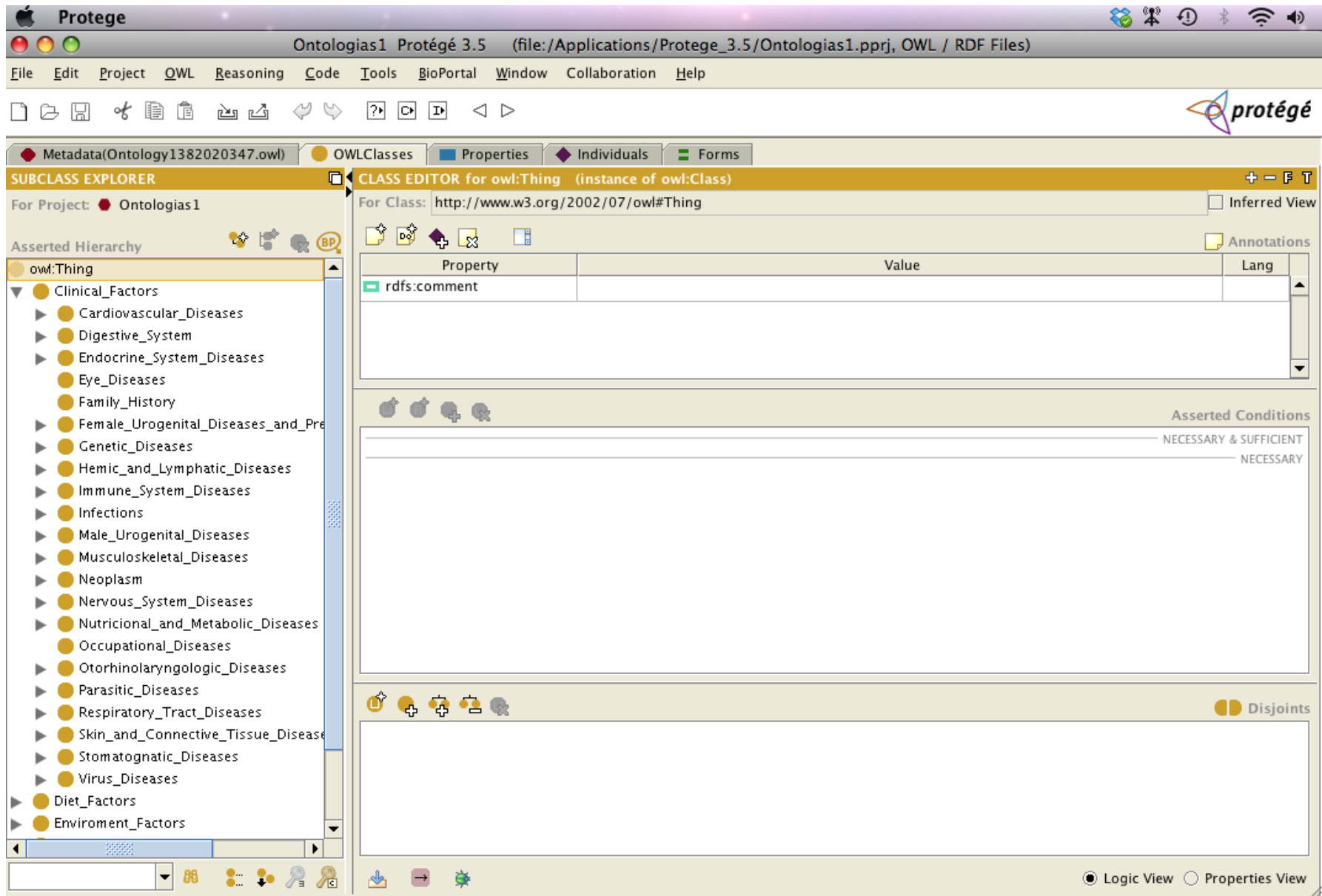


Figura 14 - Representação das Classes da Superclasse "Clinical Factors".

Em segundo lugar na organização hierárquica das Superclasses do *OntoXcancer* encontra-se a Superclasse “*Diet Factors*”. Nesta Superclasse são identificadas as ontologias referentes à dieta alimentar e bebidas, como representado pela Figura 16, página 81. Esta Superclasse hierárquica encontra-se organizada numa Classe, a “*Foodand Beverages*” e essa Classe encontra-se dividida em duas Subclasses hierárquicas inferiores organizadas em, “*Food*” e “*Beverages*”.

A Figura 15 representa um excerto do *OntoXcancer* onde é demonstrado a organização da Subclasse “*Beverages*”, que organiza hierarquicamente as subclasses ontológicas com níveis hierárquicos inferiores como: “*Alcohol Beverages*”; “*Milk*”; “*Milk Substitutes*”; “*Tea*”, que por sua vez englobam níveis hierárquicos inferiores como: “*Soy milk*”; “*greentea*”; “*maté*”.

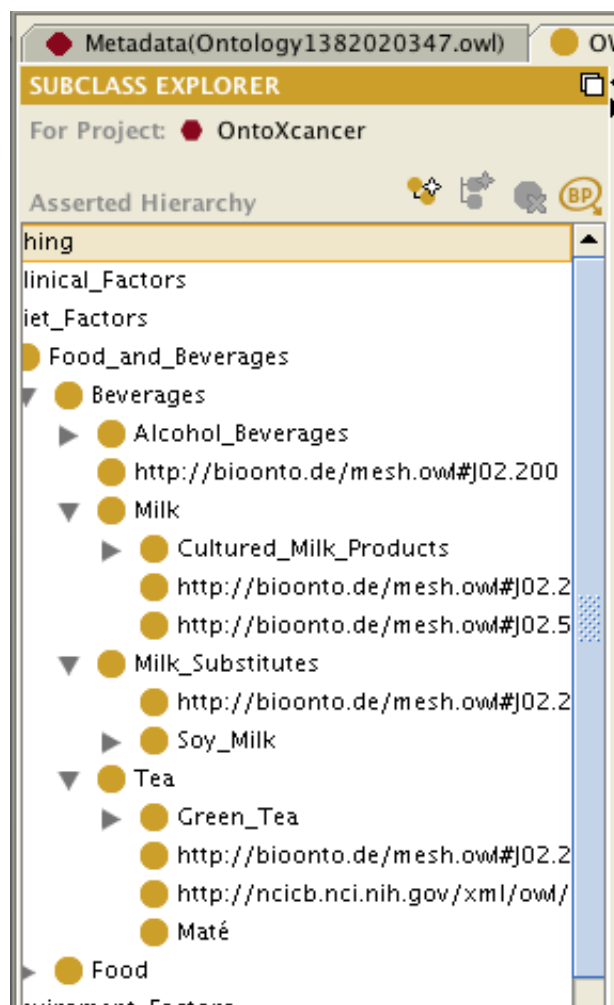


Figura 15 - Excerto do *OntoXcancer* onde é demonstrado a organização hierárquica da Subclasse “*Beverages*”.

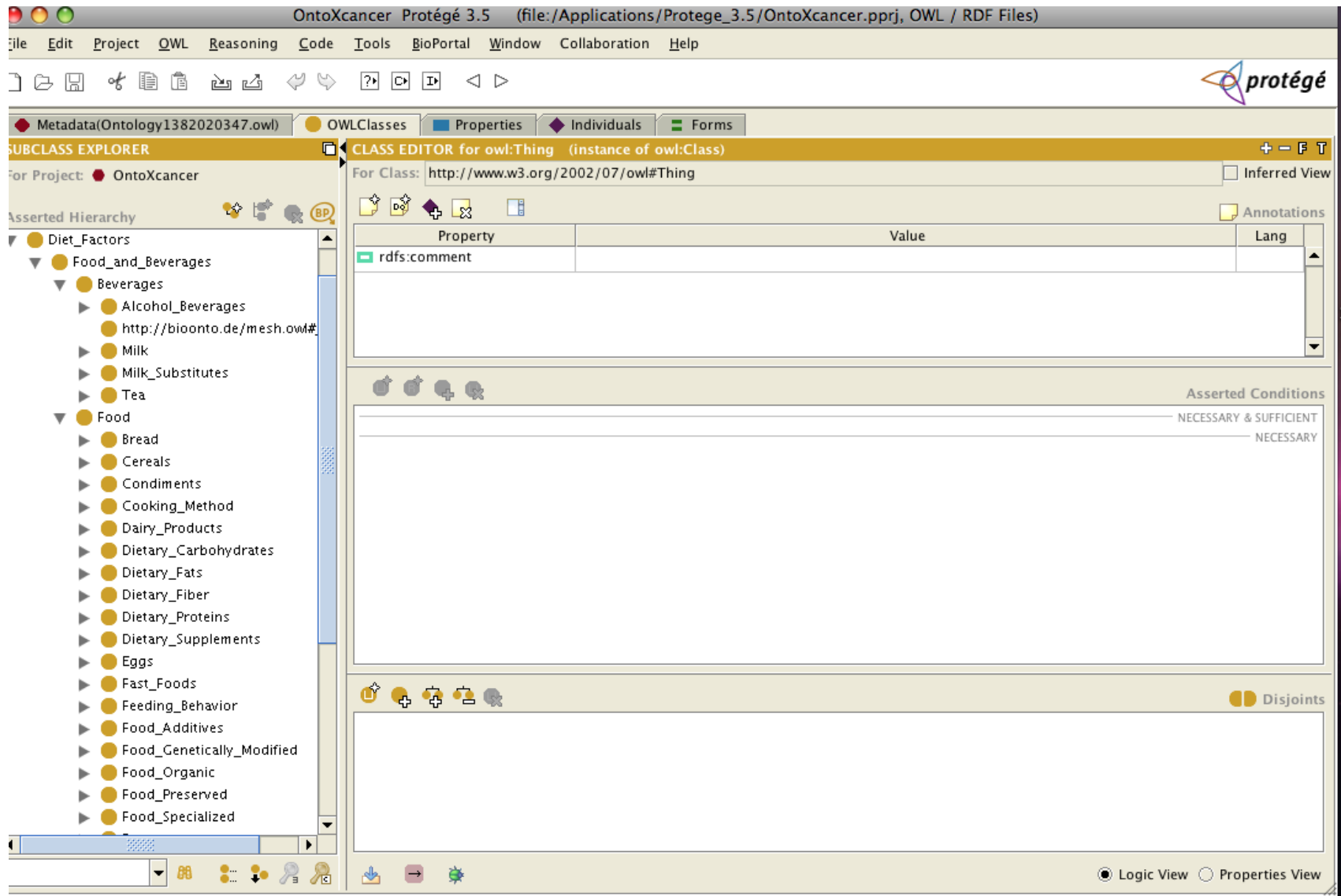


Figura 16 - Representação hierárquica das Classes da Superclasse "Diet Factors".

Seguindo este prisma, iremos de seguida caracterizar a Subclasse “*Food*”, pertencente à Superclasse “*Diet Factors*”, que classifica ontologicamente todos os fatores associados a cancro, relacionados com alimentos e forma como estes são confeccionados.

A Figura 17, evidencia a forma como esta Subclasse se encontra organizada hierarquicamente e como os alimentos se encontram organizados em “grupos” de ontologias.

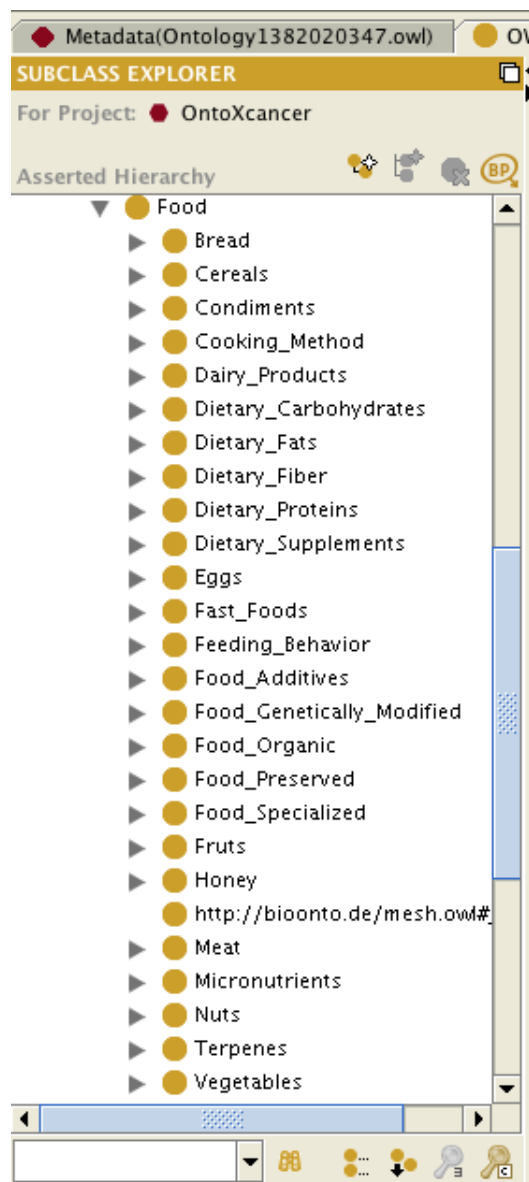


Figura 17 - Excerto do *OntoXcancer* onde é evidenciada a organização hierárquica da Subclasse “*Food*”.

Na terceira posição da organização hierárquica do *OntoXcancer*, encontra-se a Superclasse “*Enviroment Factors*”, onde são organizadas todas as ontologias que descrevem os fatores com implicação em cancro associados ao ambiente e à ocupação. Como representado pela Figura 18, as Classes que fazem parte da organização hierárquica da Superclasse “*Enviroment Factors*” são: “*Chemical Carcinogen*”; “*Chemical Modifier*”; “*Organic Chemical*”; “*Radiation*” e *Tobacco Associated Carcinogen*”.

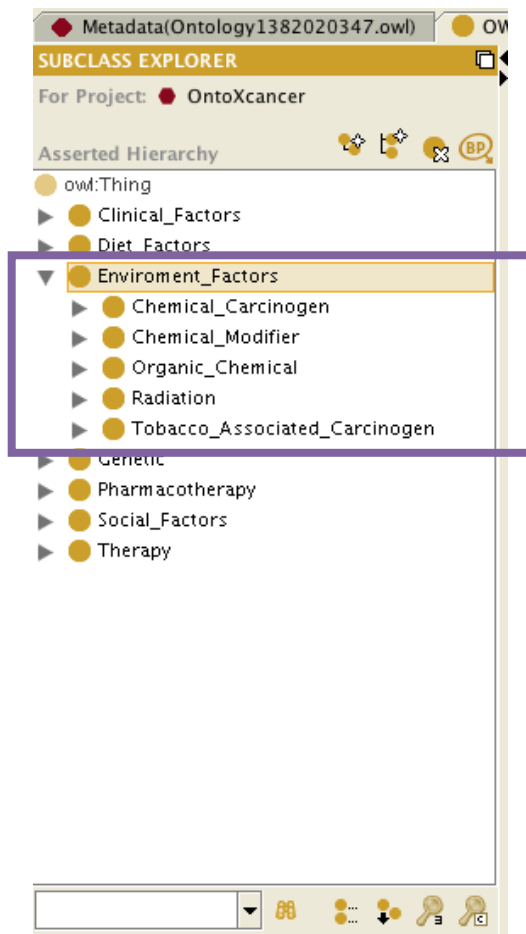


Figura 18 – Excerto do *OntoXcancer* onde é evidenciado a organização hierárquica da Classe “*Enviroment Factors*”.

Como anteriormente constatado por outras Superclasses, as Classes anteriormente descritas possuem ontologias organizadas em hierarquias inferiores conotadas como subclasses.

A Classe “*Chemical Carcinogen*” abrange as Subclasses “*Inorganic Carcinogen*”; “*Organic Carcinogen*”. Dentro de cada uma destas Subclasses ainda existem ontologias com uma organização hierárquica inferior, como a Subclasse “*Inorganic Carcinogen*”, que é organizada ontologicamente em divisões como “*Carcinogen Mineral*” (da qual fazem parte as ontologias dos “*asbestos*”, “*Benzine*” e “*Coalddust*”), “*Carcinogem Metal*” (da qual fazem parte as ontologias “*Cadmium*”; “*Cobalt*”; *Nickel Compound* como o *nickel refining*”; *Respirable*, onde fazem parte vários elementos, como: “*Cement Dust*”, “*Metal Dust, Textile Dust*”, “*Wood Dust*” e “*Plastic Byproducts*”. Já dentro da Subclasse “*Organic Carcinogen*” temos as divisões de ontologias em hierarquias inferiores como; “*Carcinogen Nitrosamoines*”; “*Carcinogenic Chlorinated Hydrocarbon*” (que abrange os termos ontológicos “*Tetrachloroethylene*” e “*Vinyl Chloride*”); e por último “*Organo-nitrogen Carcinogen*” (que enquadra a ontologia “*Acrylamide*”) como representa a Figura 19.

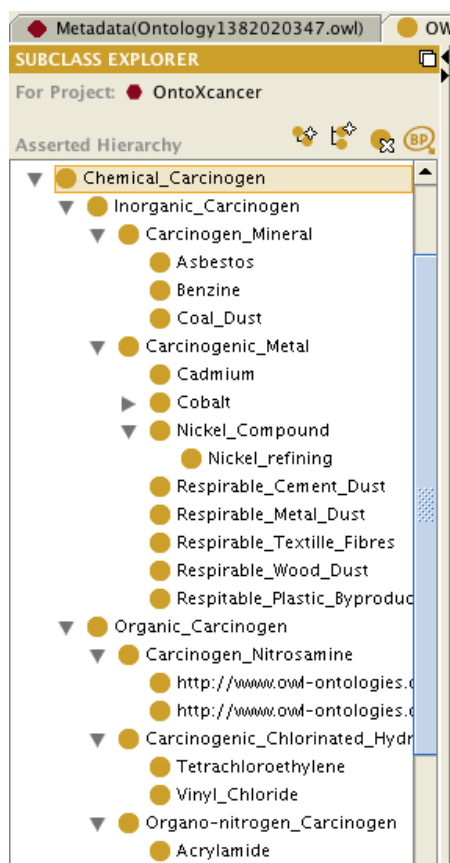


Figura 19 - Excerto do OntoXcancer onde é evidenciado a organização hierárquica da Classe “*Chemical Carcinogen*”

Seguindo nesta caracterização da organização hierárquica da Classe “*Chemical Modifier*” pertencente à superclasse “*Environment Factors*”, esta só abrange as ontologias referentes à Classe “*Toxin*”, que por sua vez foi dividida em “*Mycotoxin*” como a “*Aflatoxin*”. Já a Classe “*Organic Chemical*” encontra-se dividida em “*Aldehyde*” (que possui as Subclasses “*Acetaldehyde*” e “*Formaldehyde*”), e “*Heterocyclic Amines*”.

A organização ontológica da classe “*Radiation*”, é organizada hierarquicamente em duas Subclasses: “*Ionizing Radiation*” e “*Non Ionizing Radiation*”. Dentro da Subclasse “*Ionizing Radiation*” temos catalogadas ontologias pertencentes a fatores como: “*Radiation*”; “*Nuclear*”, “*Radon*”; “*Therapeutic and Diagnostic*”. Enquanto que, dentro da Subclasse “*Non Ionizing Radiation*” encontram-se catalogadas ontologias pertencentes a fatores como: “*Natural Radiation*”.

Por último resta a caracterização da Classe “*Tobacco Associated Carcinogen*” onde se encontram descritas as ontologias correspondente aos fatores com implicação em cancro, sendo estas organizadas em 2 Subclasses: “*B- naphthylamine e Polyaromatic hydrocarbons*”. A partir da Figura 20 podemos verificar toda esta organização hierárquica.

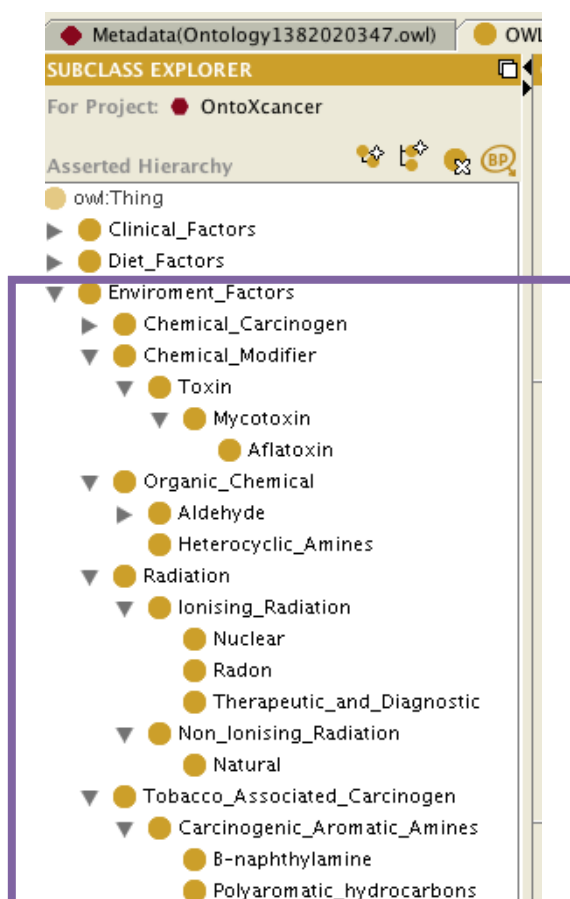


Figura 20 - Excerto do OntoXcancer onde é evidenciado a organização hierárquica da Classe “*Chemical Modifier*”, “*Radiation*” e “*Tobacco Associated Carcinogen*”.

Prosseguindo na caracterização da organização hierárquica do *OntoXcancer*, a superclasse “*Genetic*” encontra-se em quarta posição. Nesta Superclasse, encontra-se organizado hierarquicamente todos os descritores ontológicos referentes a genes e doenças genéticas. Para a organização e caracterização dos genes catalogados, contamos com a ajuda do *HUGO Gene Nomenclature Committee* (HGNC), que é a organização oficial de terminologia de genes. HUGO Gene atribuiu símbolos genéticos únicos e nomes a quase 38000 *loci* humanos dos quais cerca de 19000 codificam proteínas. É uma biblioteca on-line, como ilustra a Figura 21 (genenames.org), com recursos associados, incluindo links para informação genómica, proteómica e fenotípica, assim como páginas dedicadas a famílias de genes⁽⁶⁹⁾.

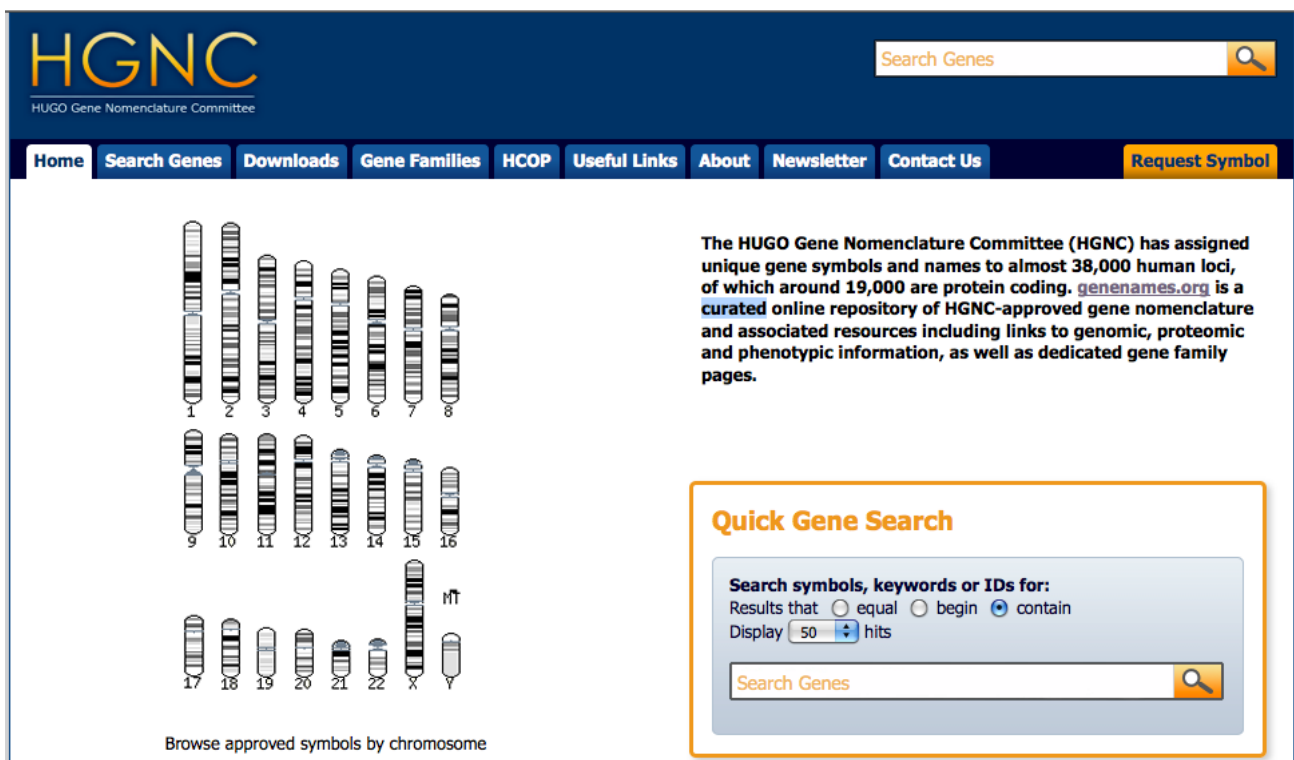


Figura 21 - Página Web HUGO Gene NomenclatureCommittee.

Como se pode verificar na Figura 22, página 88, esta Superclasse é organizada em três classes: “*Gene or Polymorphism*”; “*Genes by Function*” e “*Genetic Diseases*”.

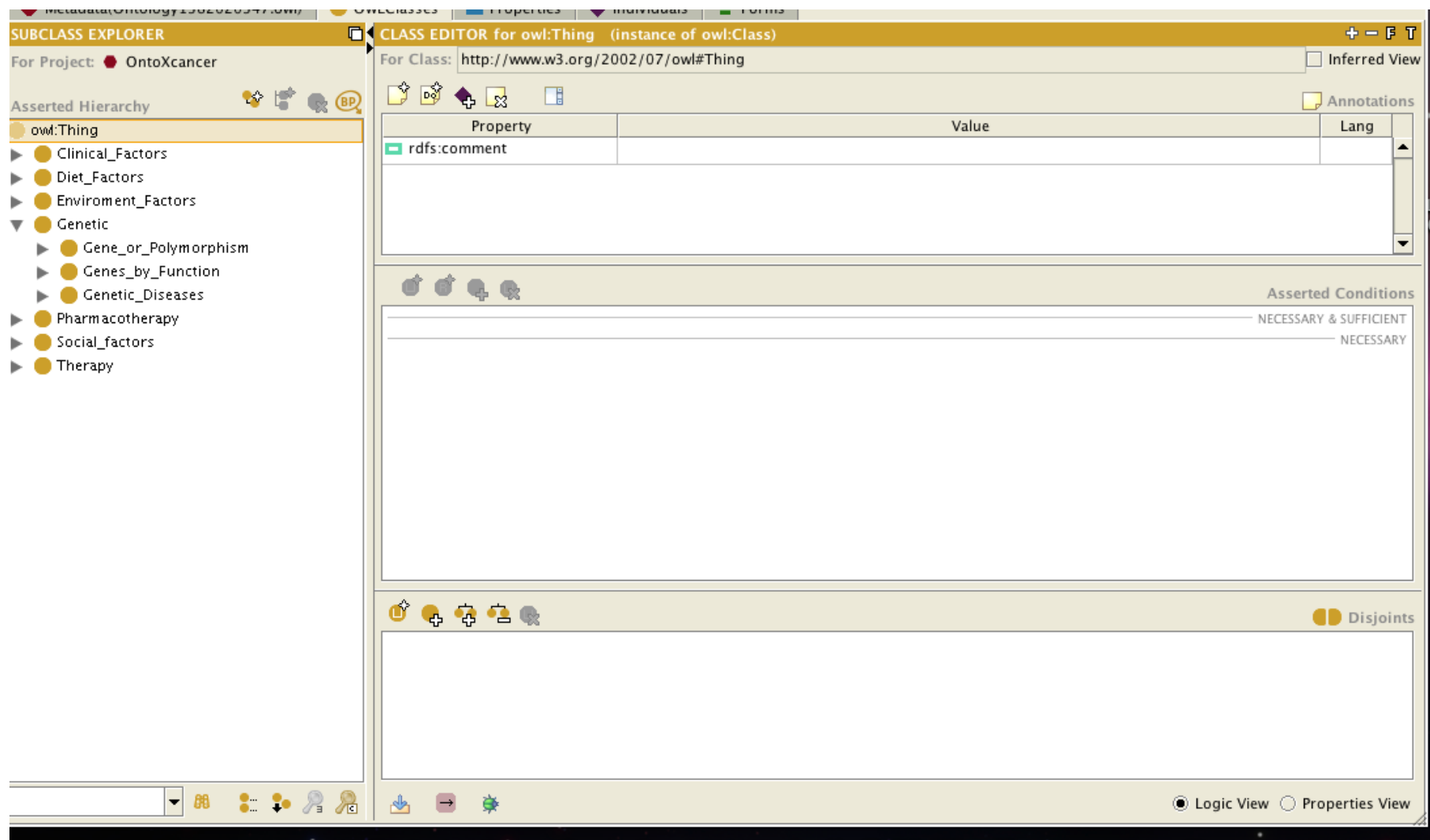


Figura 22 - Representação da organização hierárquica ontológica das classes pertencentes à superclasse “Genetic” no OntoXcancer.

A Classe “*Gene or Polymorphism*” reúne toda a catalogação de genes/polimorfismos como ilustrado na Figura 23.

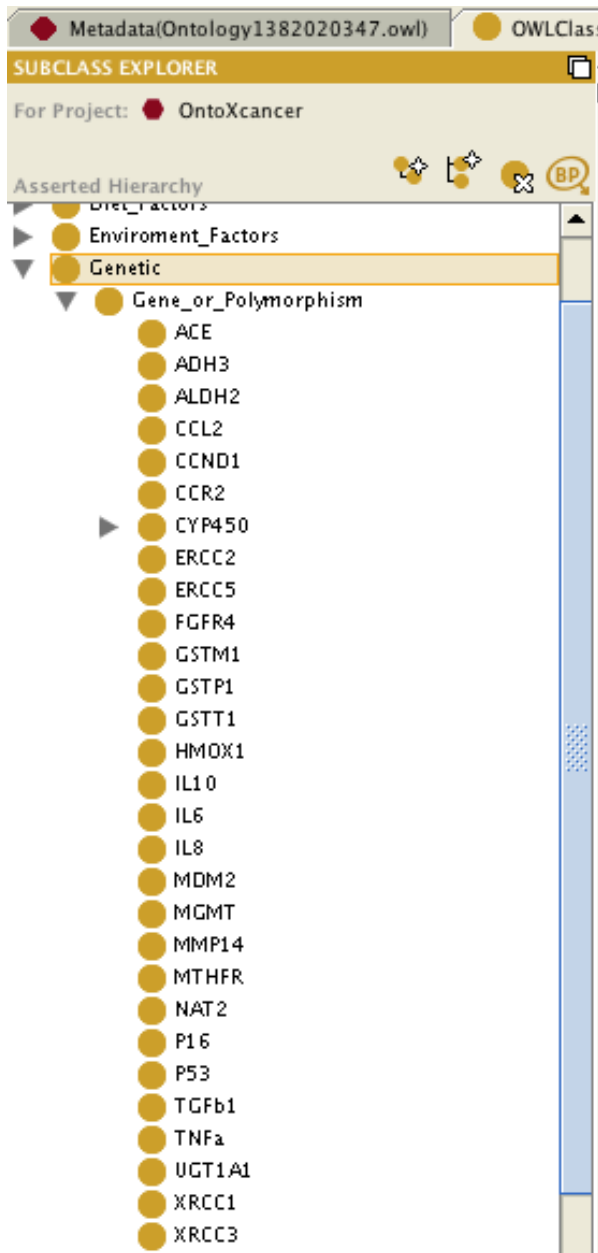


Figura 23 - Excerto do *OntoXcancer* onde é evidenciado a organização hierárquica da Classe “*Gene or Polymorphism*”, pertencente à Superclasse “*Genetic*”.

Já dentro das classes hierárquicas “*Genes by Function*” e “*Genetic Diseases*”, estas encontram-se representadas pelo excerto da Figura 24. É de salientar que os genes/polimorfismos englobados nas Classes anteriores também se encontram descritos na Classe “*Genes by Function*”, de acordo com a sua função biológica.

Por último dentro da superclasse “*Genetic*” temos a Classe “*Genetic Diseases*”, que abrange todas as patologias genéticas, anteriormente descritas pela mesma Classe inserida na Superclasse “*Clinical Factors*”, página 74 do presente capítulo.

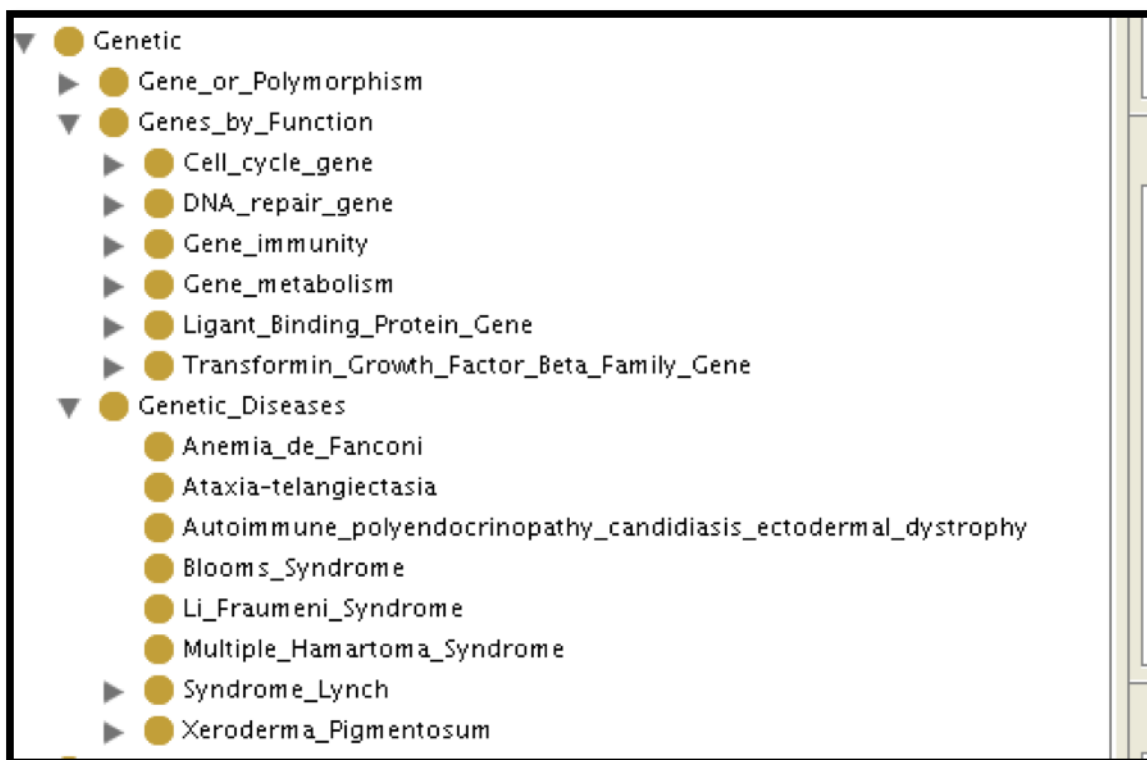


Figura 24 - Excerto do OntoXcancer onde é evidenciado a organização hierárquica da Classe “*Genes by Function*” e “*Genetic Diseases*”, pertencente à Superclasse “*Genetic*”.

Na Superclasse, em quinta posição “*Pharmacotherapy*”, como já foi referido anteriormente (página 72), utilizamos a organização proposta pela *Medical Subject Headings (Mesh)*, para facilitar o processo.

“*Pharmacotherapy*” organiza as ontologias referentes aos fármacos e aos materiais médicos utilizados intrinsecamente pelos indivíduos, catalogados como fator com implicação em cancro pelo local de acção. Esta Superclasse encontra-se hierarquicamente organizada em duas classes “*Biomedical and Dental Materials*” e “*Chemical and Drug Category*”. Na Figura 25, ilustra-se a organização hierárquica explicada anteriormente, de cada componente pertencente a cada Classe. Por exemplo, dentro da classe “*Biomedical and Dental Materials*” a Subclasse que se encontra subjacente é “*mouthwashes*”, que pode ou não conter álcool.

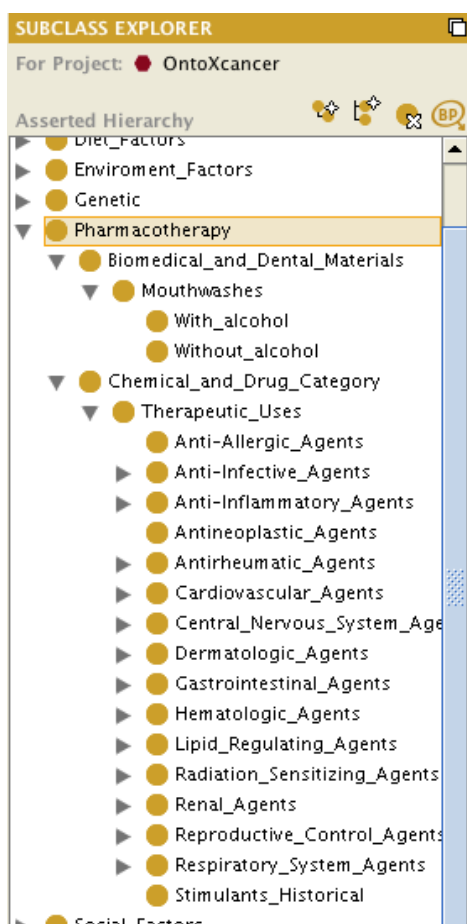


Figura 25 - Excerto do *OntoXcancer* onde é evidenciado a organização hierárquica da classe “*Pharmacotherapy*”.

Prossegue-se para a caracterização da organização hierárquica da penúltima Superclasse, “*Social Factors*”. Esta Superclasse encontra-se organizada em Classes com hierarquias inferiores referentes a caracterização individual a nível social como: “*Age*”, “*Gender*”; “*Geographical continent*”; “*Living Area*”; “*Marital Status*”; “*Occupation*” e “*Social and Economic Status*”, e comportamentos individuais sociais como: “*Alcohol Beverages*” (que também pertence à classe “*Beverages*” anteriormente descrita); “*Dental Appointments*”; “*Motor activity*”; “*Oral Hygiene*”; “*Sexual Behavioral*” e “*Smoking Habits*”, como se pode averiguar pela Figura 26, página 93.

Metadata(Ontology1382020347.owl) OWLClasses Properties Individuals Forms

SUBCLASS EXPLORER

For Project: **OntoXcancer**

Asserted Hierarchy

- owl:Thing
 - ▶ Clinical_Factors
 - ▶ Diet_Factors
 - ▶ Enviroment_Factors
 - ▶ Genetic
 - ▶ Pharmacotherapy
 - ▼ Social_factors
 - ▶ Age
 - ▶ Alcohol_Beverages
 - ▶ Dental_Appointments
 - ▶ Gender
 - ▶ Geographical_continent
 - ▶ Living_Area
 - ▶ Marital_Status
 - ▶ Motor_Activity
 - ▶ Occupation
 - ▶ Oral_Hygiene
 - ▶ Sexual_Behavioral
 - ▶ Smoking_Habits
 - ▶ Social_and_Economic_Status
 - ▶ Therapy

CLASS EDITOR for owl:Thing (instance of owl:Class)

For Class: <http://www.w3.org/2002/07/owl#Thing> Inferred View

Property	Value	Lang
rdfs:comment		

Annotations

Asserted Conditions

- NECESSARY & SUFFICIENT
- NECESSARY

Disjoints

Logic View Properties View

Figura 26 - Representação da organização hierárquica ontológica da superclasse “Social Factors” no *OntoXcancer*.

Por último caracterizamos a Superclasse mais inferior na hierarquia do *OntoXcancer*, “*Therapy*” onde se encontram descritos as ontologias referentes a terapias. Esta Superclasse enquadra três classes hierárquicas, a “*Gene Therapy Vectors*”, “*Hormone Replacement*” e “*Stem Cell Transplantation*”. Esta última classe identificada ainda organiza termos ontológicos em Subclasses, tais como: “*Cord Blod*”; “*Hematopoietic*”; “*Mesenchymal*” e “*Peripheral blood*” que caracterizam formas de transplantação de células estaminais, como se pode inferir pela Figura 27.

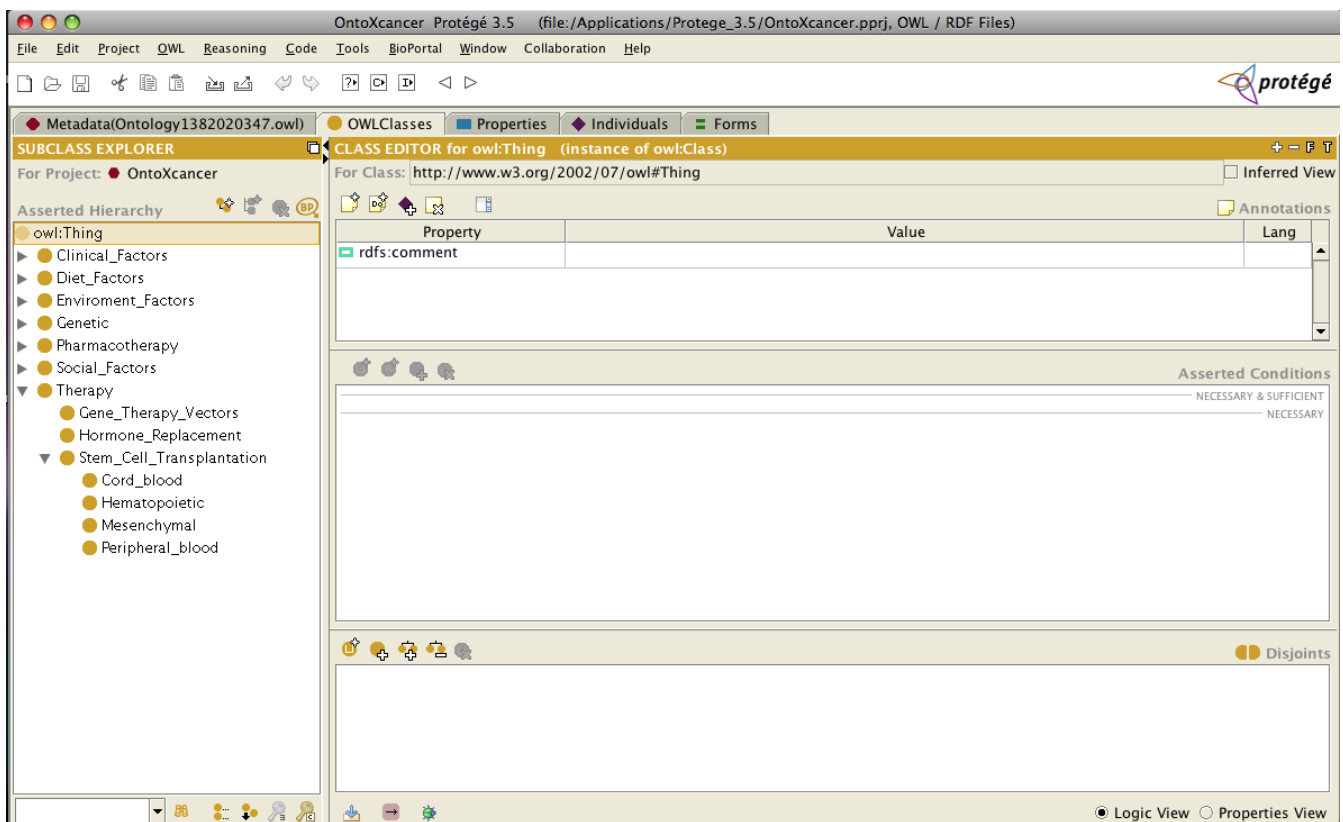


Figura 27 - Representação da organização hierárquica ontológica da superclasse “*Therapy*” no *OntoXcancer*.

Pelas Figuras 28 e 29, evidencia-se a forma como utilizamos a organização hierárquica, na Superclasse “*Clinical Factors*”, mais especificamente na Classe “*Stomatognathic diseases*”. Nas Figuras 30 e 31, página 96 demonstra-se essa utilização na Superclasse “*Pharmacotherapy*”. É de salientar que no *OntoXcancer* se encontram inseridas mais ontologias para além das correspondentes a fatores associados a cancro catalogados pela bibliografia consultada neste trabalho. Desta forma a *OntoXcancer* manter-se-á atualizada.

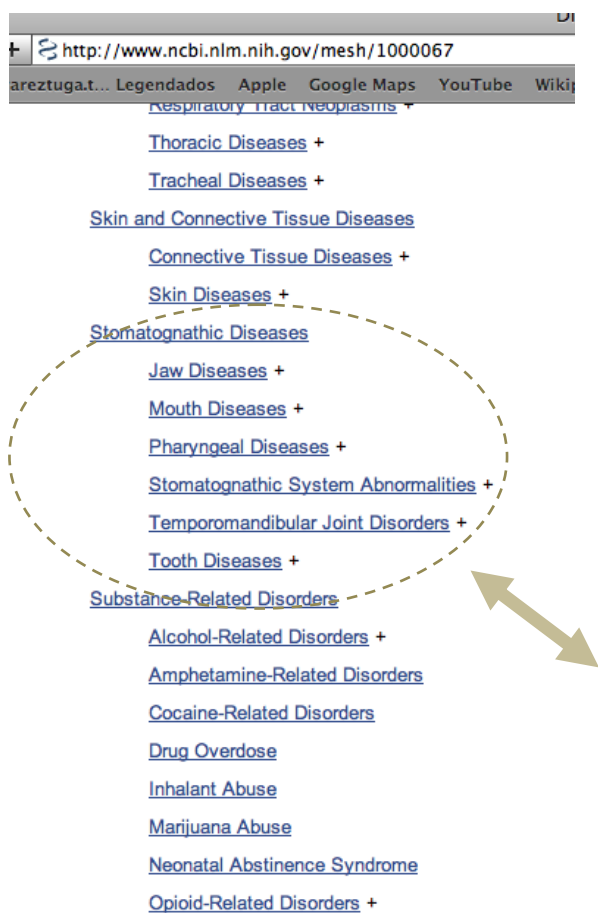


Figura 29 – Excerto representativo da organização hierárquica *Mesh*, referente ao grupo “*Stomatognathic Diseases*”.

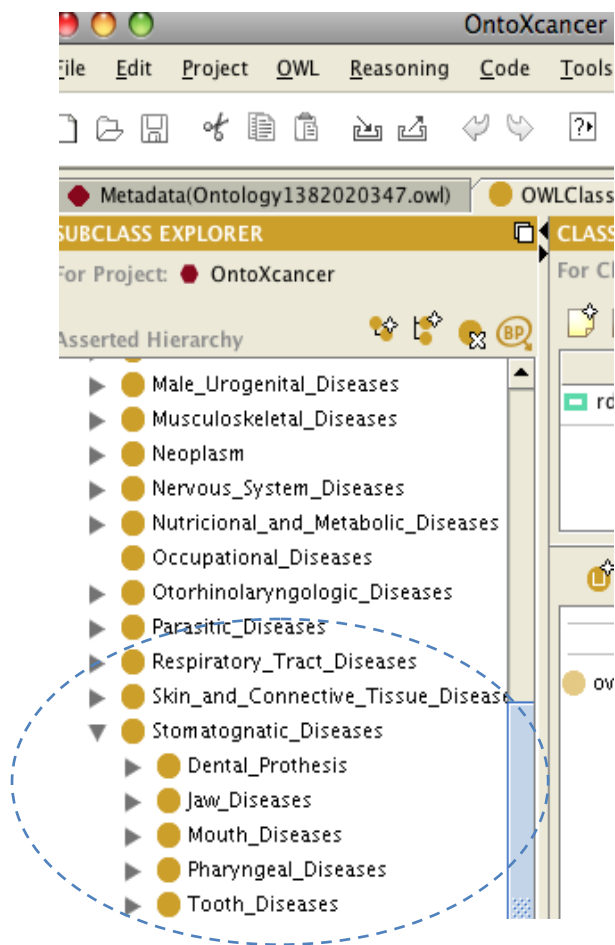


Figura 29 – Excerto representativo da organização hierárquica ontológica referente à classe “*Stomatognathic Diseases*” no *OntoXcancer*.

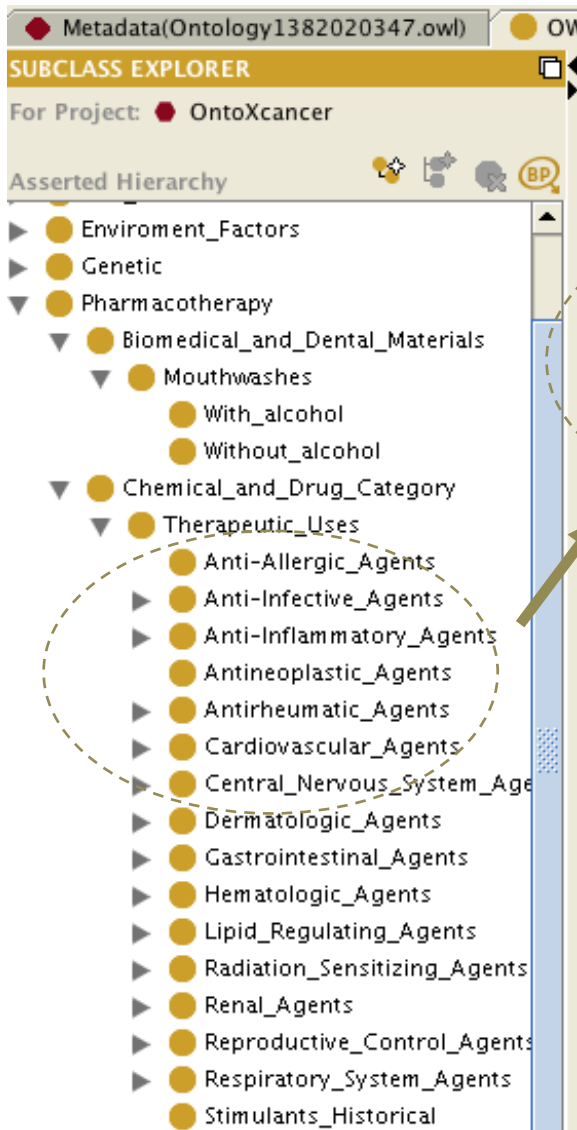


Figura 31 – Excerto representativo da organização hierárquica ontológica da superclasse “Pharmacotherapy” no OntoXcancer.

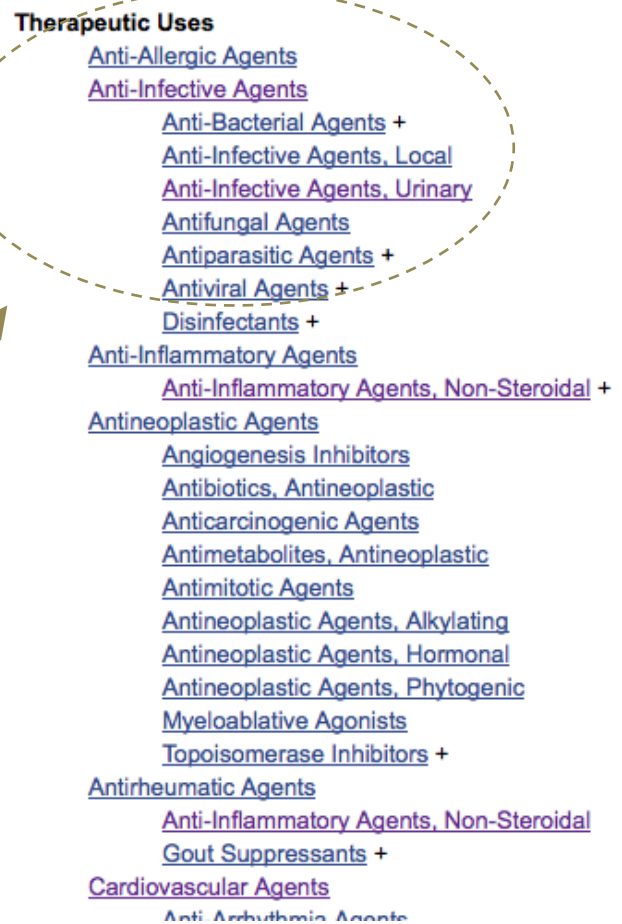


Figura 31 – Excerto representativo da organização hierárquica Mesh, referente ao grupo “Therapeutic agents”.

Não obstante a metodologia descrita para a realização de uma ontologia de raiz, quisemos ir mais longe no trabalho. Dado que quase toda a terminologia catalogada a partir da literatura dedicada aos fatores associados a cancro, descrita anteriormente, já se encontrar, de algum modo, inserida em bibliotecas ontológicas. Assim, justifica-se a utilização da versão menos recente do *Protégé*, explicado nas páginas 68 e 69 do presente capítulo. A versão *Protégé 3.5* detém uma ferramenta que nos permite averiguar se as terminologias em causa já teriam classificação ontológica, através da ligação à biblioteca pública de ontologias *Bioportal*, já as outras versões do programa de criação de ontologias *Protégé* não possuem. Este permite realizar a pesquisa para que o utilizador descubra qual o conceito ontológico que melhor se enquadra à terminologia inicialmente proposta pela base de dados. Toda a informação existente foi posteriormente importada para a biblioteca *OntoXcancer*, como ilustrado na Figura 32, página 98. A mesma figura revela as principais bibliotecas ontológicas a partir das quais adaptámos as ontologias para o *OntoXcancer*, que foram o *Mesh* e a versão modificada do *Mesh* por *Robert Hoehndorf*.

Na construção de toda a hierarquia das Superclasses da ferramenta *OntoXcancer* a partir da bibliografia consultada referente a fatores associados a cancro, aquando da adaptação de ontologias já pré-definidas por outras bibliotecas ontológicas como o *Mesh*, verificámos a existência de terminologias ainda não caracterizadas ontologicamente por nenhuma biblioteca ontológica no *Bioportal*. Achamos então essencial criar ontologias de raiz, definindo-lhe conceitos/*slots*. Como ilustrado pela Figura 33, página 100, o termo que foi necessário introduzir na biblioteca ontológica foi o termo “*Maté*”, ao qual se associou o conceito: ”Is a tea like beverage consumed habitually in South America and among South Americans throughout the world. It is brewed from the dried leaves and stem lets of the perennial tree *Ilexparaguariensis* (yerba mate), a species that belongs to the Aquifoliaceae family⁽²⁶³⁾.”

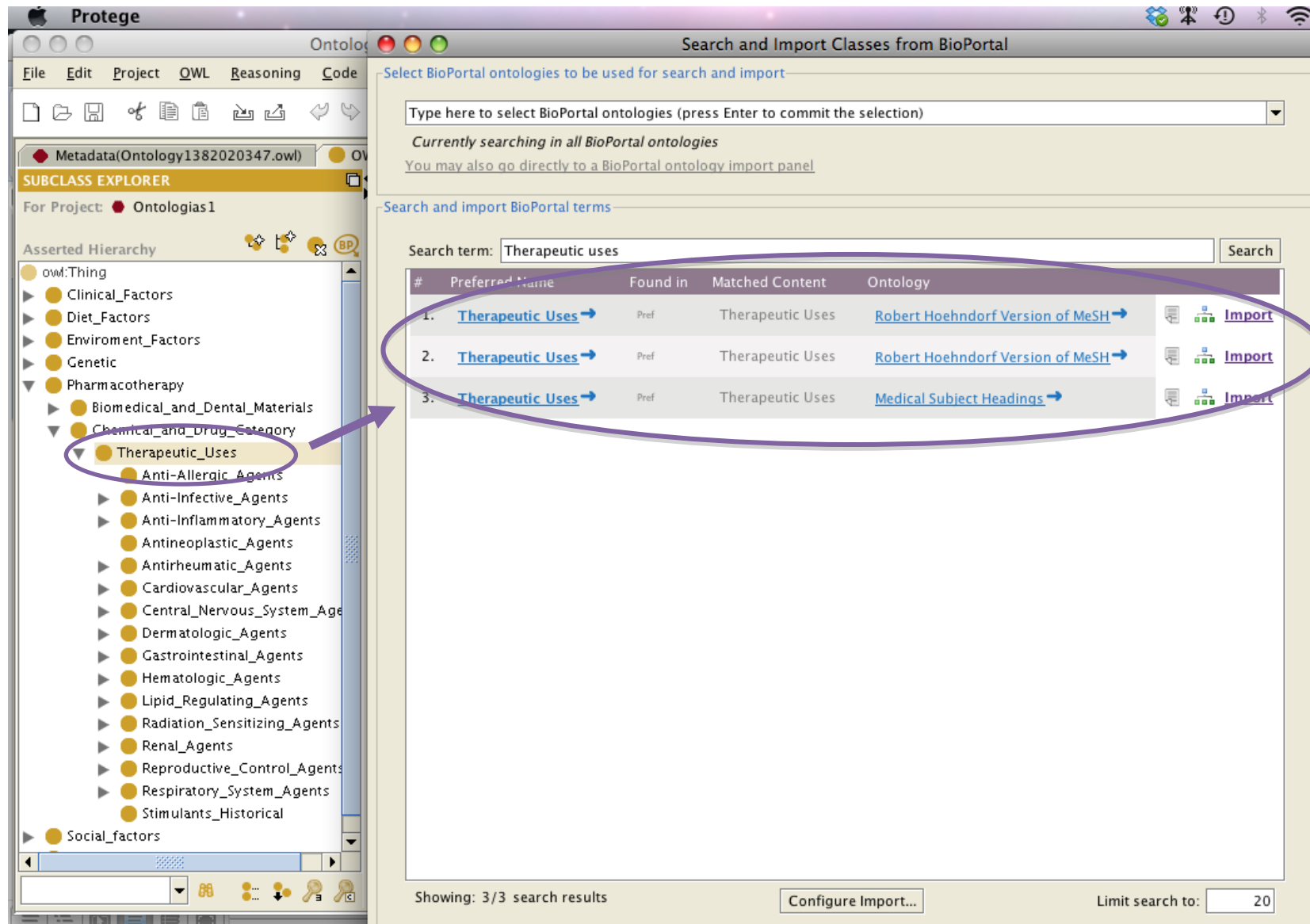


Figura 32 - Representação do método de importação/reutilização de ontologias a partir do programa informático Protégé.

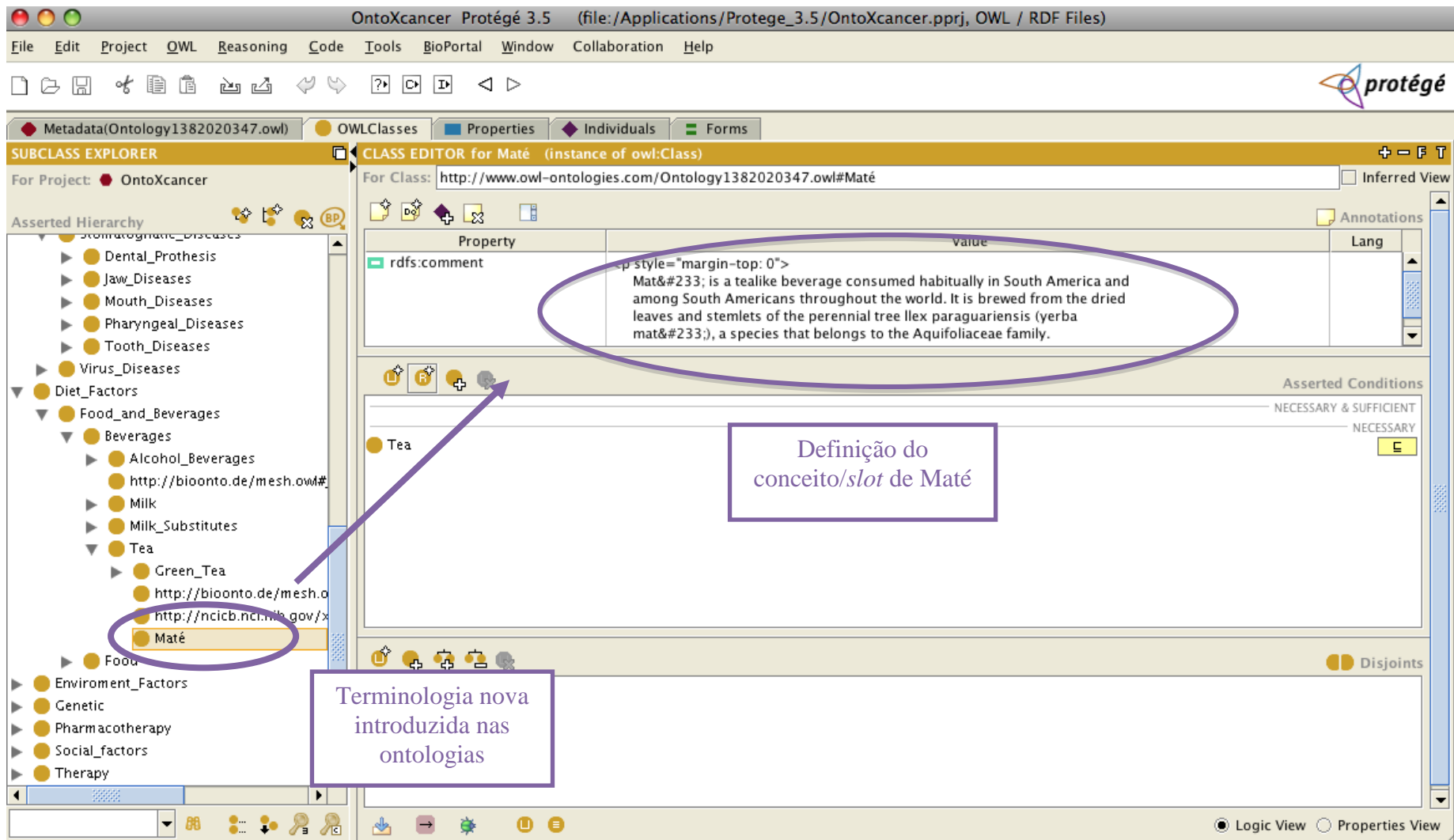


Figura 33 - Representação da construção da ontologia de raiz do termo *Maté*.

Depois de estabelecermos a estrutura ontológica, foi essencial criar redes semânticas, que se associam às ontologias. Já que uma ontologia pode estar inserida em várias Classes como já referido pela presente dissertação e como ilustrado pela Figura 34, página 102. Onde a título de exemplo se apresenta a classe “*Alcohol Beverages*” que tanto se encontra inserida na Superclasse “*Diet Factors*” como na Superclasse “*Social Factors*”.

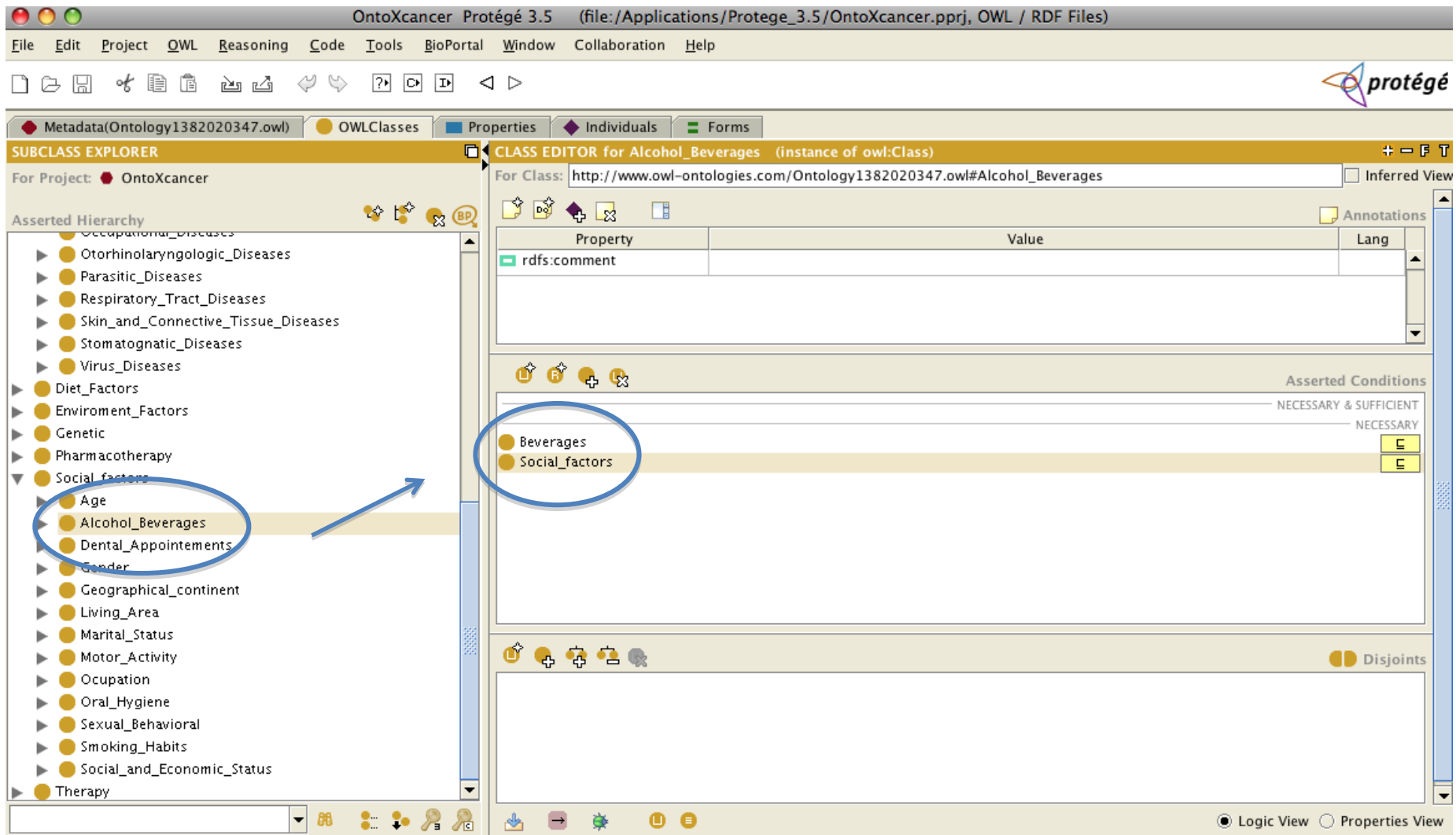


Figura 34 - Representação de execução de redes semânticas no *OntoXcancer*.

4.4.3 Determinação da metodologia para análise estatística da base de dados -*FactorXcancer*

Após terminada a anotação manual da estrutura da base de dados, procedeu-se à análise dos dados inseridos. Esta análise permitiu concluir que, face ao número de citações averiguadas, era possível demonstrar como aplicar métodos estatísticos a este tipo de informação, de forma a obter exemplos de conclusões.

Numa primeira fase procedeu-se à contagem das citações totais depositadas para cada tipo de fator de risco, sem distinção da neoplasia referente, para ser possível um termo de comparação.

Seguiu-se a contagem das citações de cada fator de risco/ protetor referente a cada neoplasia que nos deu como resultado o valor total das citações independentes dos fatores de risco/ protetor característicos para cada tipo de neoplasia. A aplicação de filtros à base de dados permite retirar a informação característica de cada cancro. Com os resultados obtidos executa-se uma comparação dos valores de forma a ser observada a existência de fatores específicos ou comuns de cancro. Para esta comparação primeiro utiliza-se uma fórmula “*Ratio*” em vez da habitual subtração, dado que, pela subtração, não seria possível realizar uma relação numérica entre os valores necessários para obtenção de resultados. Dado que os valores de preenchimento da base de dados são positivos e negativos e abrangem uma grande escala aplica-se posteriormente a fórmula “*Logarithmic*”. A utilização dos logaritmos dos valores em vez dos valores reais reduz a amostra para um tamanho mais exequível.

Para evitar a obtenção de extremos quando existe uma ausência de amostra (ou da ausência de estudos) aplica-se a fórmula “*additivesmoothing*”, de modo a evitar erros. Por último calcula-se a diferença entre $\log_2\text{Ratio}$ de todos os tipos de cancro, de forma a obter a relevância do fator comparado com toda a população.

Numa segunda fase utilizou-se o programa informático SPSS que é um *software* aplicativo científico da IMB que permitiu relacionar as variáveis/ontologias de forma a obtermos os fatores com maior preponderância na carcinogénese dos cancros caracterizados como sendo da “cabeça e pescoço”, dado que eram os cancros com maior número de citações.

Após a inserção em SPSS dos dados constantes na base de dados realizada em Excel foi feita uma análise fatorial e uma análise de componentes principais. Estas técnicas estatísticas têm como objetivo representar ou descrever um número de variáveis iniciais a partir de um menor número de variáveis hipotéticas (os fatores/ontologias). Isto é, permite identificar novas variáveis (os fatores/ontologias principais), em menor número que o conjunto inicial, mas sem perda significativa da informação contida neste conjunto.

Foi realizada uma pesquisa dos dados em análise que permitiu identificar 310 variáveis no registo correspondente à base de dados dos cancros pertencentes ao termo “cabeça e pescoço” e 97 variáveis na informação correspondente ao cancro do esófago, que poderiam ser usadas para levar a cabo tal propósito. Obviamente seria intratável interpretar 310 descritores para ter acesso aos mais significativos. Este trabalho seria então simplificado se estas 310 variáveis fossem reduzidas a um número inferior de novas variáveis, tornando-se necessário usar uma técnica de redução de dados. O propósito geral destas técnicas é encontrar uma forma de condensar a informação contida num conjunto de variáveis originais, num conjunto menor de variáveis perdendo o mínimo possível de informação. Tratam-se portanto de técnicas de redução de dados que investigam as correlações entre as variáveis e as descrevem, se possível, em termos de um menor número de variáveis.

Os resultados desta análise são analisados e discutidos no capítulo Resultados e Discussão (páginas 107-124).

Resultados e Discussão

5. RESULTADOS E DISCUSSÃO

A parte mais importante de uma terminologia, de acordo com as normas International Standards Organization (ISO), é todo o procedimento que determina o conceito que lhe está subjacente, ou seja todo o trabalho e pesquisa necessário para o surgimento de uma terminologia. Infelizmente é muito difícil realizar uma análise em profundidade de todas as terminologias presentes no meio científico, precisamente por causa das muitas ambiguidades ao nível dos seus termos e definições. Além disso, em muitos casos, pode apenas especular-se quais as verdadeiras entidades que devem qualificar-se como exemplos para os conceitos que estão incluídos nas ontologias⁽⁶⁾. O volume de bibliografia analisada na presente dissertação forneceu uma base coerente de informação a este nível (de ambiguidades entre termos) No entanto, mantém-se o problema da falta de uniformidade de termos e medidas no meio científico. Mesmo entre bibliotecas ontológicas existe alguma ambiguidade entre definições de termos, o que dificulta o caminho para um consenso na comunidade científica.

Com o desenvolvimento da base de dados inicial deparámo-nos com uma grande dispersão de informação, conceitos e terminologias referentes a fatores associados a cancro, surgindo deste modo a necessidade de definir um sistema de ontologias que colmatasse estas falhas.

Acreditamos que a proposta de organização de ontologia que apresentamos, o *OntoXcancer*, proporciona uma uniformização da terminologia catalogada.

O desenvolvimento do *OntoXcancer* baseou-se na reutilização de terminologias ontológicas e de conhecimentos pré-existentes em bibliotecas ontológicas (como o *Mesh*, *Robert Hoehndorf version of Mesh* o *HUGO Gene Nomenclature Committe*) bem como na criação de novas terminologias de raíz. Isto facilitou a construção de um sistema, que proporciona novos avanços no domínio da investigação de fatores associados a cancro a partir de uma gestão adequada da imensa quantidade de informação disponível atualmente na web.

A falta de uma infra-estrutura de informação constantemente atualizada, aberta e compartilhada com toda a comunidade científica constitui um dos maiores entraves à evolução na partilha de dados científicos. Isto impede que se extraia e analise diferentes

fontes de dados, dificultando a aquisição de conhecimento e tornando mais lento o processo de divulgação. Outro dos obstáculos encontrados foi a dificuldade na aquisição de todas as ontologias existentes na pesquisa biomédica. Talvez devido ao isolamento de várias comunidades científicas, que não colocam em domínio público as suas bibliotecas ontológicas. Face a todos estes obstáculos, o desenvolvimento e implementação de uma infraestrutura de ontologias, *OntoXcancer*, adquire especial importância, principalmente ao ser compartilhada com toda a comunidade científica, tornando-se pública pelo *Bioportal*.

O *OntoXcancer* caracteriza-se então como sendo uma ferramenta bioinformática com o objectivo de catalogação de ontologias que integra Classes/conceitos, alguns reutilizados de várias bibliotecas ontológicas de fatores associados a cancro. Esta ferramenta bioinformática foi criada a partir da base de dados inicial, resultante do levantamento da informação presente na literatura. Uma vez estabelecido o sistema de ontologias a base de dados inicial foi convertida de modo a integrar o novo sistema classificativo, criando-se uma base de dados dedicada a fatores com implicação em cancro, baseado em ontologias, o *FactorXcancer*.

As informações detalhadas fornecidas sobre uma ampla gama de termos médicos e científicos constituem uma importante e útil fonte de dados para o meio científico, apenas e só, se esta se encontrar numa ferramenta organizada, atualizada e compartilhada. Dada a evolução rápida de resultados de investigações e de melhoramento de processos clínicos relacionados com a prevenção em cancro, a ferramenta *FactorXcancer* deve ser continuamente preenchida/atualizada.

5.1 Exemplos de redes semânticas do *OntoXcancer*

Uma rede semântica é a ligação que existe entre as ontologias e as diferentes classes hierárquicas. Como referido anteriormente, foi necessário originar estas redes para perceber as associações entre as diferentes ontologias.

A Figura 35 representa uma rede semântica da ontologia “Anemia de Fanconi”, que é uma patologia de origem genética, que tanto se encontra organizada hierarquicamente na Superclasse “*Genetic*” como na Superclasse “*Clinical Factors*”.

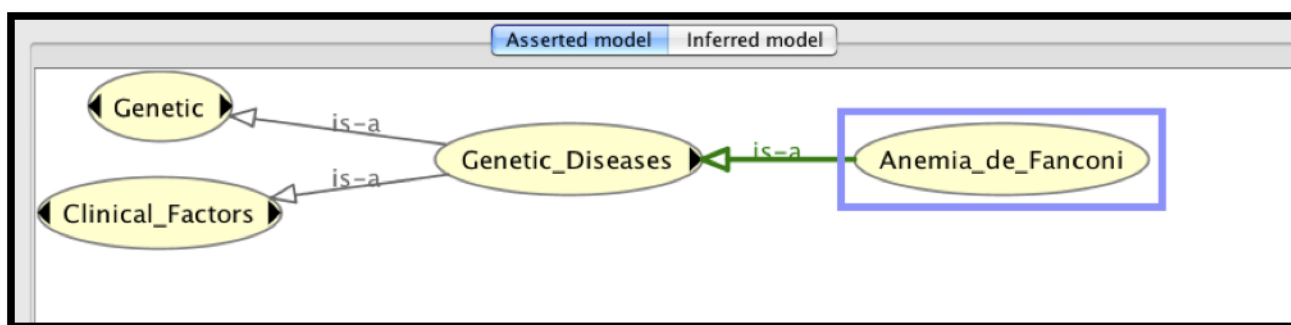


Figura 35 - Excerto representativo da rede semântica da ontologia “Anemia de Fanconi”, que se encontra inserida tanto na Superclasse “*Genetic*” como na Superclasse “*Clinical Factors*”.

Outro exemplo de uma rede semântica é apresentado pela Figura 36, onde a ontologia “*Gastric Neoplasm*” é inserida tanto na Classe “*Digestive System*”, como na Classe “*Neoplasm*” ambas pertencentes à Superclasse hierárquica “*Clinical Factors*”.

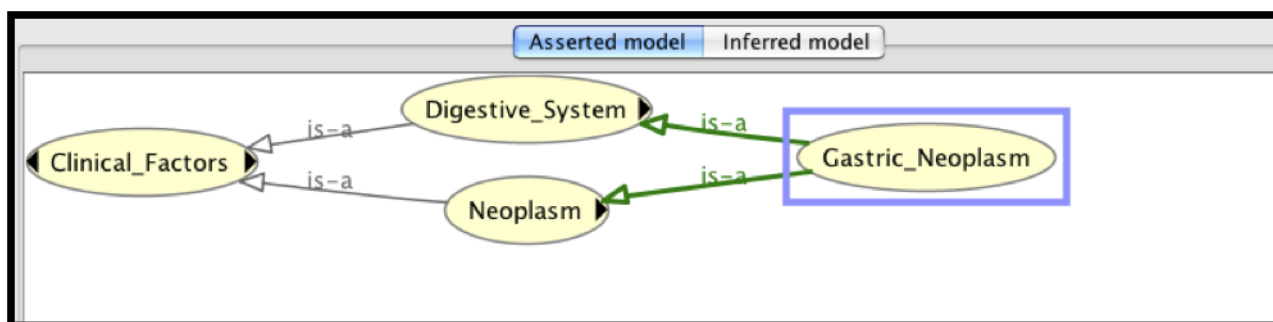


Figura 36 - Excerto representativo da rede semântica da ontologia “*Gastric Neoplasm*”, que se encontra inserida tanto na Classe “*Digestive system*” como na Classe “*Neoplasm*”, ambas pertencentes à Superclasse “*Clinical Factors*”.

Neste exemplo ilustramos a relação existente na ontologia “*Xeroderma Pigmentosum*” que é uma Subclasse na organização hierárquica no *OntoXcancer*. Como se pode averiguar pela Figura 37, a subclasse em questão encontra-se inserida na classe “*Genetic Diseases*” que pertence tanto à Superclasse “*Clinical Factors*” como à Superclasse “*Genetic*”. Além disso, a ontologia “*Xeroderma Pigmentosum*” já contém genes identificados com relação à patologia, tais como: XRCC3, XRCC1, ERCC5, ERCC2 E MGMT.

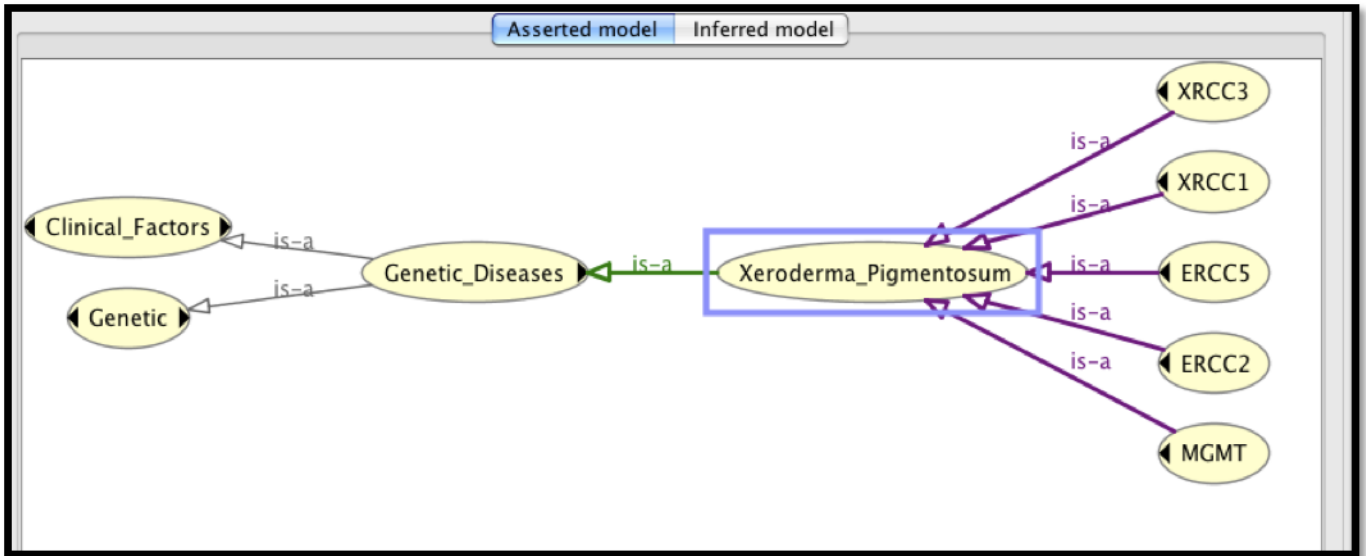


Figura 38 - Excerto representativo da rede semântica da ontologia “*Xeroderma Pigmentosum*”, que se encontra inserida tanto na superclasse “*Genética*” como na superclasse “*Clinical Factors*”, ambas pertencentes à Superclasse “*Clinical Factors*”.

Noutro exemplo um pouco mais complexo, onde a ontologia “*Carcinogenic Aromatic amines*”, pertencente à Classe “*Tobacco Associated Carcinogen*” e à Superclasse “*Environment Factors*”, também inserida na Subclasse “*Tobacco*” que pertence à Classe “*Smoking Habits*” e subsequente Superclasse “*Social Habits*”. Aqui vemos a importância das redes semânticas na extração de evidências, já que observa-se claramente uma relação entre as diferentes subclasses e classes. Ou seja, a mesma ontologia pode pertencer a diferentes Classes e Subclasses.

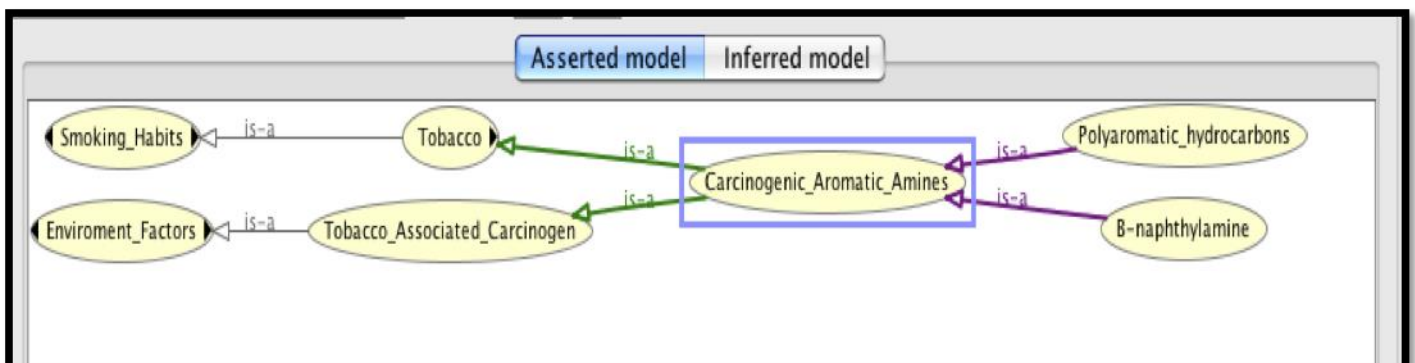


Figura 37 - Excerto representativo da rede semântica da ontologia “*Carcinogenic Aromatic amines*”, que se encontra inserida tanto na classe “*Tobacco Associated Carcinogen*” como na subclasse “*tobacco*”, dado à sua relação.

5.2 Exemplos de como a informação disponível no –*FactorXcancer* – pode ser consultada

A estrutura da base de dados *FactorXcancer* foi preenchida de acordo com os critérios enumerados anteriormente no capítulo Materiais e Métodos, na página 62. Após a anotação de toda a informação disponível na bibliografia consultada dedicada a fatores associados a cancro, exemplifica-se de como esta informação pode ser consultada.

FactorXcancer contém informação referente a vários tipos de cancros (como descrito pelas tabelas I e II, capítulo Materiais e Métodos páginas 56-104) e respectivos fatores implicados.

É pertinente reforçar a ideia de que o termo “cabeça e pescoço” é um termo coletivo que inclui o cancro oral, laringe, orofaringe, hipofaringe e nasofaringe, mas na catalogação realizada para preenchimento do *FactorXcancer* foram tratados de forma independente, ou seja, foram catalogados consoante a sua descrição na bibliografia consultada, sem ter em conta o facto de pertencerem a um termo coletivo.

Um dos objetivos desta dissertação era exemplificar formas de consulta de informação depositada na base de dados construída a *FactorXcancer*. Por forma a se poder comparar a informação existente sobre cancro oral comparativamente a outros cancros “cabeça e pescoço”, os dados referentes às ontologias que descrevem os fatores associados a cancros pertencentes ao termo coletivo “cabeça e pescoço”, foram consideradas em conjunto. Ou seja, nesta análise, o termo “cabeça e pescoço” abrange as ontologias a que correspondem os fatores associados aos cancros da laringe, orofaringe, hipofaringe e nasofaringe, deixando de parte os dados referentes a cancro oral.

No seguimento da análise realizada, no que diz respeito ao número de cancros referentes ao termo “cabeça e pescoço”, o gráfico 1 representa o número de citações depositadas no *FactorXcancer* para cada tipo de cancro que integra o termo anteriormente descrito. Com a análise deste gráfico é possível inferir que, para os diferentes cancros pertencentes ao grupo “cabeça e pescoço”, o número de citações é bastante distinto, sendo escassa a informação sobre alguns deles. Não será de mais voltar a referir que as conclusões apresentadas estão dependentes do número de referências que estão até ao momento depositadas na *FactorXcancer*. À medida que

mais informação for depositada este tipo de gráfico vai traçando o panorama da informação existente sobre os diferentes tipos de cancro.

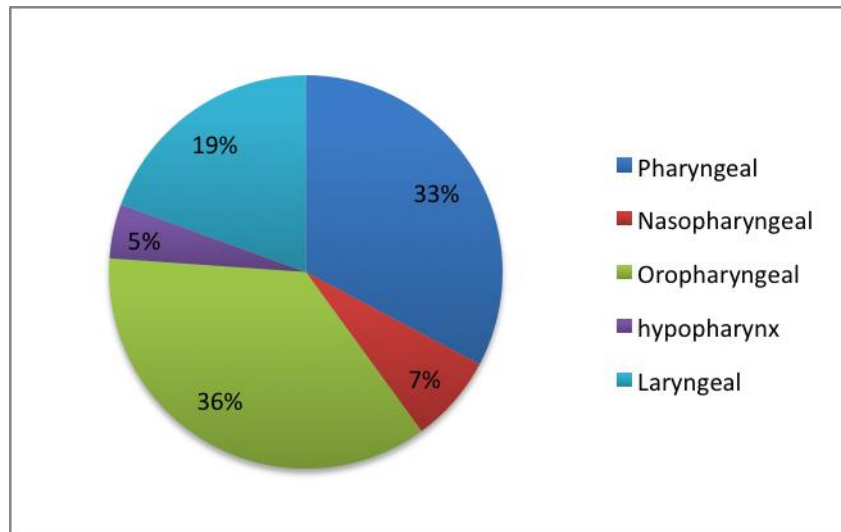


Gráfico 1 - Comparação entre o número de citações depositadas no *FactorXcancer* para os diferentes tipos de cancros que integram a designação “cabeça e pescoço”.

Se quiséssemos agora saber o número de citações depositados no *FactorXcancer*, referentes ao cancro oral e para os cancros que integram o termo “cabeça e pescoço”, a análise do gráfico 2 permite inferir que atualmente se encontram mais dados depositados sobre os cancros inseridos no termo “cabeça e pescoço” do que referente ao cancro oral.

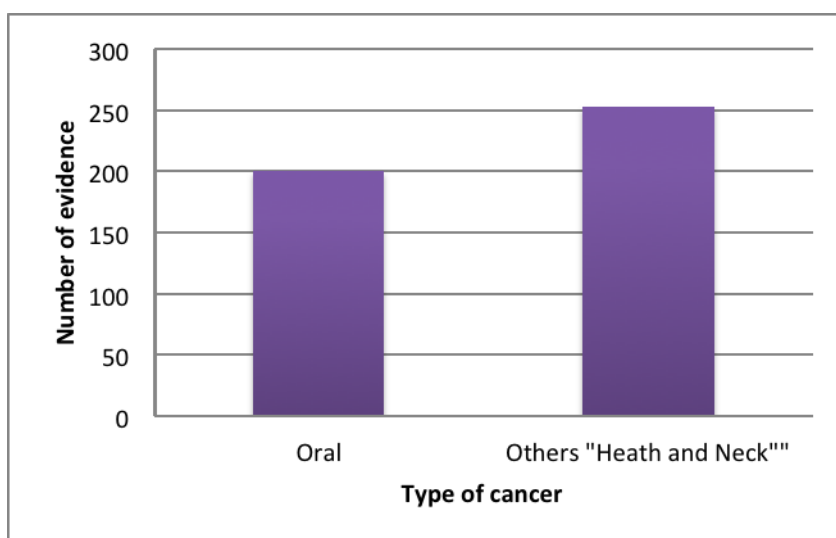


Gráfico 2 - Comparação do número de citações depositadas no *FactorXcancer*, entre o cancro oral e os outros tipos de cancros classificados como pertencendo ao termo “cabeça e pescoço”.

Por outro lado, pela análise do gráfico 3, podemos averiguar qual das várias Superclasses existentes, relativas a cancro oral e aos outros cancros classificados com o termo “cabeça e pescoço”, tem maior número de citações. Verificamos que as Superclasses com mais citações bibliográficas atualmente depositadas no *FactorXcancer* são: “*Social Habits*”, “*Clinical Factors*” e “*Genetic*”.

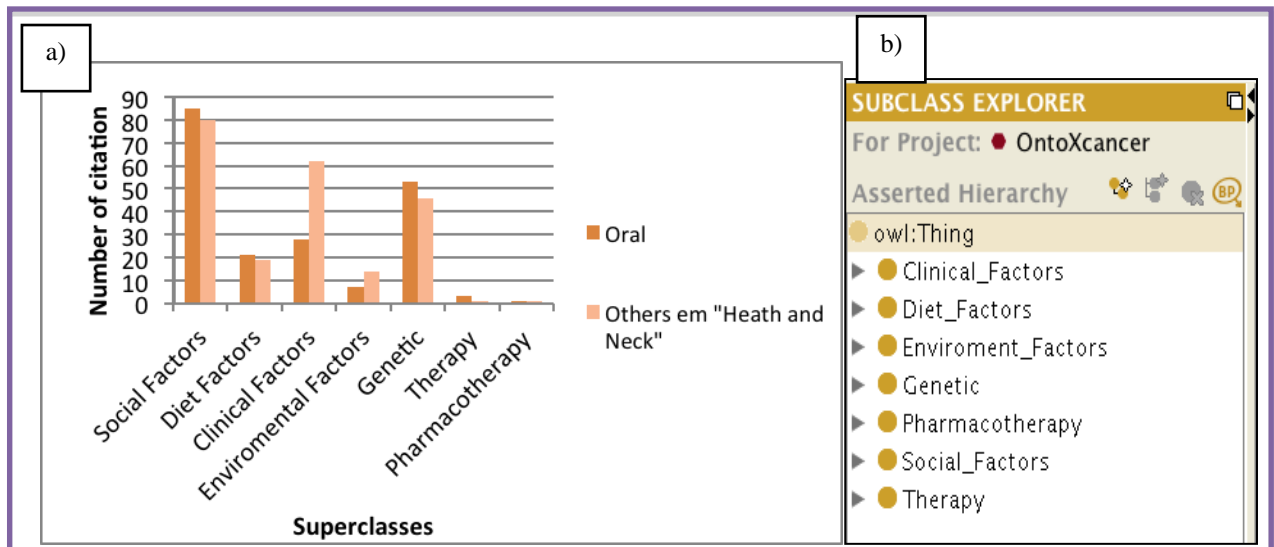


Gráfico 3 - a) Comparação do número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros da “cabeça e pescoço”, correspondente às diferentes Superclasses.. b) Excerto da rede de ontologias em que se inserem as correspondentes aos fatores associados às Superclasses representadas em a).

No sentido de acompanharmos hierarquicamente as ontologias com maior número de citações, iremos comparar este número de citações entre o cancro oral e os outros tipos de cancros que integram o termo “cabeça e pescoço, para as Superclasses “*Social Habits*”, “*Clinical Factors*” e “*Genetic*”.

Pelo gráfico 4, página 114, podemos ver que existem mais citações alusivas a ontologias referentes ao cancro oral do que relativas aos vários tipos de cancro pertencentes ao termo “cabeça pescoço”, nas várias Superclasses apresentadas. Indicando deste modo que atualmente estão depositados mais dados sobre cancro oral do que sobre cancros que integram o termo “cabeça e pescoço”. Também se verifica que o cancro orofaríngeo possui o mesmo número de citações referentes a ontologias que o cancro oral na Superclasse “*Clinical Factors*”.

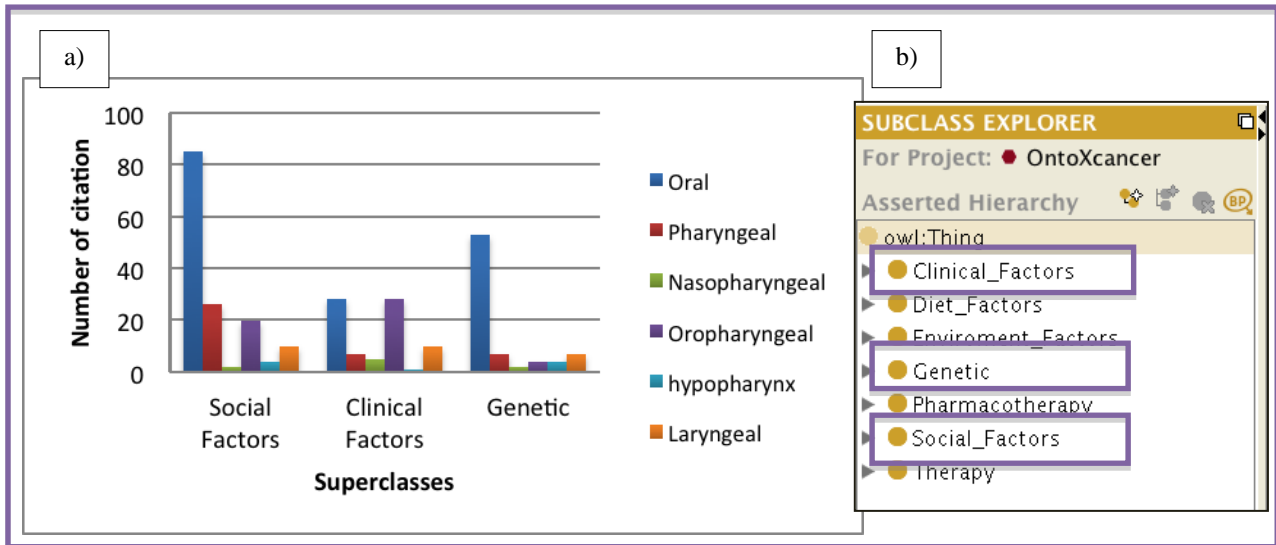


Gráfico 4 - Comparação entre o número de citações depositadas atualmente no FactorXcancer referente a cancro oral, faríngeo, nasofaríngeo, orofaríngeo, hipofaríngeo e laríngeo correspondente às Superclasses com maior número de citações depositadas até à data. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às Superclasses representadas em a).

Seguindo a mesma linha de análise, é também possível saber o número de citações, referentes às ontologias a que correspondem os fatores de risco depositadas para cada Classe com maior número de citações, que integram as Superclasses referidas no gráfico 4.

O gráfico 5, página 115 mostra que dentro da primeira Superclasse com maior número de citações: “*Social Factors*”, as classes de ontologias como maior número de citações depositadas até a data no FactorXcancer são “*Smoking and Alcohol Beverages*” no cancro oral e “*Alcohol Beverages*” em cancros inseridos no termo “cabeça e pescoço”. É de salientar que estas classes constituem descritores mais específicos da Superclasse “*Social Factors*”.

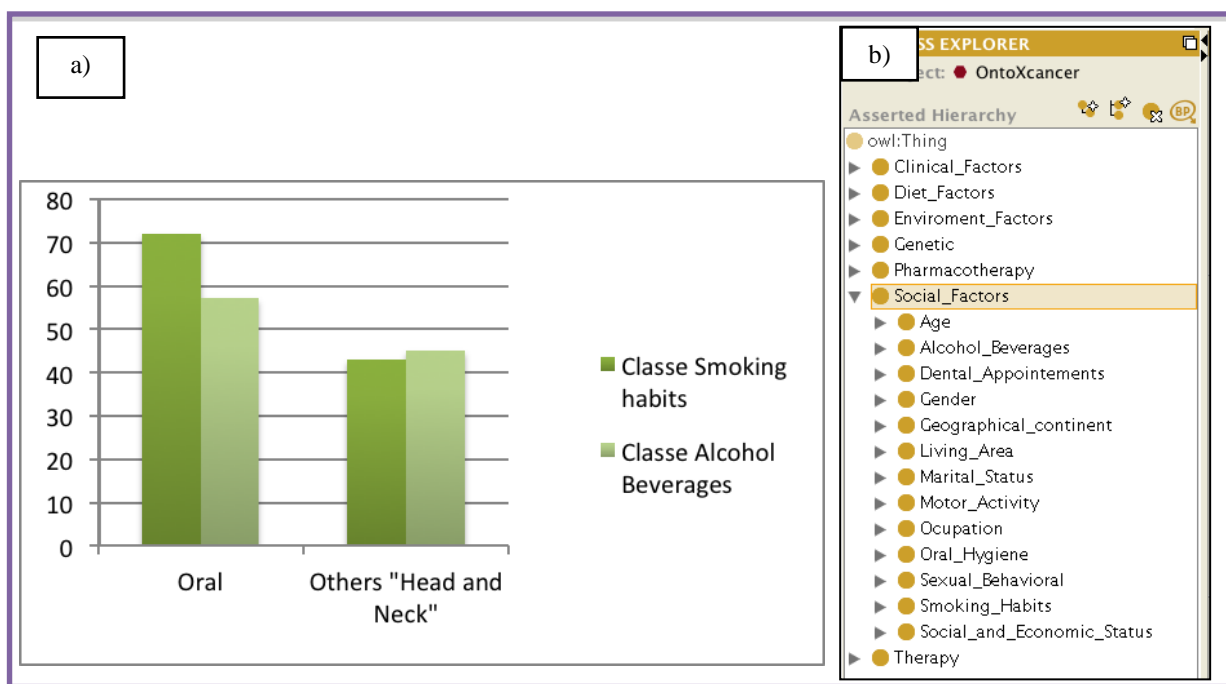


Gráfico 5 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo "cabeça e pescoço", alusivos às classes "Smoking and Alcohol Beverages. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).

Em relação à segunda Superclasse com maior número de citações: "Clinical Factors", a partir da análise dos dados depositados no FactorXcancer, verifica-se no gráfico 6, página 116, que as ontologias a que correspondem os fatores com maior número de citações se encontram organizadas na classe "Infections", tanto em cancro oral como nos outros cancros que totalizam o termo "cabeça e pescoço". A análise deste gráfico permite também inferir que existem poucos dados referentes às Subclasses "Bacterial", e "Eukaryota", depositados no FactorXcancer.

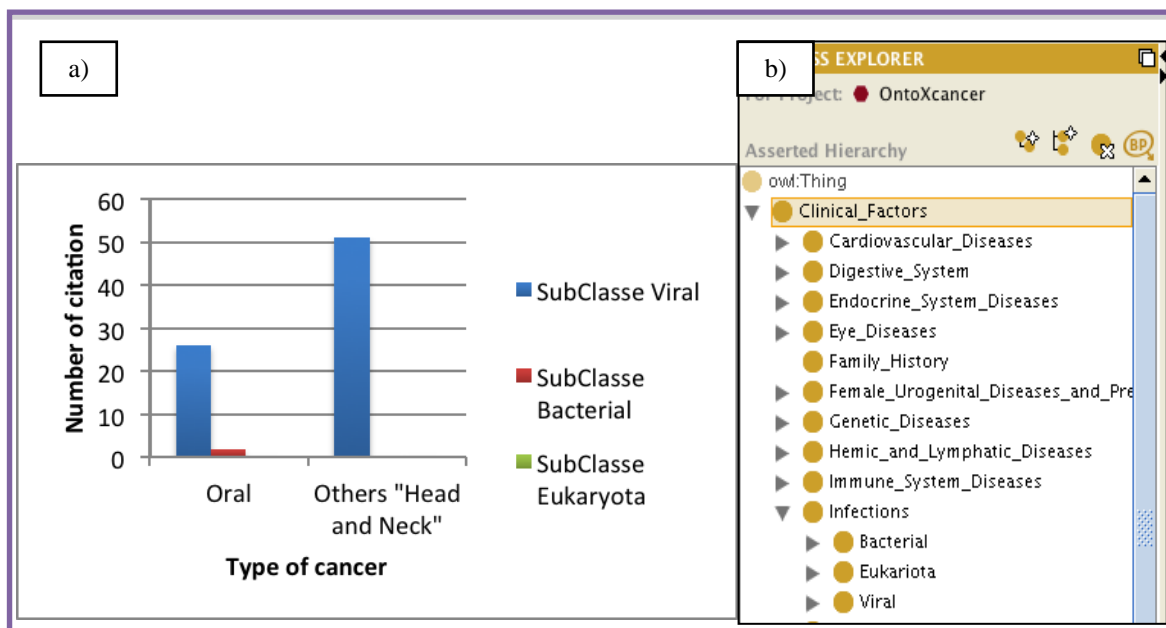


Gráfico 6 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivos às subclasses que pertencem à classe “*Infections*”, b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às subclasses representadas em a).

Pela mesma lógica em relação à Superclasse “Genetic”, esta integra as ontologias referentes aos fatores de risco integrados nas Classes “Gene or Polymorphism”, “Genes by Function e “Genetic Diseases”, o gráfico 7 compara o número de citações depositadas no *FactorXcancer* entre cancro oral e os outros cancros classificados como cancro da “cabeça e pescoço”. A análise deste gráfico permite inferir que também aqui existem poucos dados referentes à Classe “Genes By Function”.

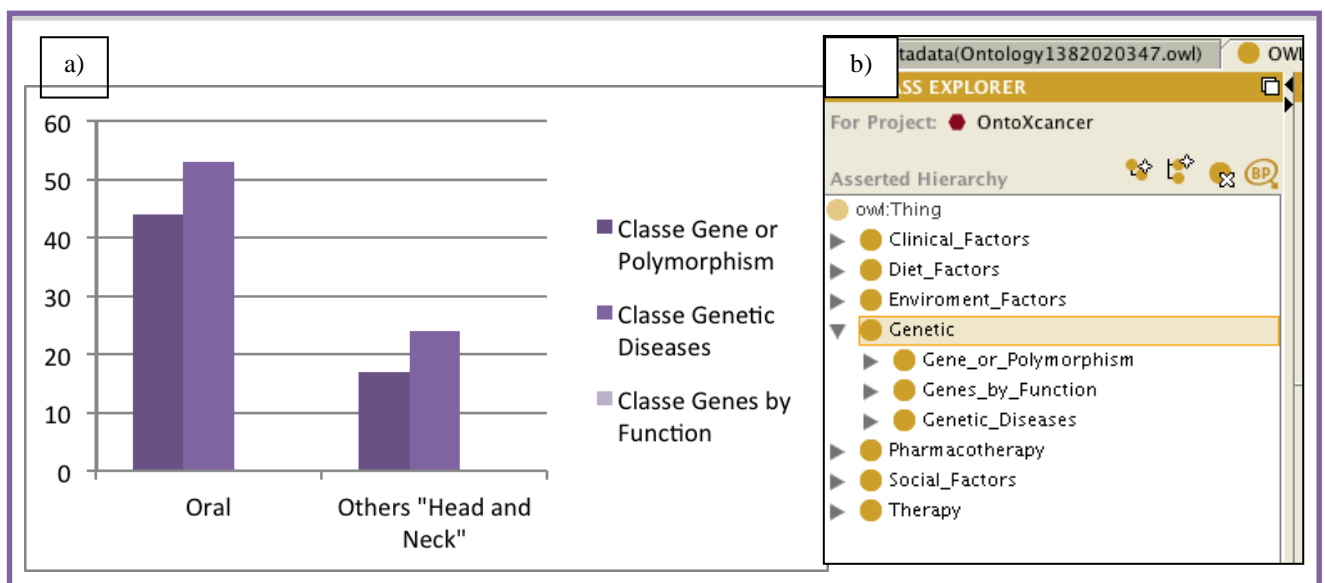


Gráfico 7 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas às classes “Gene or Polymorphism”, “Genes by Function e “Genetic Diseases”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).

Do mesmo modo, podemos também fazer este tipo de análise agora referente ao número de citações para as ontologias das subclasses, englobadas em cada classe referida anteriormente.

Continuando a avaliação do número de citações anotadas no *FactorXcancer* que integram as ontologias referentes aos factores da Subclasse que integra a Classe “*Smoking Habits*” apresenta-se o gráfico 8. Este compara o número de citações existentes até a data destas Subclasses, entre cancro oral e cancros caracterizados como “cabeça e pescoço”, onde se verifica que, as Subclasses com maior número de citações depositadas no *FactorXcancer* é a Subclasse “*Tobacco*”, tanto relativo a cancro oral como relativo aos outros cancros caracterizados, como sendo da “cabeça e pescoço”. Outra Subclasse que também apresenta elevadas citações é a correspondente a “*Smokeless*”.

Neste momento, para as restantes Subclasses como: “*Bidi*”, “*Betel quid or Areca Nut*”, “*Cigars*”, “*Opium*”, e “*Ex-Smoker*”, o número de dados depositados até à data ainda é reduzido.

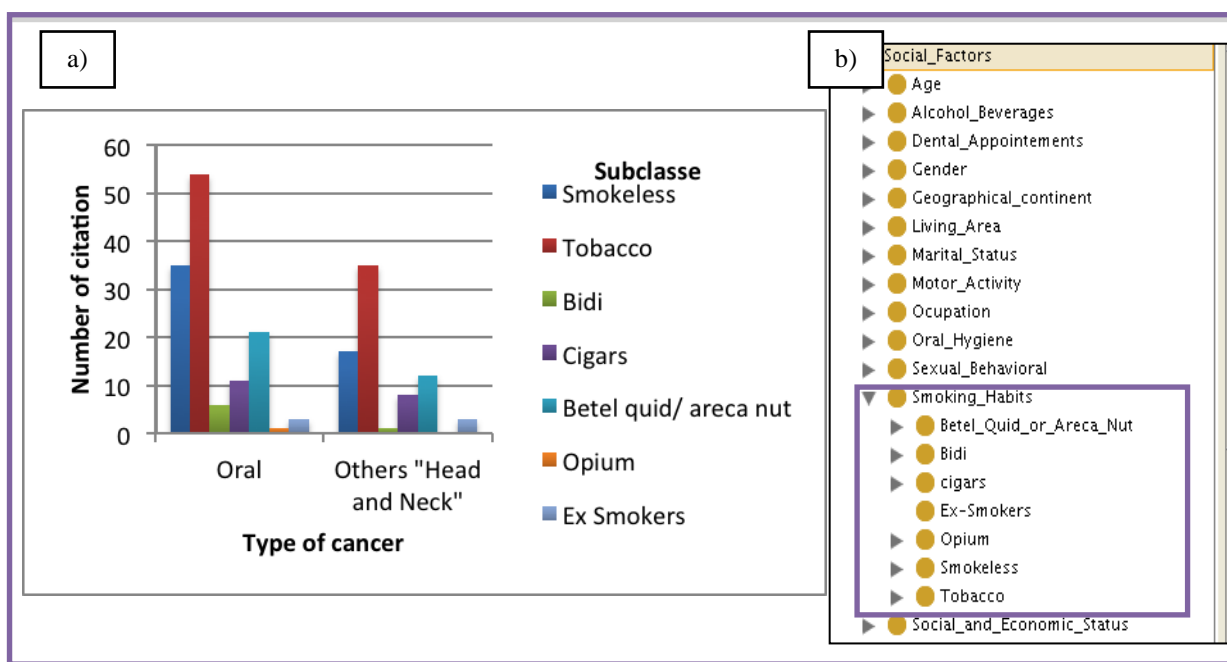


Gráfico 8 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas à classe “*Smoking Habits*”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).

Para saber o número de citações depositadas no *FactorXcancer* referentes às Subclasses que abrangem a classe “*Alcohol Beverages*”, elaboramos o gráfico 9, página 119, onde podemos constatar que o número de citações superiores depositadas no

FactorXcancer referente a ontologias referentes aos fatores associados a cancro pertence à subclasse “*Others*”, tanto no cancro oral como nos cancros pertencentes ao termo “cabeça e pescoço”. De onde inferimos que o numero de dados referente à Subclasse “*Wine*” é substancialmente reduzida em ambos os cancros representados.

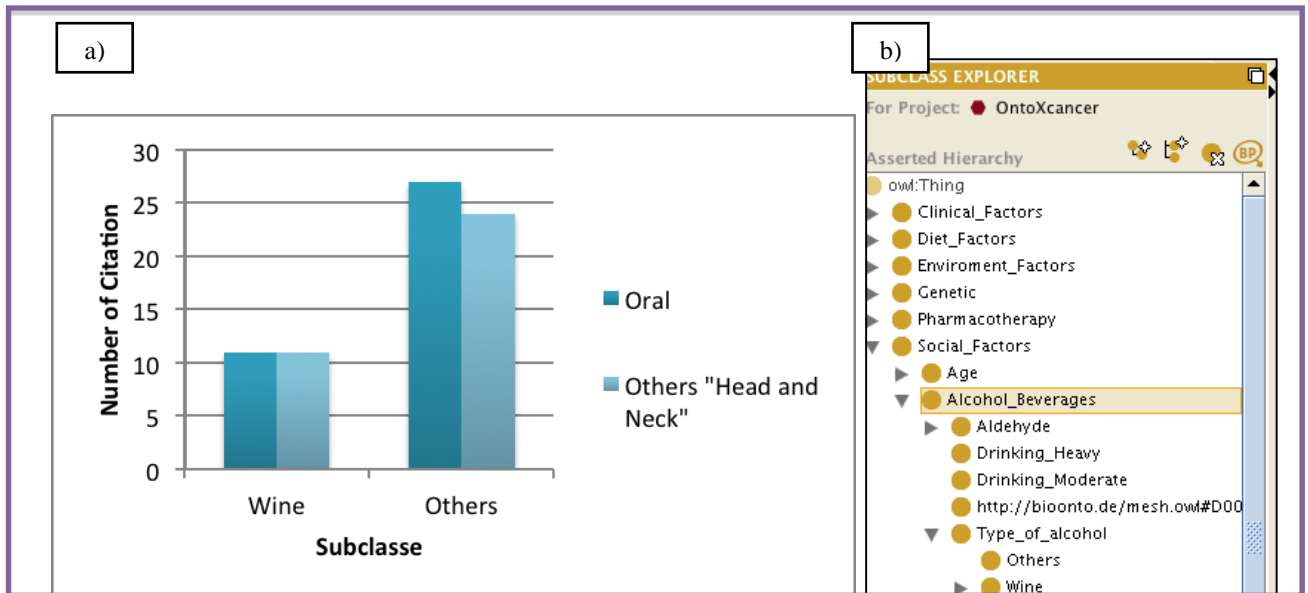


Gráfico 9 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas à classe “*Alcohol Beverages*”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).

Observando o gráfico 6, página 116, a única Subclasse com o número de citações significativo é a subclasse “*Viral*”, pois as restantes Subclasses, “*Bacterial* e “*Eukariota*”, não apresentam citações suficientes. Para averiguar qual o número de citações existentes no *FactorXcancer* referente à ontologia que corresponde a um fator viral, analisámos o gráfico 10, página 120, que mostra que o vírus com maior número de citações depositadas no *FactorXcancer*, tanto no cancro oral como nos cancros que pertencem ao termo “cabeça e pescoço”, é o HPV. Podemos ver também no gráfico 10, ausência de dados referentes a outros vírus.

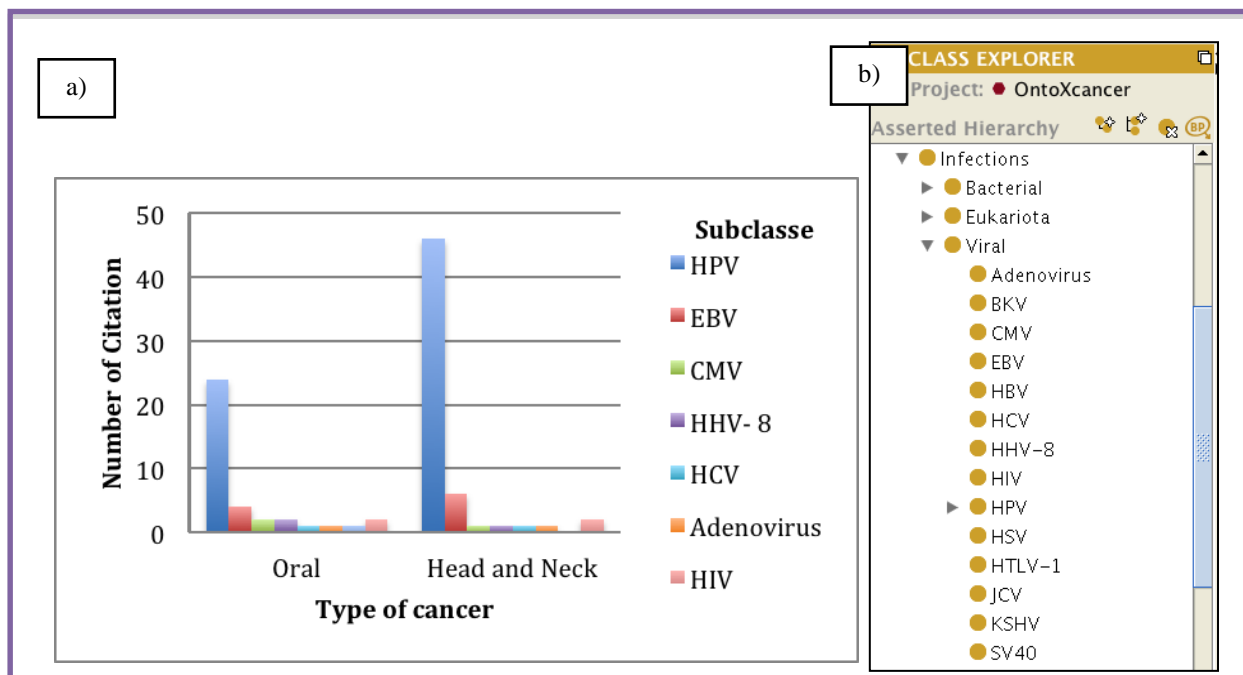


Gráfico 10 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusiva à subclasse “Viral”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).

Por último, resta avaliar o número de citações depositadas no *FactorXcancer* referente à Classe “*Genetic Diseases*” pertencente à Superclasse “*Clinical Factors*”. Analisando então o gráfico 11, página 121, podemos inferir que, das Subclasses apresentadas, a subclasse “*Anemia de Fanconi*” possui um maior número de citações depositadas nos cancros tipos da “cabeça e pescoço”. Outra evidência interessante é a presença exclusiva de citações referentes à Subclasse “*Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy*” em cancro oral e de “*Bloom Syndrome*” apenas nos cancros descritos com o termo “cabeça e pescoço”.

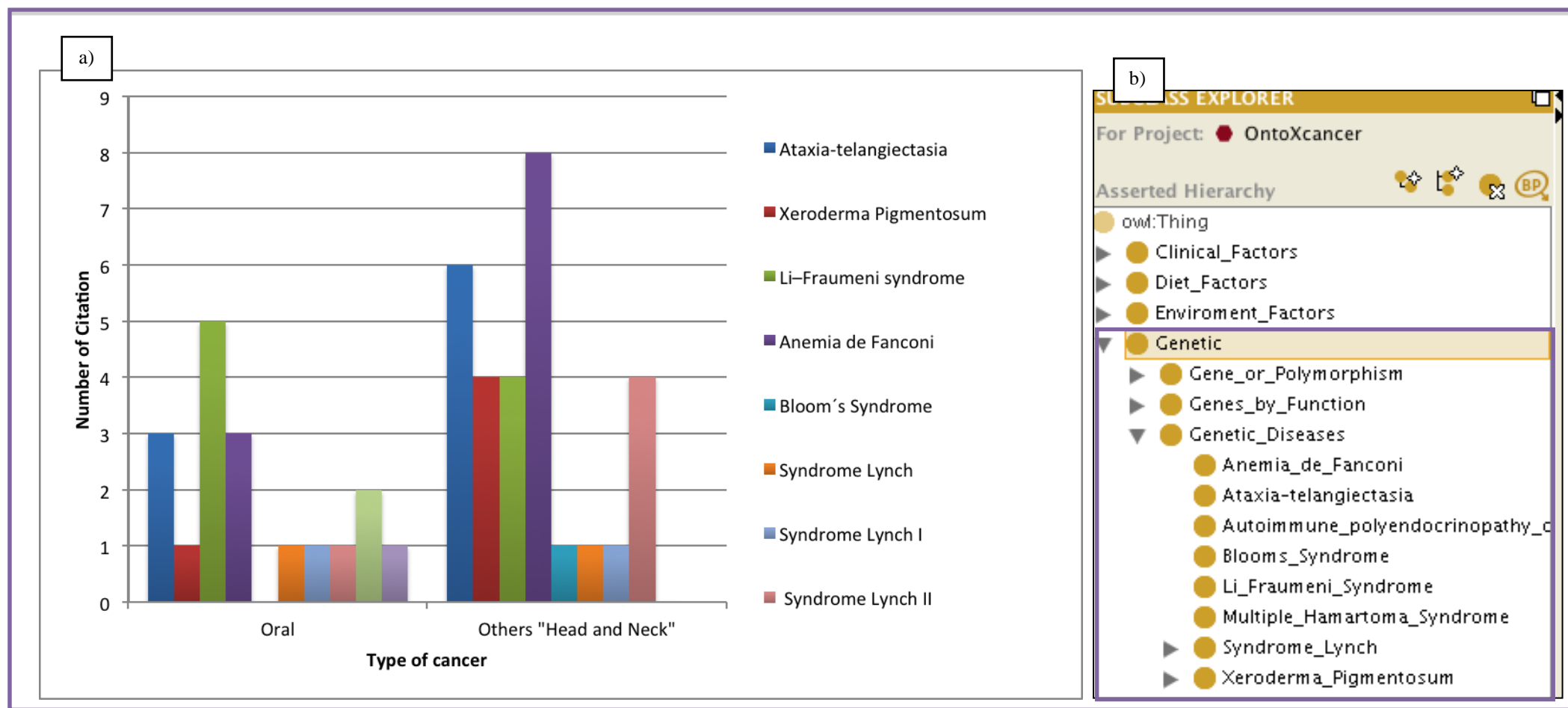


Gráfico 11 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusiva classe “Genetic Disorders”. b) Excerto da rede de ontologias em que se inserem as ontologias

5.3 Resultados obtidos pelo tratamento estatístico

Face à demonstração feita nos itens anteriores relativos ao número de citações depositadas no *FactorXcancer*, correspondentes a diversos fatores, podem ser aplicadas estratégias estatísticas para retirar evidências sobre o impacto de determinados factores em determinados cancros. Podem ser utilizados para o efeito softwares informáticos como o *Excel* da Microsoft™ e o SPSS da IBM, para apreciação da significância da informação tratada. A partir da metodologia estatística realizada no programa informático Excel é possível encontrar citações referentes à existência de fatores de risco, sendo que, o cancro oral partilha com os cancros caracterizados como sendo da “cabeça e pescoço” as principais ontologias referentes aos principais fatores de risco: “*Alcohol Beverages*”; “*Smoking habits*”. Segundo o quadro I presente no capítulo Anexos (página 159), em cada 5 pacientes que possuem hábitos tabágicos, 5 desenvolveram cancro oral e 4 desenvolveram cancros tipo da “cabeça e pescoço”. No caso das ontologias referentes aos fatores de risco “*Alcohol Beverages*” apenas 4 desenvolveram ambos os cancros. Como segundo fator com maior preponderância no desenvolvimento de cancro oral e dos cancros caracterizados com o termo “cabeça e pescoço” encontram-se as “*Genetic Diseases*”. A partir do quadro II, capítulo anexos (página 160), pode verificar-se que em cada 5 pacientes que padecem de uma doença genética anteriormente identificada, 4 desenvolveram cancro oral e 3 desenvolveram cancros tipo da “cabeça e pescoço”. Das três classes ontológicas escolhidas anteriormente para exemplo de análise da informação disponível no *FactorXcancer*, em último lugar situam-se as ontologias referentes ao fator de risco “*Infections*” onde, a partir da análise do quadro III, capítulo Anexos (página 161), se verifica que em 5 pacientes, 3 desenvolveram cancro oral e 2 desenvolveram cancros do tipo “cabeça e pescoço”. O vírus com maior evidência como fator de risco de cancro oral e cancros caracterizados como pertencendo à “cabeça e pescoço” é o HPV.

A utilização do software SPSS da IBM permitiu relacionar variáveis/ontologias/fatores de forma a identificar as ontologias que correspondem a fatores para os quais estão depositadas um maior número de citações ou seja os fatores com maior preponderância na carcinogénese do cancro oral e dos tipos de cancros da “cabeça e pescoço”. Com esta análise da informação inserida na estrutura da base de dados *FactorXcancer*, é possível inferir a existência de fatores de risco (como

anteriormente descrito) e fatores protetores com maior impacto no cancro oral e dos cancros caracterizados como sendo da “cabeça e pescoço”. A título de exemplo as ontologias a que correspondem aos fatores de risco numa ordem decrescente em termos de impacto são: “*Smoking Tobacco Heavy*”; “*Smokeless*”; “*Alcohol Habits*”; “*Viral Infection HPV-16*”. Como principais fatores protetores em cancro oral foram identificados as ontologias correspondentes aos fatores depositados no *FactorXcancer* com maior número de citações, na classe “*Beverages*” como representado pelo gráfico 13, página 124. As ontologias “*Vegetables*”, “*Fruits*”, micronutrientes como o “*Calcium*” e o “*Retinol*”; “*Carotenoids*”, a “*Vit. B12*”, “*Vit. A*”, “*Folato*” e “*Vit. C*”. Já dos cancros caracterizados como sendo da “cabeça e pescoço”, as principais ontologias que correspondem aos principais fatores protetores são os “*Vegetables*”; “*Fruits*”, e micronutrientes: “*Folato*”; “*Vit. B12*”, *C e E*”, como se verifica pelo gráfico 13.

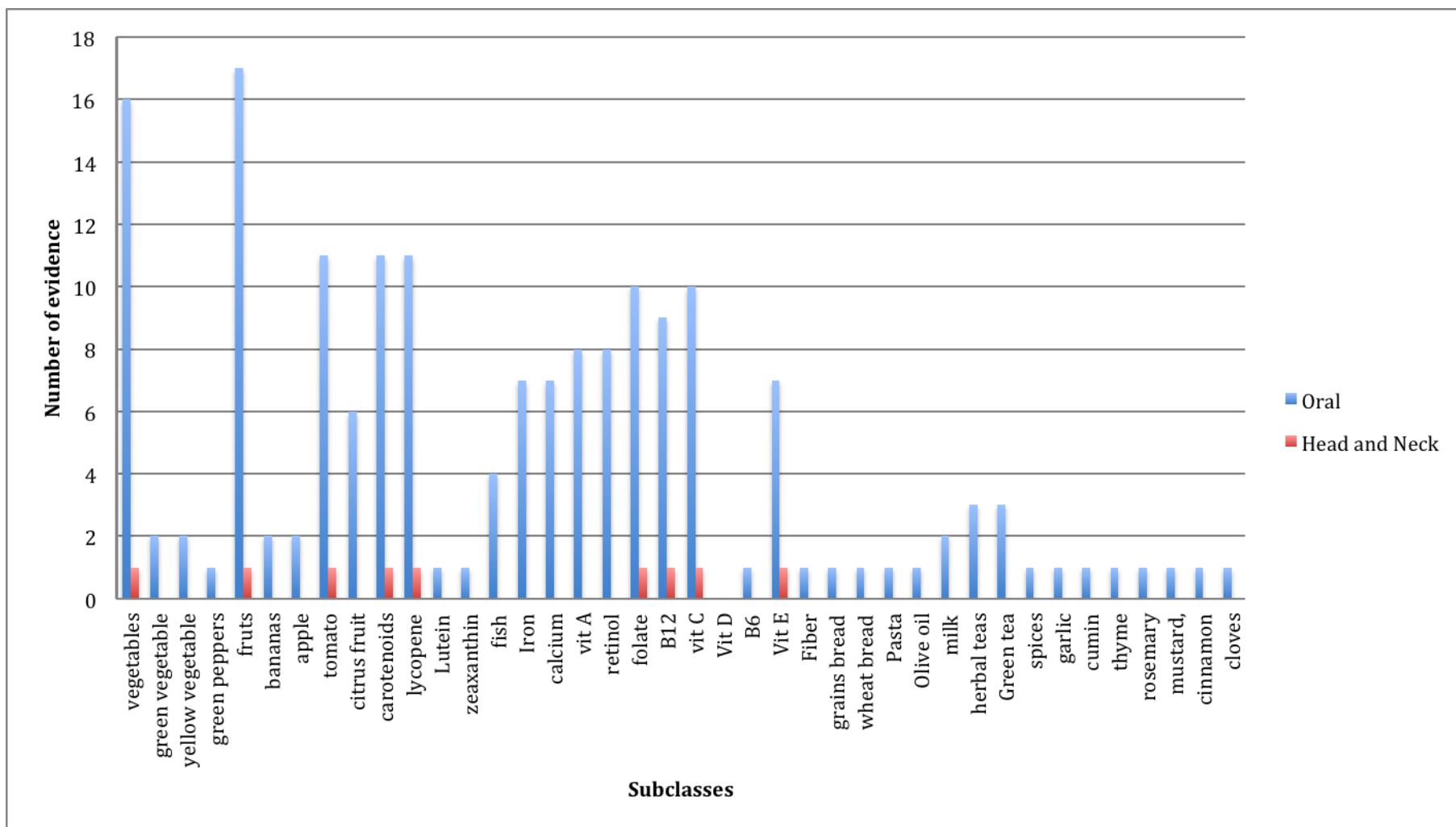


Gráfico 12 - Comparação entre o número de citações anotadas no *FactorXcancer*, referente à Subclasse ontológicas “Beverage”, correspondente às ontologias dos fatores protetores que integram a Superclasse “Diet Factors” entre o cancro oral e câncros do tipo “cabeça e pescoço”.

Conclusão

6. CONCLUSÃO

A necessidade da criação de uma terminologia abrangente referente a fatores associados a cancro, inspirou-nos na criação de um instrumento o *OntoXcancer* que reflete uma hierarquia de princípios ontológicos. Tivemos como objetivo primordial a organização, uniformização e reutilização de terminologias utilizadas no meio científico. As terminologias sobre ontologias que correspondem a fatores associados a cancro foram catalogadas a partir da literatura consultada nas bibliotecas “*Medical Subject headings, Robert Hoehndorf version of Mesh, HUGO Gene Nomenclature Committe e NCI Thersaurus*”.

OntoXcancer é uma ontologia, que pode continuar a ser incrementada por forma a poder integrar mais termos correspondentes a outros factores com aimplicação noutras patologias. A ontologia criada dedicada a factores associados a cancro foi integrada numa base de dados a *FactorXcancer*.

Após a anotação manual da informação recolhida pela bibliografia consultada referente à janela temporal e requisitos associados defenidos no Material e Métodos, apresentou-se uma metodologia demonstrativa do modo como pode ser integrada a informação e extraídas evidências com implicação na interpretação do impacto de factores nos diversos tipos de cancro. Com a metodologia estabelecida demonstrou-se que é possível relacionar a informação anotada na base de dados por forma a identificar factores de risco, factores protetores de cancro e relações de sinergia entre factores de risco causadores de carcinogénese.

Tendo por base os resultados obtidos, acreditamos que com um maior volume de citações, muito mais conclusões pertinentes serão de se esperar.

Com a criação desta ferramenta bioinformática atualizada, aberta e compartilhada com toda a comunidade científica, acreditamos que a extração e análise de informação está facilitada, sendo indiscutivelmente uma mais-valia na criação de novos caminhos de investigação em diagnóstico em saúde.

Pretendemos que a estruturação precisa e a conversão entre terminologias e ontologias permitam, através da metodologia descrita, atualizações futuras das ferramentas criadas.

Com a organização estabelecida nas ontologias referentes a factores com impacto em saúde é possível a sua integração na ferramenta *OralCard*, como contributo para que, no futuro, seja possível relacionar os fatores de risco com os dados de proteómica depositados nesta ferramenta, visando o esclarecimento de mecanismos moleculares em diversas patologias e em carcinogénese oral em particular⁽²⁶⁴⁾.

Desenvolvimentos Futuros

7. DESENVOLVIMENTOS FUTUROS

As citações obtidas com a realização deste projeto indicam que ainda há muito para desenvolver, nomeadamente na identificação de mais ontologias correspondentes a fatores associados no processo de malignização. Neste sentido, e de acordo com os resultados obtidos, o trabalho futuro deve ser direcionado na continuação da integração nestas ferramentas bioinformáticas, mas de uma forma automática, para a obtenção de mais citações, de forma a conseguir acompanhar o ritmo vertiginoso com que o conhecimento científico é publicado.

Propõe-se a adaptação desta metodologia de trabalho para o desenvolvimento de um estudo que permita apreciar o significado clínico destes possíveis fatores de risco causais de cancro, assim como compreender os mecanismos carcinogénicos específicos de cada fator nos vários tipos de cancro. Com os resultados de um estudo desta natureza será possível determinar quais os agentes causais/risco potenciadores dos mecanismos fisiopatológicos associados ao desenvolvimento de cancro e outras patologias na população portuguesa. O estudo constituirá um contributo para o diagnóstico e tratamento da patologia e criação de novas formas de prevenção.

Outro projeto seria alargar a base de dados *FactorXcancer* a outras patologias e integrá-la noutras redes de conhecimento biomédico.

Referências Bibliográficas

8. REFERÊNCIAS BIBLIOGRÁFICAS

1. Mauricio B. Almeida, Marcello P. Bax. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT [Internet]. 2003;32(3). Available from: <http://dx.doi.org/10.1590/S0100-19652003000300002>
2. Kusnierczyk W. Augmenting Bioinformatics Research with Biomedical Ontologies. [Internet] [Doctoral thesis, monograph (Other scientific)]. [Norwegian]: Norwegian University of Science and Technology; 2008. Available from: <http://urn.kb.se/resolve?urn=urn:nbn:no:ntnu:diva-2001>
3. Berners Lee T, Lassila O. The Semantic Web. *Scientific American*. 2001 May;(284):34–43.
4. F. Noy N, L. McGuinness D. Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report [Internet]. 2001 Mar; Available from: <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>
5. Leite J, Breitman K. *Ontologias - Como e porquê criá-las*. Salvador, Brasil; 2008.
6. Ceusters W, Smith B, Goldberg L. A terminological and ontological analysis of the NCI Thesaurus. *Methods Inf Med*. 2005;44(4):498–507.
7. Chicurel M. Bioinformatics: bringing it all together. *Nature* [Internet]. 2002;419(6908). Available from: <http://dx.doi.org/10.1038/419751a>
8. Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. *AMIA Annu Symp Proc*. 2003;609–13.
9. Smith B, Rosse C. The role of foundational relations in the alignment of biomedical ontologies. *Stud Health Technol Inform*. 2004;107(Pt 1):444–8.
10. Ceusters W, Smith B, Kumar A, Dhaen C. Ontology-based error detection in SNOMED-CT. *Stud Health Technol Inform*. 2004;107(Pt 1):482–6.
11. Ceusters W, Smith B, Kumar A, Dhaen C. Mistakes in medical ontologies: where do they come from and how can they be detected? *Stud Health Technol Inform*. 2004;102:145–63.
12. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, et al. Advancing translational research with the Semantic Web. *BMC Bioinformatics*. 2007;8 Suppl 3:S2.

13. Romero M. Una Aproximación Multi-criterio para la Selección Automática de Ontologías Biomédicas Utilizando Conocimiento Colectivo. [Coruña]: UNIVERSIDADE DA CORUÑA; 2010.
14. Juristo N, Gómez-Pérez A, Prazos J. Evaluation and assessment of Knowledge sharing technology. Towards Very Large Knowledge Bases. IOS Press; 1995. p. 289–96.
15. ALFONSE M, AREF M, SALEM A. An Ontology-Based Cancer Diseases Diagnostic Methodology. Available from: <http://www.wseas.us/e-library/conferences/2013/Dubrovnik/ECC/ECC-14.pdf>
16. Scully C. Oral cancer aetiopathogenesis; past, present and future aspects. *Med Oral Patol Oral Cir Bucal*. 2011 May;16(3):e306–311.
17. Stein CJ, Colditz GA. Modifiable risk factors for cancer. *Br. J. Cancer*. 2004 Jan 26;90(2):299–303.
18. Pontén J. Precancer biology: importance and possible prevention. Introduction. *Cancer Surv*. 1998;32:1–3.
19. Meurman JH, Scully C. Oral cancer: comprehending the condition, causes, controversies, control and consequences. 3. Other risk factors. *Dent Update*. 2011 Feb;38(1):66–8.
20. Ram H, Sarkar J, Kumar H, Konwar R, Bhatt MLB, Mohammad S. Oral cancer: risk factors and molecular pathogenesis. *J Maxillofac Oral Surg*. 2011 Jun;10(2):132–7.
21. Liga Portuguesa contra o cancro. [Web page] Portugal, 2009 [updated 2013; cited.
22. Wang Z, Feng X, Liu X, Jiang L, Zeng X, Ji N, et al. Involvement of potential pathways in malignant transformation from oral leukoplakia to oral squamous cell carcinoma revealed by proteomic analysis. *BMC Genomics*. 2009;10:383.
23. Kim J, Shin DM, El-Naggar A, Lee JS, Corrales C, Lippman SM, et al. Chromosome polysomy and histological characteristics in oral premalignant lesions. *Cancer Epidemiol. Biomarkers Prev*. 2001 Apr;10(4):319–25.
24. David AS, Hess LM. *Fundamentals of Cancer Prevention*. Second ed. Springer; 2008.
25. Amin W, Kang HP, Egloff AM, Singh H, Trent K, Ridge-Hetrick J, et al. An informatics supported web-based data annotation and query tool to expedite translational research for head and neck malignancies. *BMC Cancer*. 2009;9:396.
26. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat. Rev. Cancer*. 2007 Jul;7(7):545–53.

27. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief. Bioinformatics*. 2008 Jan;9(1):75–90.
28. Hey T, Tansley, Stewart, Tolle K. *The fourth paradigm data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research; 2009.
29. Woolf SH. The meaning of translational research and why it matters. *JAMA*. 2008 Jan 9;299(2):211–3.
30. Albani S, Prakken B. The advancement of translational medicine—from regional challenges to global solutions. *Nat. Med.* 2009 Sep;15(9):1006–9.
31. Webb CP, Pass HI. Translation research: from accurate diagnosis to appropriate treatment. *J Transl Med*. 2004 21;2(1):35.
32. Davidson SB, Overton C, Buneman P. Challenges in integrating biological data sources. *J. Comput. Biol.* 1995;2(4):557–72.
33. Machado CM, Rebholz-Schuhmann D, Freitas AT, Couto FM. The semantic web in translational medicine: current applications and future directions. *Brief. Bioinformatics*. 2013 Nov 6;
34. Louie B, Mork P, Martin-Sanchez F, Halevy A, Tarczy-Hornoch P. Data integration and genomic medicine. *J Biomed Inform.* 2007 Feb;40(1):5–16.
35. Signore O. Representing Knowledge in the semantic web. W3C Office in Italy at CNR.
36. Custódio M. Sistema de informação e extração de conhecimento para análise e gestão de informação médica. [Internet]. [Lisboa]: Instituto Superior Técnico - Universidade Técnica de Lisboa; 2012. Available from: https://dspace.ist.utl.pt/bitstream/2295/1285898/1/Dissertacao_50136.pdf
37. Yu A. Methods in biomedical ontology. *Journal of biomedical informatics*. 2006;39(3):252–66.
38. Nunes M. Biomarcadores para o processo de malignização de lesões pré-malignas da cavidade oral. [Internet]. [Viseu]: Universidade Católica Portuguesa; 2012. Available from: <repositorio.ucp.pt/bitstream/10400.14/10267/1/finalissima.pdf>
39. Gruber TR. A translation approach to portable ontology specifications. *Knowledge Acquisition* [Internet]. 1993;5. Available from: <http://dx.doi.org/10.1006/knac.1993.1008>
40. Lambrix P, Habbouche M, Pérez M. Evaluation of ontology development tools for bioinformatics. *Bioinformatics*. 2003 Aug 12;19(12):1564–71.
41. Fonseca FT, Egenhofer MJ, et al. Using Ontologies for Integrated Geographic Information Systems. *Transaction in GIS*. 2002 Jun;6(3):231–57.

42. Nieto M. An Overview of Ontologies. Interactive and Cooperative Technologies Lab: Universidad De Las Américas Puebla – México; 2003 Mar p. 1–18.
43. Gómez-Pérez, A. 1998. Knowledge Sharing and Reuse In The handbook on Applied Ex.
44. Guarino N. Formal ontology and information systems. Trento, Italy; 1998. p. 3–15.
45. Martínez-Romero M, Vázquez-Naya JM, Rabuñal JR, Pita-Fernández S, Macenlle R, Castro-Alvariño J, et al. Artificial intelligence techniques for colorectal cancer drug metabolism: ontology and complex network. *Curr. Drug Metab.* 2010 May;11(4):347–68.
46. Blake J. Bio-ontologies-fast and furious. *Nat. Biotechnol.* 2004 Jun;22(6):773–4.
47. Noy NF, Tudorache T, de Coronado S, Musen MA. Developing biomedical ontologies collaboratively. *AMIA Annu Symp Proc.* 2008;520–4.
48. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res.* 2001 Aug;11(8):1425–33.
49. Sioutos N, de Coronado S, Haber MW, Hartel FW, Shaiu W-L, Wright LW. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform.* 2007 Feb;40(1):30–43.
50. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA.* 1994 Apr 13;271(14):1103–8.
51. Bodenreider O. Using UMLS semantics for classification purposes. *Proc AMIA Symp.* 2000;86–90.
52. Hersh W, Hickam DH, Haynes RB, McKibbin KA. Evaluation of SAPHIRE: an automated approach to indexing and retrieving medical literature. *Proc Annu Symp Comput Appl Med Care.* 1991;808–12.
53. Suomela BP, Andrade MA. Ranking the whole MEDLINE database according to a large training set using text indexing. *BMC Bioinformatics.* 2005;6:75.
54. F. Noy N, et all. Creating Semantic Web Contents with Protégé-2000. *Intelligent Systems, IEEE.* 2001;16(2):60–71.
55. Staab S, et al. *Handbook on Ontologies.* 2nd ed. Springer; 2009.
56. Côté RG, Jones P, Apweiler R, Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics.* 2006;7:97.

57. Côté RG, Jones P, Martens L, Apweiler R, Hermjakob H. The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res.* 2008 Jul 1;36(Web Server issue):W372–376.
58. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.* 2009 Jul;37(Web Server issue):W170–173.
59. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, et al. Bioportal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* 2011 Jul;39(Web Server issue):W541–545.
60. Döbrossy L. Epidemiology of head and neck cancer: magnitude of the problem. *Cancer Metastasis Rev.* 2005 Jan;24(1):9–17.
61. International Classification of diseases for oncology (ICD-O), second ed. Geneva.
62. Burt BA. Concepts of risk in dental public health. *Community Dent Oral Epidemiol.* 2005 Aug;33(4):240–7.
63. Meurman JH, Scully C. Oral cancer: comprehending the condition, causes, controversies, control and consequences. 3. Other risk factors. *Dent Update.* 2011 Feb;38(1):66–8.
64. Argiris A, Eng C. Epidemiology, staging, and screening of head and neck cancer. *Cancer Treat. Res.* 2003;114:15–60.
65. D’Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N. Engl. J. Med.* 2007 May 10;356(19):1944–56.
66. Vineis P, Alavanja M, Buffler P, Fontham E, Franceschi S, Gao YT, et al. Tobacco and cancer: recent epidemiological evidence. *J. Natl. Cancer Inst.* 2004 Jan 21;96(2):99–106.
67. Ko YC, Huang YL, Lee CH, Chen MJ, Lin LM, Tsai CC. Betel quid chewing, cigarette smoking and alcohol consumption related to oral cancer in Taiwan. *J. Oral Pathol. Med.* 1995 Nov;24(10):450–3.
68. Enwonwu CO, Meeks VI. Bionutrition and oral cancer in humans. *Crit. Rev. Oral Biol. Med.* 1995;6(1):5–17.
69. Park JY, Muscat JE, Ren Q, Schantz SP, Harwick RD, Stern JC, et al. CYP1A1 and GSTM1 polymorphisms and oral cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 1997 Oct;6(10):791–7.
70. La Vecchia C, Tavani A, Franceschi S, Levi F, Corrao G, Negri E. Epidemiology and prevention of oral cancer. *Oral Oncol.* 1997 Sep;33(5):302–12.

71. Sato M, Sato T, Izumo T, Amagasa T. Genetic polymorphism of drug-metabolizing enzymes and susceptibility to oral cancer. *Carcinogenesis*. 1999 Oct;20(10):1927–31.
72. Katoh T, Kaneko S, Kohshi K, Munaka M, Kitagawa K, Kunugita N, et al. Genetic polymorphisms of tobacco- and alcohol-related metabolizing enzymes and oral cavity cancer. *Int. J. Cancer*. 1999 Nov 26;83(5):606–9.
73. Cloos J, Spitz MR, Schantz SP, Hsu TC, Zhang ZF, Tobi H, et al. Genetic susceptibility to head and neck squamous cell carcinoma. *J. Natl. Cancer Inst.* 1996 Apr 17;88(8):530–5.
74. Trizna Z, Clayman GL, Spitz MR, Briggs KL, Goepfert H. Glutathione s-transferase genotypes as risk factors for head and neck cancer. *Am. J. Surg.* 1995 Nov;170(5):499–501.
75. Oude Ophuis MB, van Lieshout EM, Roelofs HM, Peters WH, Manni JJ. Glutathione S-transferase M1 and T1 and cytochrome P4501A1 polymorphisms in relation to the risk for benign and malignant head and neck lesions. *Cancer*. 1998 Mar 1;82(5):936–43.
76. McKaig RG, Baric RS, Olshan AF. Human papillomavirus and head and neck cancer: epidemiology and molecular biology. *Head Neck*. 1998 May;20(3):250–65.
77. Schwartz SM, Daling JR, Doody DR, Wipf GC, Carter JJ, Madeleine MM, et al. Oral cancer risk in relation to sexual history and evidence of human papillomavirus infection. *J. Natl. Cancer Inst.* 1998 Nov 4;90(21):1626–36.
78. Schildt EB, Eriksson M, Hardell L, Magnuson A. Oral snuff, smoking habits and alcohol consumption in relation to oral cancer in a Swedish case-control study. *Int. J. Cancer*. 1998 Jul 29;77(3):341–6.
79. Lazarus P, Sheikh SN, Ren Q, Schantz SP, Stern JC, Richie JP Jr, et al. p53, but not p16 mutations in oral squamous cell carcinomas are associated with specific CYP1A1 and GSTM1 polymorphic genotypes and patient tobacco use. *Carcinogenesis*. 1998 Mar;19(3):509–14.
80. Sturgis EM, Castillo EJ, Li L, Zheng R, Eicher SA, Clayman GL, et al. Polymorphisms of DNA repair gene XRCC1 in squamous cell carcinoma of the head and neck. *Carcinogenesis*. 1999 Nov;20(11):2125–9.
81. Larsson LG, Sandström A, Westling P. Relationship of Plummer-Vinson disease to cancer of the upper alimentary tract in Sweden. *Cancer Res*. 1975 Nov;35(11 Pt. 2):3308–16.
82. Fioretti F, Bosetti C, Tavani A, Franceschi S, La Vecchia C. Risk factors for oral and pharyngeal cancer in never smokers. *Oral Oncol*. 1999 Jul;35(4):375–8.
83. Zheng W, Blot WJ, Diamond EL, Norkus EP, Spate V, Morris JS, et al. Serum micronutrients and the subsequent risk of oral and pharyngeal cancer. *Cancer Res*. 1993 Feb 15;53(4):795–8.

84. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res.* 1988 Jun 1;48(11):3282-7.
85. Jefferies S, Eeles R, Goldgar D, A'Hern R, Henk JM, Gore M. The role of genetic factors in predisposition to squamous cell cancer of the head and neck. *Br. J. Cancer.* 1999 Feb;79(5-6):865-7.
86. Gold EB, Gordis L, Diener MD, Seltser R, Boitnott JK, Bynum TE, et al. Diet and other risk factors for cancer of the pancreas. *Cancer.* 1985 Jan 15;55(2):460-7.
87. Wu AH, Paganini-Hill A, Ross RK, Henderson BE. Alcohol, physical activity and other risk factors for colorectal cancer: a prospective study. *Br. J. Cancer.* 1987 Jun;55(6):687-94.
88. Bostick RM, Potter JD, Kushi LH, Sellers TA, Steinmetz KA, McKenzie DR, et al. Sugar, meat, and fat intake, and non-dietary risk factors for colon cancer incidence in Iowa women (United States). *Cancer Causes Control.* 1994 Jan;5(1):38-52.
89. Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. *Lancet.* 1998 May 9;351(9113):1393-6.
90. Ursin G, Pike MC, Preston-Martin S, d'Abalain G 3rd, Peters RK. Sexual, reproductive, and other risk factors for adenocarcinoma of the cervix: results from a population-based case-control study (California, United States). *Cancer Causes Control.* 1996 May;7(3):391-401.
91. Chichareon S, Herrero R, Muñoz N, Bosch FX, Jacobs MV, Deacon J, et al. Risk factors for cervical cancer in Thailand: a case-control study. *J. Natl. Cancer Inst.* 1998 Jan 7;90(1):50-7.
92. Chaouki N, Bosch FX, Muñoz N, Meijer CJ, El Gueddari B, El Ghazi A, et al. The viral origin of cervical cancer in Rabat, Morocco. *Int. J. Cancer.* 1998 Feb 9;75(4):546-54.
93. Hatanaka H, Shimomura K. [Immunosuppressive activity of a microbial product]. *Nippon Yakurigaku Zasshi.* 1997 Oct;110 Suppl 1:69P-74P.
94. Homann N, Tillonen J, Meurman JH, Rintamäki H, Lindqvist C, Rautio M, et al. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. *Carcinogenesis.* 2000 Apr;21(4):663-8.
95. Talamini R, Vaccarella S, Barbone F, Tavani A, La Vecchia C, Herrero R, et al. Oral hygiene, dentition, sexual habits and risk of oral cancer. *Br. J. Cancer.* 2000 Nov;83(9):1238-42.

96. Sato M, Sato T, Izumo T, Amagasa T. Genetically high susceptibility to oral squamous cell carcinoma in terms of combined genotyping of CYP1A1 and GSTM1 genes. *Oral Oncol.* 2000 May;36(3):267–71.
97. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ.* 2000 Sep 9;321(7261):624–8.
98. Mori M, Saitoh S, Takagi S, Obara F, Ohnishi H, Akasaka H, et al. A Review of Cohort Studies on the Association Between History of Diabetes Mellitus and Occurrence of Cancer. *Asian Pac. J. Cancer Prev.* 2000;1(4):269–76.
99. Wong DT, Münger K. Association of human papillomaviruses with a subgroup of head and neck squamous cell carcinomas. *J. Natl. Cancer Inst.* 2000 May 3;92(9):675–7.
100. Kuper H, Adami HO, Trichopoulos D. Infections as a major preventable cause of human cancer. *J. Intern. Med.* 2000 Sep;248(3):171–83.
101. Butel JS. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. *Carcinogenesis.* 2000 Mar;21(3):405–26.
102. Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJ, van den Brandt PA. Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study: The Netherlands Cohort Study on Diet and Cancer. *Am. J. Epidemiol.* 2000 Dec 1;152(11):1081–92.
103. Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J. Natl. Cancer Inst.* 2000 May 3;92(9):709–20.
104. McWilliams JE, Evans AJ, Beer TM, Andersen PE, Cohen JI, Everts EC, et al. Genetic polymorphisms in head and neck cancer risk. *Head Neck.* 2000 Sep;22(6):609–17.
105. Kooiman GG, Martin FL, Williams JA, Grover PL, Phillips DH, Muir GH. The influence of dietary and environmental factors on prostate cancer risk. *Prostate Cancer Prostatic Dis.* 2000 Dec;3(4):256–8.
106. Zain RB. Cultural and dietary risk factors of oral cancer and precancer--a brief overview. *Oral Oncol.* 2001 Apr;37(3):205–10.
107. Garrote LF, Herrero R, Reyes RM, Vaccarella S, Anta JL, Ferbeye L, et al. Risk factors for cancer of the oral cavity and oro-pharynx in Cuba. *Br. J. Cancer.* 2001 Jul 6;85(1):46–54.
108. Sreelekha TT, Ramadas K, Pandey M, Thomas G, Nalinakumari KR, Pillai MR. Genetic polymorphism of CYP1A1, GSTM1 and GSTT1 genes in Indian oral cancer. *Oral Oncol.* 2001 Oct;37(7):593–8.
109. Tanabe H, Yokota K, Shibata N, Satoh T, Watari J, Kohgo Y. Alcohol consumption as a major risk factor in the development of early esophageal

cancer in patients with head and neck cancer. *Intern. Med.* 2001 Aug;40(8):692–6.

110. Phukan RK, Ali MS, Chetia CK, Mahanta J. Betel nut and tobacco chewing; potential risk factors of cancer of oesophagus in Assam, India. *Br. J. Cancer.* 2001 Sep 1;85(5):661–7.

111. Starr JR, Daling JR, Fitzgibbons ED, Madeleine MM, Ashley R, Galloway DA, et al. Serologic evidence of herpes simplex virus 1 infection and oropharyngeal cancer risk. *Cancer Res.* 2001 Dec 1;61(23):8459–64.

112. van Houten VM, Snijders PJ, van den Brekel MW, Kummer JA, Meijer CJ, van Leeuwen B, et al. Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. *Int. J. Cancer.* 2001 Jul 15;93(2):232–5.

113. Mayne ST, Risch HA, Dubrow R, Chow WH, Gammon MD, Vaughan TL, et al. Nutrient intake and risk of subtypes of esophageal and gastric cancer. *Cancer Epidemiol. Biomarkers Prev.* 2001 Oct;10(10):1055–62.

114. Gallus S, Bosetti C, Franceschi S, Levi F, Simonato L, Negri E, et al. Oesophageal cancer in women: tobacco, alcohol, nutritional and hormonal factors. *Br. J. Cancer.* 2001 Aug 3;85(3):341–5.

115. Wu MT, Lee YC, Chen CJ, Yang PW, Lee CJ, Wu DC, et al. Risk of betel chewing for oesophageal cancer in Taiwan. *Br. J. Cancer.* 2001 Sep 1;85(5):658–60.

116. Ko Y, Abel J, Harth V, Bröde P, Antony C, Donat S, et al. Association of CYP1B1 codon 432 mutant allele in head and neck squamous cell cancer is reflected by somatic mutations of p53 in tumor tissue. *Cancer Res.* 2001 Jun 1;61(11):4398–404.

117. Zheng Y, Shen H, Sturgis EM, Wang LE, Eicher SA, Strom SS, et al. Cyclin D1 polymorphism and risk for squamous cell carcinoma of the head and neck: a case-control study. *Carcinogenesis.* 2001 Aug;22(8):1195–9.

118. Muto M, Nakane M, Hitomi Y, Yoshida S, Sasaki S, Ohtsu A, et al. Association between aldehyde dehydrogenase gene polymorphisms and the phenomenon of field cancerization in patients with head and neck cancer. *Carcinogenesis.* 2002 Oct;23(10):1759–65.

119. Olshan AF, Watson MA, Weissler MC, Bell DA. XRCC1 polymorphisms and head and neck cancer. *Cancer Lett.* 2002 Apr 25;178(2):181–6.

120. Buch SC, Notani PN, Bhisey RA. Polymorphism at GSTM1, GSTM3 and GSTT1 gene loci and susceptibility to oral cancer in an Indian population. *Carcinogenesis.* 2002 May;23(5):803–7.

121. Xing D, Tan W, Wei Q, Lin D. Polymorphisms of the DNA repair gene XPD and risk of lung cancer in a Chinese population. *Lung Cancer.* 2002 Nov;38(2):123–9.

122. Hahn M, Hagedorn G, Kuhlisch E, Schackert HK, Eckelt U. Genetic polymorphisms of drug-metabolizing enzymes and susceptibility to oral cavity cancer. *Oral Oncol.* 2002 Jul;38(5):486–90.
123. Coussens LM, Werb Z. Inflammation and cancer. *Nature.* 2002 Dec 19;420(6917):860–7.
124. Migliore L, Coppedè F. Genetic and environmental factors in cancer and neurodegenerative diseases. *Mutat. Res.* 2002 Dec;512(2-3):135–53.
125. Mayne ST, Navarro SA. Diet, obesity and reflux in the etiology of adenocarcinomas of the esophagus and gastric cardia in humans. *J. Nutr.* 2002 Nov;132(11 Suppl):3467S–3470S.
126. Yokoyama A, Kato H, Yokoyama T, Tsujinaka T, Muto M, Omori T, et al. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and glutathione S-transferase M1 and drinking, smoking, and diet in Japanese men with esophageal squamous cell carcinoma. *Carcinogenesis.* 2002 Nov;23(11):1851–9.
127. Daling JR, Madeleine MM, Schwartz SM, Shera KA, Carter JJ, McKnight B, et al. A population-based study of squamous cell vaginal cancer: HPV and cofactors. *Gynecol. Oncol.* 2002 Feb;84(2):263–70.
128. Mortazavi S, Zali M, Raoufi M, Nadji M, Kowsarian P, Nowroozi A. The Prevalence of Human Papillomavirus in Cervical Cancer in Iran. *Asian Pac. J. Cancer Prev.* 2002;3(1):69–72.
129. Riboli E, Norat T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am. J. Clin. Nutr.* 2003 Sep;78(3 Suppl):559S–569S.
130. Herrero R, Castellsagué X, Pawlita M, Lissowska J, Kee F, Balaram P, et al. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J. Natl. Cancer Inst.* 2003 Dec 3;95(23):1772–83.
131. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer.* 2003 Oct;3(10):733–44.
132. Pelucchi C, Talamini R, Negri E, Levi F, Conti E, Franceschi S, et al. Folate intake and risk of oral and pharyngeal cancer. *Ann. Oncol.* 2003 Nov;14(11):1677–81.
133. Yokoyama A, Omori T. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers. *Jpn. J. Clin. Oncol.* 2003 Mar;33(3):111–21.
134. Stolzenberg-Solomon RZ, Qiao Y-L, Abnet CC, Ratnasinghe DL, Dawsey SM, Dong ZW, et al. Esophageal and gastric cardia cancer risk and folate- and vitamin B(12)-related polymorphisms in Linxian, China. *Cancer Epidemiol. Biomarkers Prev.* 2003 Nov;12(11 Pt 1):1222–6.
135. Rosenberg PS, Greene MH, Alter BP. Cancer incidence in persons with Fanconi anemia. *Blood.* 2003 Feb 1;101(3):822–6.

136. Sewram V, De Stefani E, Brennan P, Boffetta P. Maté consumption and the risk of squamous cell esophageal cancer in Uruguay. *Cancer Epidemiol. Biomarkers Prev.* 2003 Jun;12(6):508–13.
137. Murray L, Watson P, Johnston B, Sloan J, Mainie IML, Gavin A. Risk of adenocarcinoma in Barrett's oesophagus: population based study. *BMJ.* 2003 Sep 6;327(7414):534–5.
138. Stolzenberg-Solomon RZ, Dodd KW, Blaser MJ, Virtamo J, Taylor PR, Albanes D. Tooth loss, pancreatic cancer, and *Helicobacter pylori*. *Am. J. Clin. Nutr.* 2003 Jul;78(1):176–81.
139. Dahlstrom KR, Adler-Storthz K, Etzel CJ, Liu Z, Dillon L, El-Naggar AK, et al. Human papillomavirus type 16 infection and squamous cell carcinoma of the head and neck in never-smokers: a matched pair analysis. *Clin. Cancer Res.* 2003 Jul;9(7):2620–6.
140. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* 2003 Feb 6;348(6):518–27.
141. Toporcov TN, Antunes JLF, Tavares MR. Fat food habitual intake and risk of oral cancer. *Oral Oncol.* 2004 Oct;40(9):925–31.
142. Stein CJ, Colditz GA. Modifiable risk factors for cancer. *Br. J. Cancer.* 2004 Jan 26;90(2):299–303.
143. Ha PK, Califano JA. The role of human papillomavirus in oral carcinogenesis. *Crit. Rev. Oral Biol. Med.* 2004;15(4):188–96.
144. Tae K, Lee HS, Park BJ, Park CW, Kim KR, Cho HY, et al. Association of DNA repair gene XRCC1 polymorphisms with head and neck cancer in Korean population. *Int. J. Cancer.* 2004 Sep 20;111(5):805–8.
145. Chang K-W, Lee T-C, Yeh W-I, Chung M-Y, Liu C-J, Chi L-Y, et al. Polymorphism in heme oxygenase-1 (HO-1) promoter is related to the risk of oral squamous cell carcinoma occurring on male areca chewers. *Br. J. Cancer.* 2004 Oct 18;91(8):1551–5.
146. Drummond SN, De Marco L, Noronha JCM, Gomez RS. GSTM1 polymorphism and oral squamous cell carcinoma. *Oral Oncol.* 2004 Jan;40(1):52–5.
147. Wogan GN, Hecht SS, Felton JS, Conney AH, Loeb LA. Environmental and chemical carcinogenesis. *Semin. Cancer Biol.* 2004 Dec;14(6):473–86.
148. English DR, MacInnis RJ, Hodge AM, Hopper JL, Haydon AM, Giles GG. Red meat, chicken, and fish consumption and risk of colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* 2004 Sep;13(9):1509–14.

149. Pagano JS, Blaser M, Buendia M-A, Damania B, Khalili K, Raab-Traub N, et al. Infectious agents and cancer: criteria for a causal relation. *Semin. Cancer Biol.* 2004 Dec;14(6):453–71.
150. Talbot SJ, Crawford DH. Viruses and tumours--an update. *Eur. J. Cancer.* 2004 Sep;40(13):1998–2005.
151. Abbas A, Delvinquiere K, Lechevrel M, Lebailly P, Gauduchon P, Launoy G, et al. GSTM1, GSTT1, GSTP1 and CYP1A1 genetic polymorphisms and susceptibility to esophageal cancer in a French population: different pattern of squamous cell carcinoma and adenocarcinoma. *World J. Gastroenterol.* 2004 Dec 1;10(23):3389–93.
152. Solaymani-Dodaran M, Logan RFA, West J, Card T, Coupland C. Risk of oesophageal cancer in Barrett's oesophagus and gastro-oesophageal reflux. *Gut.* 2004 Aug;53(8):1070–4.
153. Evans AJ, Henner WD, Eilers KM, Montalto MA, Wersinger EM, Andersen PE, et al. Polymorphisms of GSTT1 and related genes in head and neck cancer risk. *Head Neck.* 2004 Jan;26(1):63–70.
154. Riedel F, Goessler UR, Hormann K. Alcohol-related diseases of the mouth and throat. *Dig Dis.* 2005;23(3-4):195–203.
155. Petti S, Scully C. Oral cancer: the association between nation-based alcohol-drinking profiles and oral cancer mortality. *Oral Oncol.* 2005 Sep;41(8):828–34.
156. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol. Biomarkers Prev.* 2005 Feb;14(2):467–75.
157. Ragin CCR, Taioli E. Survival of squamous cell carcinoma of the head and neck in relation to human papillomavirus infection: review and meta-analysis. *Int. J. Cancer.* 2007 Oct 15;121(8):1813–20.
158. Yang C, Wang H, Wang Z, Du H, Tao D, Mu X, et al. Risk factors for esophageal cancer: a case-control study in South-western China. *Asian Pac. J. Cancer Prev.* 2005 Mar;6(1):48–53.
159. Neumann AS, Lyons HJ, Shen H, Liu Z, Shi Q, Sturgis EM, et al. Methylenetetrahydrofolate reductase polymorphisms and risk of squamous cell carcinoma of the head and neck: a case-control analysis. *Int. J. Cancer.* 2005 May 20;115(1):131–6.
160. Döbrossy L. Epidemiology of head and neck cancer: magnitude of the problem. *Cancer Metastasis Rev.* 2005 Jan;24(1):9–17.
161. Huang W-Y, Olshan AF, Schwartz SM, Berndt SI, Chen C, Llaca V, et al. Selected genetic polymorphisms in MGMT, XRCC1, XPD, and XRCC3 and risk of head and neck cancer: a pooled analysis. *Cancer Epidemiol. Biomarkers Prev.* 2005 Jul;14(7):1747–53.

162. D'Souza G, Sugar E, Ruby W, Gravitt P, Gillison M. Analysis of the effect of DNA purification on detection of human papillomavirus in oral rinse samples by PCR. *J. Clin. Microbiol.* 2005 Nov;43(11):5526–35.
163. Sakaguchi S, Yokokawa Y, Hou J, Zhang X-L, Li X-P, Li S-S, et al. Environmental exposure and p53 mutations in esophageal cancer patients in areas of low and high incidence of esophageal cancer in China. *Tohoku J. Exp. Med.* 2005 Dec;207(4):313–24.
164. Sepehr A, Kamangar F, Fahimi S, Saidi F, Abnet CC, Dawsey SM. Poor oral health as a risk factor for esophageal squamous dysplasia in northeastern Iran. *Anticancer Res.* 2005 Feb;25(1B):543–6.
165. Lu X-M, Zhang Y-M, Lin R-Y, Arzi G, Wang X, Zhang Y-L, et al. Relationship between genetic polymorphisms of metabolizing enzymes CYP2E1, GSTM1 and Kazakh's esophageal squamous cell cancer in Xinjiang, China. *World J. Gastroenterol.* 2005 Jun 28;11(24):3651–4.
166. Farhadi M, Tahmasebi Z, Merat S, Kamangar F, Nasrollahzadeh D, Malekzadeh R. Human papillomavirus in squamous cell carcinoma of esophagus in a high-risk population. *World J. Gastroenterol.* 2005 Feb 28;11(8):1200–3.
167. Slots J, Saygun I, Sabeti M, Kubar A. Epstein-Barr virus in oral diseases. *J. Periodont. Res.* 2006 Aug;41(4):235–44.
168. Pavia M, Pileggi C, Nobile CGA, Angelillo IF. Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. *Am. J. Clin. Nutr.* 2006 May;83(5):1126–34.
169. Purdue MP, Järholm B, Bergdahl IA, Hayes RB, Baris D. Occupational exposures and head and neck cancers among Swedish construction workers. *Scand J Work Environ Health.* 2006 Aug;32(4):270–5.
170. Castro TPPG, Bussoloti Filho I. Prevalence of human papillomavirus (HPV) in oral cavity and oropharynx. *Braz J Otorhinolaryngol.* 2006 Apr;72(2):272–82.
171. Gattás GJF, de Carvalho MB, Siraque MS, Curioni OA, Kohler P, Eluf-Neto J, et al. Genetic polymorphisms of CYP1A1, CYP2E1, GSTM1, and GSTT1 associated with head and neck cancer. *Head Neck.* 2006 Sep;28(9):819–26.
172. Kietthubthew S, Sriplung H, Au WW, Ishida T. Polymorphism in DNA repair genes and oral squamous cell carcinoma in Thailand. *Int J Hyg Environ Health.* 2006 Jan;209(1):21–9.
173. Vairaktaris E, Yapijakis C, Kessler P, Vylliotis A, Ries J, Wiltfang J, et al. Methylenetetrahydrofolate reductase polymorphism and minor increase of risk for oral cancer. *J. Cancer Res. Clin. Oncol.* 2006 Apr;132(4):219–22.
174. Karin M, Lawrence T, Nizet V. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. *Cell.* 2006 Feb 24;124(4):823–35.

175. Fortuny J, Kogevinas M, Garcia-Closas M, Real FX, Tardón A, Garcia-Closas R, et al. Use of analgesics and nonsteroidal anti-inflammatory drugs, genetic predisposition, and bladder cancer risk in Spain. *Cancer Epidemiol. Biomarkers Prev.* 2006 Sep;15(9):1696–702.
176. Nishikawa M, Kojima N, Komori K, Yamamoto T, Fujii T, Sakai Y. Enhanced maintenance and functions of rat hepatocytes induced by combination of on-site oxygenation and coculture with fibroblasts. *J. Biotechnol.* 2008 Jan 20;133(2):253–60.
177. Wu M, Zhao J-K, Hu X-S, Wang P-H, Qin Y, Lu Y-C, et al. Association of smoking, alcohol drinking and dietary factors with esophageal cancer in high- and low-risk areas of Jiangsu Province, China. *World J. Gastroenterol.* 2006 Mar 21;12(11):1686–93.
178. Lins e Horta H de, Guimarães FF, Rocha LOS, Guimarães RES, Valadares ER. Squamous cell carcinoma of the hypopharynx in a young woman with Fanconi's anemia. *Braz J Otorhinolaryngol.* 2006 Dec;72(6):845–8.
179. Taghavi N, Yazdi I. Type of food and risk of oral cancer. *Arch Iran Med.* 2007 Apr;10(2):227–32.
180. Weitkunat R, Sanders E, Lee PN. Meta-analysis of the relation between European and American smokeless tobacco and oral cancer. *BMC Public Health.* 2007;7:334.
181. Tezal M, Sullivan MA, Reid ME, Marshall JR, Hyland A, Loree T, et al. Chronic periodontitis and the risk of tongue cancer. *Arch. Otolaryngol. Head Neck Surg.* 2007 May;133(5):450–4.
182. Pinar T, Akdur R, Tuncbilek A, Altundag K, Cengiz M. The relationship between occupations and head and neck cancers. *J Natl Med Assoc.* 2007 Jan;99(1):64, 68–71.
183. Rautemaa R, Hietanen J, Niissalo S, Pirinen S, Perheentupa J. Oral and oesophageal squamous cell carcinoma--a complication or component of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED, APS-I). *Oral Oncol.* 2007 Jul;43(6):607–13.
184. Campisi G, Panzarella V, Giuliani M, Lajolo C, Di Fede O, Falaschini S, et al. Human papillomavirus: its identity and controversial role in oral oncogenesis, premalignant and malignant lesions (review). *Int. J. Oncol.* 2007 Apr;30(4):813–23.
185. Anantharaman D, Chaubal PM, Kannan S, Bhisey RA, Mahimkar MB. Susceptibility to oral cancer by genetic polymorphisms at CYP1A1, GSTM1 and GSTT1 loci among Indians: tobacco exposure as a risk modulator. *Carcinogenesis.* 2007 Jul;28(7):1455–62.
186. Hiyama T, Yoshihara M, Tanaka S, Chayama K. Genetic polymorphisms and esophageal cancer risk. *Int. J. Cancer.* 2007 Oct 15;121(8):1643–58.

187. Cross AJ, Leitzmann MF, Gail MH, Hollenbeck AR, Schatzkin A, Sinha R. A prospective study of red and processed meat intake in relation to cancer risk. *PLoS Med.* 2007 Dec;4(12):e325.
188. Yang S-J, Wang H-Y, Li X-Q, Du H-Z, Zheng C-J, Chen H-G, et al. Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China. *World J. Gastroenterol.* 2007 Nov 21;13(43):5760–4.
189. Genkinger JM, Koushik A. Meat consumption and cancer risk. *PLoS Med.* 2007 Dec;4(12):e345.
190. D’Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N. Engl. J. Med.* 2007 May 10;356(19):1944–56.
191. Khan FY, El-Hiday AH, Morad NA. Plummer-Vinson syndrome associated with solid-pseudopapillary tumor of the pancreas. *Chin. Med. J.* 2007 Sep 5;120(17):1553–5.
192. Fortuny J, Kogevinas M, Zens MS, Schned A, Andrew AS, Heaney J, et al. Analgesic and anti-inflammatory drug use and risk of bladder cancer: a population based case control study. *BMC Urol.* 2007;7:13.
193. Báez A. Genetic and environmental factors in head and neck cancer genesis. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev.* 2008 Jun;26(2):174–200.
194. Argiris A, Karamouzis MV, Raben D, Ferris RL. Head and neck cancer. *Lancet.* 2008 May 17;371(9625):1695–709.
195. Meyer MS, Joshipura K, Giovannucci E, Michaud DS. A review of the relationship between tooth loss, periodontal disease, and cancer. *Cancer Causes Control.* 2008 Nov;19(9):895–907.
196. Serefoglou Z, Yapijakis C, Nkenke E, Vairaktaris E. Genetic association of cytokine DNA polymorphisms with head and neck cancer. *Oral Oncol.* 2008 Dec;44(12):1093–9.
197. Park S, Bae J, Nam B-H, Yoo K-Y. Aetiology of cancer in Asia. *Asian Pac. J. Cancer Prev.* 2008 Sep;9(3):371–80.
198. Michaud DS, Liu Y, Meyer M, Giovannucci E, Joshipura K. Periodontal disease, tooth loss, and cancer risk in male health professionals: a prospective cohort study. *Lancet Oncol.* 2008 Jun;9(6):550–8.
199. Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan P, Aghcheli K, et al. Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev.* 2008 Nov;17(11):3062–8.
200. Mitri J, Castillo J, Pittas AG. Diabetes and risk of Non-Hodgkin’s lymphoma: a meta-analysis of observational studies. *Diabetes Care.* 2008 Dec;31(12):2391–7.

201. Petersen PE. Oral cancer prevention and control--the approach of the World Health Organization. *Oral Oncol.* 2009 May;45(4-5):454-60.
202. Lucenteforte E, Garavello W, Bosetti C, La Vecchia C. Dietary factors and oral and pharyngeal cancer risk. *Oral Oncol.* 2009 Jun;45(6):461-7.
203. Sandoval M, Font R, Mañós M, Dicenta M, Quintana MJ, Bosch FX, et al. The role of vegetable and fruit consumption and other habits on survival following the diagnosis of oral cancer: a prospective study in Spain. *Int J Oral Maxillofac Surg.* 2009 Jan;38(1):31-9.
204. Petti S. Lifestyle risk factors for oral cancer. *Oral Oncol.* 2009 May;45(4-5):340-50.
205. Herrel NR, Johnson NL, Cameron JE, Leigh J, Hagensee ME. Development and validation of a HPV-32 specific PCR assay. *Virol. J.* 2009;6:90.
206. Loria D, Barrios E, Zanetti R. Cancer and yerba mate consumption: a review of possible associations. *Rev. Panam. Salud Publica.* 2009 Jun;25(6):530-9.
207. Lubin JH, Purdue M, Kelsey K, Zhang Z-F, Winn D, Wei Q, et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. *Am. J. Epidemiol.* 2009 Oct 15;170(8):937-47.
208. Jayalekshmi PA, Gangadharan P, Akiba S, Nair RRR, Tsuji M, Rajan B. Tobacco chewing and female oral cavity cancer risk in Karunagappally cohort, India. *Br. J. Cancer.* 2009 Mar 10;100(5):848-52.
209. Schouten LJ, Hogervorst JGF, Konings EJM, Goldbohm RA, van den Brandt PA. Dietary acrylamide intake and the risk of head-neck and thyroid cancers: results from the Netherlands Cohort Study. *Am. J. Epidemiol.* 2009 Oct 1;170(7):873-84.
210. Cancela M de C, Ramadas K, Fayette J-M, Thomas G, Muwonge R, Chapuis F, et al. Alcohol intake and oral cavity cancer risk among men in a prospective study in Kerala, India. *Community Dent Oral Epidemiol.* 2009 Aug;37(4):342-9.
211. Campisi G, Giovannelli L. Controversies surrounding human papilloma virus infection, head & neck vs oral cancer, implications for prophylaxis and treatment. *Head Neck Oncol.* 2009;1:8.
212. Kubo A, Block G, Quesenberry CP Jr, Buffler P, Corley DA. Effects of dietary fiber, fats, and meat intakes on the risk of Barrett's esophagus. *Nutr Cancer.* 2009;61(5):607-16.
213. De Stefani E, Aune D, Boffetta P, Deneo-Pellegrini H, Ronco AL, Acosta G, et al. Salted meat consumption and the risk of cancer: a multisite case-control study in Uruguay. *Asian Pac. J. Cancer Prev.* 2009;10(5):853-7.
214. Wen S, Moss SF. *Helicobacter pylori* virulence factors in gastric carcinogenesis. *Cancer Lett.* 2009 Sep 8;282(1):1-8.

215. Murphy G, Shu X-O, Gao Y-T, Ji B-T, Cook MB, Yang G, et al. Family cancer history affecting risk of colorectal cancer in a prospective cohort of Chinese women. *Cancer Causes Control*. 2009 Oct;20(8):1517–21.
216. Tezal M, Sullivan MA, Hyland A, Marshall JR, Stoler D, Reid ME, et al. Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. *Cancer Epidemiol. Biomarkers Prev*. 2009 Sep;18(9):2406–12.
217. Meurman JH. Infectious and dietary risk factors of oral cancer. *Oral Oncol*. 2010 Jun;46(6):411–3.
218. Scully C, Petti S. Overview of cancer for the healthcare team: aetiopathogenesis and early diagnosis. *Oral Oncol*. 2010 Jun;46(6):402–6.
219. Meurman JH. Oral microbiota and cancer. *J Oral Microbiol*. 2010;2.
220. Baykul T, Yilmaz HH, Aydin U, Aydin MA, Aksoy M, Yildirim D. Early diagnosis of oral cancer. *J. Int. Med. Res*. 2010 Jun;38(3):737–49.
221. Fitzpatrick SG, Katz J. The association between periodontal disease and cancer: a review of the literature. *J Dent*. 2010 Feb;38(2):83–95.
222. Szymańska K, Levi JE, Menezes A, Wünsch-Filho V, Eluf-Neto J, Koifman S, et al. TP53 and EGFR mutations in combination with lifestyle risk factors in tumours of the upper aerodigestive tract from South America. *Carcinogenesis*. 2010 Jun;31(6):1054–9.
223. Nguyen NP, Chi A, Nguyen LM, Ly BH, Karlsson U, Vinh-Hung V. Human papillomavirus-associated oropharyngeal cancer: a new clinical entity. *QJM*. 2010 Apr;103(4):229–36.
224. Hong AM, Grulich AE, Jones D, Lee CS, Garland SM, Dobbins TA, et al. Squamous cell carcinoma of the oropharynx in Australian males induced by human papillomavirus vaccine targets. *Vaccine*. 2010 Apr 26;28(19):3269–72.
225. Michl P, Pazdera J, Prochazka M, Pink R, Stosova T. Human papillomavirus in the etiology of head and neck carcinomas. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2010 Mar;154(1):9–12.
226. Wroblewski LE, Peek RM Jr, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin. Microbiol. Rev*. 2010 Oct;23(4):713–39.
227. Lacko M, Roelofs HMJ, Te Morsche RHM, Voogd AC, Ophuis MBO, Peters WHM, et al. Genetic polymorphism in the conjugating enzyme UGT1A1 and the risk of head and neck cancer. *Int. J. Cancer*. 2010 Dec 15;127(12):2815–21.
228. Scully C. Oral cancer aetiopathogenesis; past, present and future aspects. *Med Oral Patol Oral Cir Bucal*. 2011 May;16(3):e306–311.

229. Ram H, Sarkar J, Kumar H, Konwar R, Bhatt MLB, Mohammad S. Oral cancer: risk factors and molecular pathogenesis. *J Maxillofac Oral Surg.* 2011 Jun;10(2):132–7.
230. Syrjänen S, Lodi G, von Bültzingslöwen I, Aliko A, Arduino P, Campisi G, et al. Human papillomaviruses in oral carcinoma and oral potentially malignant disorders: a systematic review. *Oral Dis.* 2011 Apr;17 Suppl 1:58–72.
231. Rubenstein LM, Smith EM, Pawlita M, Haugen TH, Hamšíková E, Turek LP. Human papillomavirus serologic follow-up response and relationship to survival in head and neck cancer: a case-comparison study. *Infect. Agents Cancer.* 2011;6:9.
232. Pannone G, Santoro A, Papagerakis S, Lo Muzio L, De Rosa G, Bufo P. The role of human papillomavirus in the pathogenesis of head & neck squamous cell carcinoma: an overview. *Infect. Agents Cancer.* 2011;6:4.
233. D'Souza G, Dempsey A. The role of HPV in head and neck cancer and review of the HPV vaccine. *Prev Med.* 2011 Oct;53 Suppl 1:S5–S11.
234. Gaur P, Mittal M, Mohanti BK, Das SN. Functional genetic variants of TGF- β 1 and risk of tobacco-related oral carcinoma in high-risk Asian Indians. *Oral Oncol.* 2011 Dec;47(12):1117–21.
235. Chen M-K, Yeh K-T, Chiou H-L, Lin C-W, Chung T-T, Yang S-F. CCR2-64I gene polymorphism increase susceptibility to oral cancer. *Oral Oncol.* 2011 Jul;47(7):577–82.
236. Chen P-H, Lee K-W, Chen C-H, Shieh T-Y, Ho P-S, Wang S-J, et al. CYP26B1 is a novel candidate gene for betel quid-related oral squamous cell carcinoma. *Oral Oncol.* 2011 Jul;47(7):594–600.
237. Lambert R, Sauvaget C, de Camargo Cancela M, Sankaranarayanan R. Epidemiology of cancer from the oral cavity and oropharynx. *Eur J Gastroenterol Hepatol.* 2011 Aug;23(8):633–41.
238. Jayalekshmi PA, Gangadharan P, Akiba S, Koriyama C, Nair RRR. Oral cavity cancer risk in relation to tobacco chewing and bidi smoking among men in Karunagappally, Kerala, India: Karunagappally cohort study. *Cancer Sci.* 2011 Feb;102(2):460–7.
239. Masand RP, El-Mofty SK, Ma X-J, Luo Y, Flanagan JJ, Lewis JS Jr. Adenosquamous carcinoma of the head and neck: relationship to human papillomavirus and review of the literature. *Head Neck Pathol.* 2011 Jun;5(2):108–16.
240. Sudhoff HH, Schwarze HP, Winder D, Steinstraesser L, Görner M, Stanley M, et al. Evidence for a causal association for HPV in head and neck cancers. *Eur Arch Otorhinolaryngol.* 2011 Nov;268(11):1541–7.
241. Yu H, Huang Y, Liu Z, Wang L-E, Li G, Sturgis EM, et al. Effects of MDM2 promoter polymorphisms and p53 codon 72 polymorphism on risk and age at

- onset of squamous cell carcinoma of the head and neck. *Mol. Carcinog.* 2011 Sep;50(9):697–706.
242. Gondivkar SM, Parikh RV, Gadbail AR, Solanke V, Chole R, Mankar M, et al. Involvement of viral factors with head and neck cancers. *Oral Oncol.* 2012 Mar;48(3):195–9.
243. Wang W, Han S, Yao Z, Li X, Huang P, Zhang M, et al. A study of epidemiologic and recurrence factors of oral cancer. *J. Oral Maxillofac. Surg.* 2012 Sep;70(9):2205–10.
244. Saman DM. A review of the epidemiology of oral and pharyngeal carcinoma: update. *Head Neck Oncol.* 2012;4:1.
245. Ang KK, Sturgis EM. Human papillomavirus as a marker of the natural history and response to therapy of head and neck squamous cell carcinoma. *Semin Radiat Oncol.* 2012 Apr;22(2):128–42.
246. Zhu C, Ling Y, Dong C, Zhou X, Wang F. The relationship between oral squamous cell carcinoma and human papillomavirus: a meta-analysis of a Chinese population (1994-2011). *PLoS ONE.* 2012;7(5):e36294.
247. Zavras AI, Yoon AJ, Chen M-K, Lin C-W, Yang S-F. Association between polymorphisms of DNA repair gene ERCC5 and oral squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2012 Nov;114(5):624–9.
248. Weng C-J, Chen M-K, Lin C-W, Chung T-T, Yang S-F. Single nucleotide polymorphisms and haplotypes of MMP-14 are associated with the risk and pathological development of oral cancer. *Ann. Surg. Oncol.* 2012 Jul;19 Suppl 3:S319–327.
249. Bektas-Kayhan K, Unur M, Boy-Metin Z, Cakmakoglu B. MCP-1 and CCR2 gene variants in oral squamous cell carcinoma. *Oral Dis.* 2012 Jan;18(1):55–9.
250. Chu Y-H, Tzeng S-L, Lin C-W, Chien M-H, Chen M-K, Yang S-F. Impacts of microRNA gene polymorphisms on the susceptibility of environmental factors leading to carcinogenesis in oral cancer. *PLoS ONE.* 2012;7(6):e39777.
251. Shukla D, Dinesh Kale A, Hallikerimath S, Vivekanandhan S, Venkatakanthaiah Y. Genetic polymorphism of drug metabolizing enzymes (GSTM1 and CYP1A1) as risk factors for oral premalignant lesions and oral cancer. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub.* 2012 Sep;156(3):253–9.
252. Wang W, Zhao Y, Yang J, Lin B, Gu H, Cao X, et al. Cyclin D1 polymorphism and oral cancer: a meta-analysis. *Mol. Biol. Rep.* 2013 Jan;40(1):87–95.
253. Liu Y-T, Lin L-W, Chen C-Y, Wang C-P, Liu H-P, Houg J-Y, et al. Polymorphism of angiotensin I-converting enzyme gene is related to oral cancer and lymph node metastasis in male betel quid chewers. *Oral Oncol.* 2012 Dec;48(12):1257–62.

254. Hernando-Rodriguez M, Rey-Barja N, Marichalar-Mendia X, Rodriguez-Tojo MJ, Acha-Sagredo A, Aguirre-Urizar JM. Role of cytochrome P-450 genetic polymorphisms in oral carcinogenesis. *J. Oral Pathol. Med.* 2012 Jan;41(1):1–8.
255. Galbiatti ALS, Ruiz MT, Rodrigues JO, Raposo LS, Maníglia JV, Pavarino EC, et al. Polymorphisms and haplotypes in methylenetetrahydrofolate reductase gene and head and neck squamous cell carcinoma risk. *Mol. Biol. Rep.* 2012 Jan;39(1):635–43.
256. Liyanage SS, Segelov E, Garland SM, Tabrizi SN, Seale H, Crowe PJ, et al. Role of human papillomaviruses in esophageal squamous cell carcinoma. *Asia Pac J Clin Oncol.* 2013 Mar;9(1):12–28.
257. Zhao Y, Cao X, Zheng Y, Tang J, Cai W, Wang H, et al. Relationship between cervical disease and infection with human papillomavirus types 16 and 18, and herpes simplex virus 1 and 2. *J. Med. Virol.* 2012 Dec;84(12):1920–7.
258. Guggenheim DE, Shah MA. Gastric cancer epidemiology and risk factors. *J Surg Oncol.* 2013 Mar;107(3):230–6.
259. Sun X, Chen W, Chen Z, Wen D, Zhao D, He Y. Population-based case-control study on risk factors for esophageal cancer in five high-risk areas in China. *Asian Pac. J. Cancer Prev.* 2010;11(6):1631–6.
260. Hopkins J, Cescon DW, Tse D, Bradbury P, Xu W, Ma C, et al. Genetic polymorphisms and head and neck cancer outcomes: a review. *Cancer Epidemiol. Biomarkers Prev.* 2008 Mar;17(3):490–9.
261. Sartor MA, Ade A, Wright Z, States D, Omenn GS, Athey B, et al. Metab2MeSH: annotating compounds with medical subject headings. *Bioinformatics.* 2012 May 15;28(10):1408–10.
262. Povey S, McAlpine P, Brufdor E, et al. HGNC HUGO Gene Nomenclature Committee [Internet]. Available from: <http://www.genenames.org/>
263. Goldenberg D, Golz A, Joachims HZ. The beverage maté: a risk factor for cancer of the head and neck. *Head Neck.* 2003 Jul;25(7):595–601.
264. Arrais JP, Rosa N, Melo J, Coelho ED, Amaral D, Correia MJ, et al. OralCard: a bioinformatic tool for the study of oral proteome. *Arch. Oral Biol.* 2013 Jul;58(7):762–72.

9. ANEXOS

Quadro 1 - Resultados do tratamento estatístico referente à classe “ClinicalFactors”.

	1. SOCIAL BEHAVIORAL	1.1. Smoking habits	1.2. Smokeless	1.3. Smoking tobacco	1.4. Smoke on the water	1.5. Bidi	1.6. Cigars a pipe	1.7. betel quid/ areca nut	1.8. opium	1.9. Ex Smokers	1.9.1. Years since quitting > ou = a 10 anos	1.9.2. Years since quitting < que 10 anos	1.10. alcohol habits	
Oral	6,011227255	5,768184325	4,321928095	5,392317423	0	1,807354922	2,584962501	3,584962501	0,584962501	1,584962501	0,584962501		5,321928095	
	85	72	35	54	1	6	11	21	1	3	2	0	57	
	0	0	0	0	0	0	0	0	0	0	0	0	0	
	5,426264755	5,189824559	4,169925001	4,781359714	0	1,807354922	2,584962501	3,459431619			0,584962501		4,857980995	
	-0,584962501	-0,578359766	-0,152003093	-0,610957700	0	0	0	-0,125530882	-0,584962501	-0,584962501			-1	-0,4639471
	52	43	17	35	0	5	8	12	0	3	2	0	45	
0	0	0	0	0	0	0	0	0	0	0	0	0		
Cabeça e pescoço	4,727920455	4,459431619	3,169925000	4,169925000	-1	1,584962500	2,169925000	2,700439711			0,584962500		4,523561956	
	-1,283306801	-1,308752706	-1,152003093	-1,222392421	-1	-0,222392421	-0,415037499	-0,884522783	-1,584962500	-0,584962500		0	-1	-0,798366139

Quadro 2 - Resultado do tratamento estatístico referente à classe “GeneticDisorders”.

	8. GENETIC	8.1. genetic predisposition	8.2. genetic disorders	8.2.1. Ataxia-telangiectasia	8.2.2. Xeroderma Pigmentosum	8.2.3. Li-Fraumeni syndrome	8.2.4. Anemia de Fanconi	8.2.5. Bloom's Syndrome	8.2.6. Syndrome Lynch	8.2.7. Syndrome Lynch I	8.2.8. Syndrome Lynch II	polyendocrinopathy candidiasis ectodermal dystrophy	8.2.10. Polymorphism
	5,55458885		5,44294349			2,58496250	1,80735492	0,58496250			0,58496250		
	2	2,321928095	6	1,321928095	0,584962501	1	2	1	1	0	1	0,584962501	5,129283017
	53	1	53	3	1	5	3	0	1	1	1	2	44
	0	0	0	0	0	0	0	0	0	0	0	0	0
Oral	4,75488750		4,75488750			1,58496250							
	2	0	2	1	0	1	1	-1	0	0	0	0,584962501	4,491853096
			0,68805599				0,80735492	1,58496250			-0,58496250		
	-0,79970135	-2,321928095	4	-0,321928095	-0,584962501	-1	2	1	-1	0	1	0	-0,637429921
	21	1	21	2	2	2	4	1	1	0	1	0	17
	0	0	0	0	0	0	0	0	0	0	0	0	0
Cabeça e pescoço	3,45943161		3,45943161			0,58496250	1,32192809						
	9	0	9	0,584962501	0,584962501	1	5	0	0	-1	0	-1	3,169925001
	-2,09515723		-1,98351187				-0,48542682	-0,58496250			-0,58496250		
	3	-2,321928095	7	-0,736965594	0	-2	7	1	-1	-1	1	-1,584962501	-1,959358016

Quadro 3 - Resultados do tratamento estatístico referente à subclasse “Infections”.

	3. INFECTIONS	3.1. viral	3.1.1. HPV	3.1.2. HSV - 1	3.1.3. HSV - 2	3.1.4. EBV	3.1.5. CMV	3.1.6. HHV- 8	3.1.7. KSHV	3.1.8. HBV	3.1.9. HCV	3.1.10. Adenovirus	3.1.11. HIV	3.2. bacterial	3.2.1. streptococcus viridans	3.2.2. C. albicans
	5,89481776	5,68650052				2,58496250			0,58496250	2,32192809	2,70043971					
	3	7	5,06608919	1	1	1	1	1,321928095	1	5	8	0	2	3,321928095	0	1
	28	26	24	2	1	4	2	2	0	0	1	1	2	2	1	3
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3,85798099	3,75488750				1,32192809	0,58496250					0,58496250				
Oral	5	2	3,64385619	0,584962501	0	5	1	0,584962501	-1	-1	0	0	1	0,584962501	0	1
	-2,0368367	-1,9316130	-1,4221330			-1,2630344	-0,4150374		-1,5849625	-3,3219280	-2,7004397		-1,4150374			
	68	25	01	-0,415037499	-1	06	99	-0,736965594	01	95	18	0	99	-2,736965594	0	0
	10	10	9	0	0	1	1	1	0	0	1	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cabeça e pescoço	2,45943161	2,45943161	2,32192809													
	9	9	5	-1	-1	0	0	0	-1	-1	0	0	-1	-1	-1	-1
	-3,4353861	-3,2270689	-2,7441610			-2,5849625			-1,5849625	-3,3219280	-2,7004397					
	45	09	96	-2	-2	01	-1	-1,321928095	01	95	18	0	-3	-4,321928095	-1	-2

10. Índice de Figuras

Figura 1 - Metodologia geral para construir uma ontologia. Adaptado de Alfonse <i>et al</i> ⁽¹⁶⁾	37
Figura 2 - Arquitetura da Web Semântica. Adaptado de Sigmore ⁽³⁷⁾	40
Figura 3 - Árvore de Porfírio, com as categorias de Aristóteles (em retângulos). As linhas representam relações entre categorias. Adaptado de (AlexanderYu 2006) ⁽³⁶⁾	41
Figura 4 - Algumas linguagens de apoio à gestão de ontologias. Adaptado de María Neto ⁽⁴¹⁾	46
Figura 5 - Página web da biblioteca Bioportal.	48
Figura 6 - Página web da base de dados Pubmed.....	58
Figura 7 - Representação da base de dados, com exemplo de preenchimento da 1 fase: na coluna 2 insere o código <i>pubmed</i> , o ano, e o tipo de cancro.....	66
Figura 8 - Representação de uma secção da ontologia, com a descrição da classe “ <i>Smoking Habits</i> ”.	67
Figura 9 - Representação de uma secção da base de dados, com a descrição da classe “ <i>Smoking Habits</i> ”.	67
Figura 10 - Página web representativa do Protégé 3.5.	68
Figura 11 - Representação das Superclasses hierárquicas <i>do OntoXcancer</i> no programa <i>protégé</i>	71
Figura 12 - Página web da <i>Medical SubjectHeadings</i> (MESH).	72
Figura 13 - Representação da biblioteca de ontologias <i>Robert Hoehndorf version of Mesh</i> , a partir da página Web do <i>Bioportal</i>	73
Figura 14 - Representação das Classes da Superclasse “ <i>Clinical Factors</i> ”.	78
Figura 15 - Excerto do <i>OntoXcancer</i> onde é demonstrado a organização hierárquica da Subclasse “ <i>Beverages</i> ”.	80
Figura 16 - Representação hierárquica das Classes da Superclasse “ <i>Diet Factors</i> ”.	81
Figura 17 - Excerto do <i>OntoXcancer</i> onde é evidenciada a organização hierárquica da Subclasse “ <i>Food</i> ”.	82
Figura 18 – Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da Classe “ <i>Enviroment Factors</i> ”.	84
Figura 19 - Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da Classe “ <i>Chemical Carcinogen</i> ”	85
Figura 20 - Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da Classe “ <i>Chemical Modifier</i> ”, “ <i>Radiation</i> ” e “ <i>Tobacco Associated Carcinogen</i> ”.	86
Figura 21 - Página Web HUGO Gene NomenclatureCommittee.....	87

Figura 22 - Representação da organização hierárquica ontológica das classes pertencentes à superclasse “Genetic” no <i>OntoXcancer</i>	88
Figura 23 - Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da Classe “Gene or Polymorphism”, pertencente à Superclasse “Genetic”.	89
Figura 24 - Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da Classe “Genes by Function” e “Genetic Diseases”, pertencente à Superclasse “Genetic”.	90
Figura 25 - Excerto do <i>OntoXcancer</i> onde é evidenciado a organização hierárquica da classe “Pharmacotherapy”.	91
Figura 26 - Representação da organização hierárquica ontológica da superclasse “Social Factors” no <i>OntoXcancer</i>	93
Figura 27 - Representação da organização hierárquica ontológica da superclasse “Therapy” no <i>OntoXcancer</i>	94
Figura 28 – Excerto representativo da organização hierárquica ontológica referente à classe “Stomatognathic Diseases” no <i>OntoXcancer</i>	95
Figura 29 – Excerto representativo da organização hierárquica <i>Mesh</i> , referente ao grupo “Stomatognathic Diseases”.	95
Figura 30 – Excerto representativo da organização hierárquica ontológica da superclasse “Pharmacotherapy” no <i>OntoXcancer</i>	96
Figura 31 – Excerto representativo da organização hierárquica <i>Mesh</i> , referente ao grupo “Therapeutic agents”.	96
Figura 32 - Representação do método de importação/reutilização de ontologias a partir do programa informático Protégé.	98
Figura 33 - Representação da construção da ontologia de raiz do termo Maté.	100
Figura 34 - Representação de execução de redes semânticas no <i>OntoXcancer</i>	102
Figura 35 - Excerto representativo da rede semântica da ontologia “Anemia de Fanconi”, que se encontra inserida tanto na Superclasse “Genetic” como na Superclasse “Clinical Factors”	109
Figura 36 - Excerto representativo da rede semântica da ontologia “Gastric Neoplasm”, que se encontra inserida tanto na Classe “Digestive system” como na Classe “Neoplasm”, ambas pertencentes à Superclasse “Clinical Factors”.	109
Figura 38 - Excerto representativo da rede semântica da ontologia “Carcinogenic Aromatic amines”, que se encontra inserida tanto na classe “Tobacco Associated Carcinogen” como na subclasse “tobacco”, dado à sua relação.	110
Figura 37 - Excerto representativo da rede semântica da ontologia “Xeroderma Pigmentosum”, que se encontra inserida tanto na superclasse “Genetic” como na	

superclasse "*Clinical Factors*", ambas pertencentes à Superclasse "*Clinical Factors*".
..... 110

11. Índice de Gráficos

- Gráfico 1 - Comparação entre o número de citações depositadas no *FactorXcancer* para os diferentes tipos de cancros que integram a designação “cabeça e pescoço”. 112
- Gráfico 2 - Comparação do número de citações depositadas no *FactorXcancer*, entre o cancro oral e os outros tipos de cancros classificados como pertencendo ao termo “cabeça e pescoço”..... 112
- Gráfico 3 - a) Comparação do número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros da “cabeça e pescoço”, correspondente às diferentes Superclasses.. b) Excerto da rede de ontologias em que se inserem as correspondentes aos fatores associados às Superclasses representadas em a)..... 113
- Gráfico 4 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* referente a cancro oral, faríngeo, nasofaríngeo, orofaríngeo, hipofaríngeo e laríngeo correspondente às Superclasses com maior número de citações depositadas até à data. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às Superclasses representadas em a)..... 114
- Gráfico 5 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivos às classes “Smoking and Alcohol Beverages. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a)..... 115
- Gráfico 6 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivos às subclasses que pertencem à classe “Infections”, b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às subclasses representadas em a)..... 116
- Gráfico 7 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas às classes “Gene or Polymorphism”, “Genes by Function e “Genetic Diseases”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a). 117
- Gráfico 8 - Comparação entre o número de citações depositadas atualmente no *FactorXcancer* entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas à classe “Smoking Habits”. b)Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a)..... 118

Gráfico 9 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusivas à classe “ <i>Alcohol Beverages</i> ”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).....	119
Gráfico 10 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusiva à subclasse “Viral”. b) Excerto da rede de ontologias em que se inserem as ontologias correspondentes aos fatores associados às classes representadas em a).....	120
Gráfico 11 - Comparação entre o número de citações depositadas atualmente no FactorXcancer entre o cancro oral e os cancros classificados com o termo “cabeça e pescoço”, alusiva classe “ <i>Genetic Disorders</i> ”. b) Excerto da rede de ontologias em que se inserem as ontologias	121
Gráfico 13 - Comparação entre o número de citações anotadas no <i>FactorXcancer</i> , referente à Subclasse ontológicas “ <i>Beverage</i> ”, correspondente às ontologias dos fatores protetores que integram a Superclasse “ <i>Diet Factors</i> ” entre o cancro oral e cancros do tipo “cabeça e pescoço”.....	124

12. Índice de Tabelas

Tabela I - Representação da bibliografia consultada referente a fatores associados a cancro. 60

Tabela II - Representação do tipo cancerígeno identificado em cada artigo consultada pelo presente estudo..... 61

13. Índice de Quadros

Quadro 1 - Resultados do tratamento estatístico referente à classe “ <i>ClinicalFactors</i> ”	159
Quadro 2 - Resultado do tratamento estatístico referente à classe “ <i>GeneticDisorders</i> ”	160
Quadro 3 - Resultados do tratamento estatístico referente à subclasse “ <i>Infections</i> ”	161