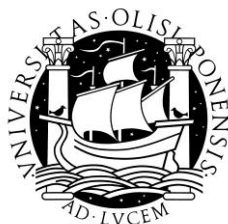


UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



ThermInfo: Sistema de Informação para Coligir e
Apresentar Propriedades Termoquímicas

Ana Isabel Lino Teixeira

Mestrado em Tecnologias de Informação Aplicadas às
Ciências Biológicas e Médicas

2009

UNIVERSIDADE DE LISBOA

Faculdade de Ciências

Departamento de Informática



ThermInfo: Sistema de Informação para Coligir e
Apresentar Propriedades Termoquímicas

Ana Isabel Lino Teixeira

Trabalho orientado pelo Prof. Doutor Francisco José Moreira Couto e co-orientado pelo
Prof. Doutor Rui Jorge Centeno Santos

Mestrado em Tecnologias de Informação Aplicadas às
Ciências Biológicas e Médicas

2009

Resumo

A investigação na área da Termoquímica origina e requer grandes quantidades de dados relativos a diversas propriedades dos compostos químicos. É necessário organizá-los e disponibilizá-los de forma estruturada, simples, coerente, permitindo a cooperação entre a comunidade científica e a evolução ao longo do tempo. Devidos aos inúmeros problemas que advêm da utilização de folhas de cálculo para este efeito, estruturou-se a proposta deste trabalho: a implementação de um Sistema de Informação para coligir e apresentar propriedades estruturais e termoquímicas de compostos orgânicos.

Procurando responder simultaneamente às necessidades de armazenamento, manipulação e visualização dos dados termoquímicos este projecto integra quatro fases de desenvolvimento interdependentes e complementares: percepção (análise do problema e de requisitos), concepção (modelação do sistema), implementação e avaliação do sistema.

O Sistema de Informação foi denominado *ThermInfo* e encontra-se disponível em <http://www.therminfo.com>. Este sistema integra uma base de dados, baseada numa especificação de dados relacional para descrever propriedades estruturais e termoquímicas de compostos orgânicos. A sua interface *Web* de fácil utilização permite a realização de pesquisas, inserção de novos compostos e gestão dos dados.

Presentemente, o *ThermInfo* contém valores de propriedades termoquímicas experimentais criticamente avaliados e termodinamicamente consistentes para cerca de 3000 compostos únicos e não redundantes. Os resultados da avaliação de usabilidade da interface mostram que é fácil de usar e de rápida aprendizagem, o que melhora a eficiência do uso do sistema.

O desenvolvimento deste trabalho sugere algumas direcções futuras que irão ampliar as capacidades do sistema *ThermInfo*, nomeadamente a expansão do conjunto de dados, integrando bases de dados de outras classes de compostos químicos e a incorporação de ferramentas *químio-informáticas* que possibilitem a combinação de uma estrutura desenhada pelo utilizador com os termos de pesquisa e métodos de previsão de propriedades termoquímicas.

Palavras-Chave: Químio-informática, Sistema de Informação, Base de Dados, Interface *Web*, Propriedades Termoquímicas

Abstract

Thermochemical research work often obtains and requires large amounts of data on many kinds of chemical compounds properties. It is necessary to organize and make them available in a structured, simple and consistent way, allowing the cooperation between the scientific community and evolving with time. Due the numerous problems derived from applying spreadsheets to this process, the proposal of this work was structured: the implementation of an Information System to collect and present structural and thermochemical properties of organic compounds.

In order to fulfill the present needs of storage, manipulation and visualization of thermochemical data this project includes four interrelated and complementary phases of development: perception (problem and requirements analysis), design (system modeling), implementation and system evaluation.

The information system was named *ThermInfo* and it is available at <http://www.therminfo.com>. This system integrates a database, based upon a relational data specification to describe structural and thermochemical properties of organic compounds. Its user-friendly Web interface allows a text-based searching, compounds insertion and data management.

Currently, *ThermInfo* contains critically evaluated and thermodynamically consistent experimental thermochemical properties values for about 3000 unique and non redundant compounds. Interface usability results show that it is easy to use and fast to learn which improves the efficiency when employing the system.

The development of this work suggests some future directions that will expand the capabilities of the ThermInfo, including the expansion of the data set, the integration of other classes of compounds databases and chemoinformatics tools that enable a combination of a structure drawn with textual search terms and prediction methods of thermochemical properties.

Keywords: Chemoinformatics, Information System, Database, Web Interface, Thermochemical Properties

Agradecimentos

Embora uma dissertação tenha como característica ser um cunho reflexivo-teórico individual, há contributos de naturezas diversas que não podem deixar de ser realçados. Entre eles a amizade, a compreensão e a colaboração de várias pessoas. Por essa razão, expresso os meus sinceros agradecimentos:

- ao meu orientador Professor Doutor Francisco M. Couto por todo o optimismo com que sempre me apoiou e incentivou durante a realização deste trabalho;
- ao meu co-orientador Doutor Rui C. Santos pela extrema dedicação e organização que sempre demonstrou, por todo o apoio, incentivo, sugestões, respeito e responsabilidade atribuídos ao meu trabalho;
- aos 43 utilizadores, que voluntariamente se disponibilizaram a realizar os testes de usabilidade e que construtivamente criticaram a interface do sistema *ThermInfo*;
- ao departamento de informática e ao LaSIGE pela oportunidade de integrar no grupo XLDB e pelas condições de trabalho oferecidas, assim como à Fundação para a Ciência e Tecnologia pelo suporte financeiro;
- a todos os membros do LaSIGE pela ajuda e companheirismo e espírito de grupo.
- ao Grupo de Energética Molecular do Centro de Química e Bioquímica pela participação activa neste trabalho, simpatia, pronta disponibilidade para esclarecimentos, assim como pelo suporte financeiro.
- a todos os colegas do Mestrado em Tecnologias de Informação Aplicadas às Ciências Biológicas e Médicas pelas discussões e longas conversas, que em muito contribuirão para a realização da parte curricular do mestrado e deste trabalho ;

- à Faculdade de Ciências da Universidade de Lisboa, que é a minha segunda casa desde que terminei o ensino secundário e a todos os professores que contribuíram para a minha formação académica;
- aos meus pais, José Alberto e Maria José, pelo amor, carinho, compreensão e carácter com que sempre me educaram e pelo esforço dedicado para que eu pudesse chegar aqui;
- ao meu irmão, José Miguel, pelo apoio, incentivo, companheirismo, humor, encorajamento e disponibilidade para ajudar
- ao Emir Catak que esteve sempre pacientemente presente apoiando-me e colaborando, tanto nas revisões do meu trabalho, como quando as minhas forças acabavam.
- aos meus amigos, por me acompanharem, incentivarem, divertirem, rirem e chorarem comigo durante todos estes anos.

Lisboa, 29 de Julho de 2009

**Aos meus pais Maria José e José Alberto,
pelo amor, apoio e compreensão.**

Conteúdo

1 Introdução	1
1.1. Enquadramento	1
1.2. Motivação	3
1.3. Objectivos	7
1.4. Metodologia	8
1.5. Organização do documento	9
2 Conceitos Básicos	11
2.1. Compostos Orgânicos	11
2.2. Importância do Conhecimento da Energética dos Compostos Químicos	15
3 Metodologia e Implementação	17
3.1. Análise do Problema	17
3.2. Análise de Requisitos	18
3.3. Base de Dados	22
3.4. Arquitectura do Sistema	25
3.5. Interface	29
3.5.1. Protótipos de Baixa-Fidelidade	29
3.5.2. Implementação e Protótipos de Alta-Fidelidade	34
3.5.3. Avaliação de Usabilidade	60
4 Avaliação	61
4.1. Base de Dados	61
4.2. Usabilidade da Interface	64
5 Conclusões	85
Bibliografia	89

Lista de Abreviaturas e Siglas

CAS	Chemical Abstracts Service
JPEG	Joint Photographic Experts Group
UML	Unified Modeling Language
MySQL	My Structured Query Language
PHP	Hypertext Preprocessor
HTML	Hypertext Markup Language
CSS	Cascading Style Sheets
ID	Identity Descriptor
IUPAC	International Union of Pure and Applied Chemistry
CASRN	Chemical Abstracts Service Registry Number
SMILES	Simplified Molecular Input Line Entry System
ASCII	American Standard Code for Information Interchange
USMILES	Unique Simplified Molecular Input Line Entry System
ASTM	American Society for Testing and Materials
SQL	Structured Language Query
SGBD	Sistema de Gestão de Bases de Dados
Perl	Practical Extraction and Report Language
JS	Java Script
DHTML	Dynamic Hypertext Markup Language
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
IP	Internet Protocol
KLM	Keystroke Level Model
GOMS	Goals, Operators, Methods, Selection
KS	Kolmogorov-Smirnov

Lista de Figuras

Figura 1- Representação da curva de crescimento do número de artigos/ano.	2
Figura 2 – Armazenamento dos dados das características nas folhas de cálculo.	5
Figura 3 - Folhas de cálculo a consultar para obter a informação de um composto..	6
Figura 4 - Metodologia adoptada para o desenvolvimento do sistema <i>ThermInfo</i>	9
Figura 5 - Funcionalidades disponibilizadas pelo sistema <i>ThermInfo</i>	20
Figura 6 - Visão global das funcionalidades do <i>ThermInfo</i> (Use-cases).....	22
Figura 7 - Diagrama de classes UML da base de dados do <i>ThermInfo</i>	23
Figura 8 - Administração das tabelas do sistema <i>ThermInfo</i> (<i>phpMyAdmin</i>).	26
Figura 9 - Utilização do <i>phpMyAdmin</i> para realizar queries ao sistema <i>ThermInfo</i>	26
Figura 10 - Modelo simplificado do funcionamento do sistema <i>ThermInfo</i>	28
Figura 11 - Protótipo de Baixa-Fidelidade para a página de entrada no sistema.	29
Figura 12 - Protótipo de Baixa-Fidelidade para a página de pesquisa de compostos ...	30
Figura 13 - Protótipo de Baixa-Fidelidade para a página de pesquisa avançada.	31
Figura 14 - Protótipo de Baixa-Fidelidade para a listagem de resultados.....	31
Figura 15 - Protótipo de Baixa-Fidelidade para a ficha de um composto.....	32
Figura 16 - Protótipo de Baixa-Fidelidade para a página de inserção de dados.....	33
Figura 17 - Protótipo de Baixa-Fidelidade para a página de actualização/eliminação..	33
Figura 18 – Página Principal do sistema <i>ThermInfo</i>	35
Figura 19 - Página de apresentação da equipa do sistema <i>ThermInfo</i>	36
Figura 20 – Código de segurança utilizado no sistema <i>ThermInfo</i>	37

Figura 21 - Aviso indicador de que o código de segurança não foi digitado.	37
Figura 22 - Mensagem de erro quando o código inserido não está correcto.....	37
Figura 23 - Diagrama de Colaboração do Caso de Uso: Pesquisar um Composto.	38
Figura 24 – Página de Pesquisa Simples	39
Figura 25 - Mensagem de aviso de que o utilizador não inseriu o termo da pesquisa. .	39
Figura 26 - Formulário da ‘Pesquisa Simples’ por nome.....	40
Figura 27 - Formulário da ‘Pesquisa Simples’ por fórmula molecular.....	40
Figura 28 - Mensagem de erro no formato do ID molecular.....	41
Figura 29 - Formulário da ‘Pesquisa Simples’ por ID molecular.	41
Figura 30 - Formulário da ‘Pesquisa Simples’ por CASRN.	42
Figura 31 - Mensagem de erro no formato do CASRN.	43
Figura 32 - Mensagem de erro no dígito controlo do CASRN.....	43
Figura 33 – Formulário da ‘Pesquisa Simples’ por SMILES.....	43
Figura 34 - Parte da listagem de resultados obtida para uma pesquisa.	44
Figura 35 - Ficha de um composto contendo toda a informação disponível.....	45
Figura 36 - Formulário de Pesquisa Estrutural.....	46
Figura 37 – Formulário de registo no sistema <i>ThermInfo</i>	47
Figura 38 - Diagrama de Colaboração do Caso de Uso: Registrar uma conta..	48
Figura 39 - Formulário que permite a entrada no sistema.....	48
Figura 40 - Formulário que permite a inserção de novos compostos orgânicos.	49
Figura 41 - Diagrama de Colaboração do Caso de Uso: Inserir um composto.....	50

Figura 42 – Caixa de pesquisa por ID molecular de um composto a apagar.	52
Figura 43 - Formulário de actualização dos dados dos compostos existentes.	53
Figura 44 – Informação sobre o composto a remover da base de dados.....	54
Figura 45 - Funcionalidade para a monitorização dos utilizadores registados	56
Figura 46 – Estatísticas da base de dados.....	56
Figura 47 – Interface que permite a monitorização do uso das funcionalidades.....	57
Figura 48 - Conteúdos do manual de instruções do sistema <i>ThermInfo</i>	58
Figura 49 – Conteúdos do manual de instruções do sistema <i>ThermInfo</i>	59
Figura 50 – Formulário para o envio de mensagens aos administradores.....	60
Figura 51 - Representação gráfica das estatísticas da Base de Dados.....	62
Figura 52 - Avaliação dos protótipos de Baixa Fidelidade	65
Figura 53 - Verificação do código HTML num sistema de validação do W3C.....	66
Figura 54 - Apresentação de mensagens adequadas para transmitir erros.	67
Figura 55 - Prevenção de erros durante a pesquisa de um composto.....	67
Figura 56 - Avaliação dos protótipos de Alta Fidelidade.....	74
Figura 57 - Representação gráfica, utilizando <i>Box Plots</i> , da variável tempo.....	76
Figura 58 – Distribuição dos utilizadores (%) pelo número de erros cometidos..	81
Figura 59 – Distribuição dos utilizadores (%) pelas pontuações aos atributos de usabilidade.....	82

Lista de Tabelas

Tabela 1 – Desempenho temporal das <i>queries</i> realizadas para pesquisar um composto e inserir dados.....	63
Tabela 2 - Tempo de execução para cada um dos comandos utilizados durante a realização de tarefas segundo o modelo KLM.	68
Tabela 3 - Quadro resumo dos tempos de execução totais para cada uma das tarefas delineadas com base no modelo KLM-GOMS.	72
Tabela 4 - Resultados dos testes de usabilidade.	75
Tabela 5 – Valores de p-value obtido para cada um dos testes de KS realizados..	78
Tabela 6 - Valores de <i>p-value</i> obtidos para os testes de <i>Levene</i> e <i>t-Student</i>	79
Tabela 7 – Valores de tempo para desempenhar uma tarefa esperados (KLM-GOMS) e a média de tempo obtidas em cada um dos grupos de utilizadores.	80

1

Introdução

“While much bioscience is published with the knowledge that machines will be expected to understand at least part of it, almost all chemistry is published purely for humans to read.”

- Murray-Rust

1.1. Enquadramento

Em grande parte, a Química continua a ser uma ciência empírica, desenvolvendo uma crescente quantidade de dados e informação. A informação química quantificável cresce exponencialmente devido ao constante refinamento e optimização das tecnologias experimentais [1]. De acordo com o *Chemical Abstracts Service (CAS)*¹, actualmente existem mais de 35 milhões de compostos químicos conhecidos,

¹ CAS Statistical Summary:
[<http://www.cas.org/ASSETS/836E3804111B49BFA28B95BD1B40CD0F/casstats.pdf>]

aumentando mais de 1 milhão todos os anos. Em termos de literatura, mais de 1 milhão de publicações por ano estão relacionadas com informação química (Figura 1) [2].

Por este motivo, o desenvolvimento das Tecnologias de Informação aplicadas à informação química é essencial e torna possível a manipulação de grandes quantidades de dados, dinamiza os processos de procura, partilha e obtenção de informação e ainda a aplicação de algoritmos, técnicas computacionais e métodos estatísticos aos dados [3-6].

A esta área do conhecimento dá-se o nome de “químio-informática”, tradução do inglês *cheminformatics*².

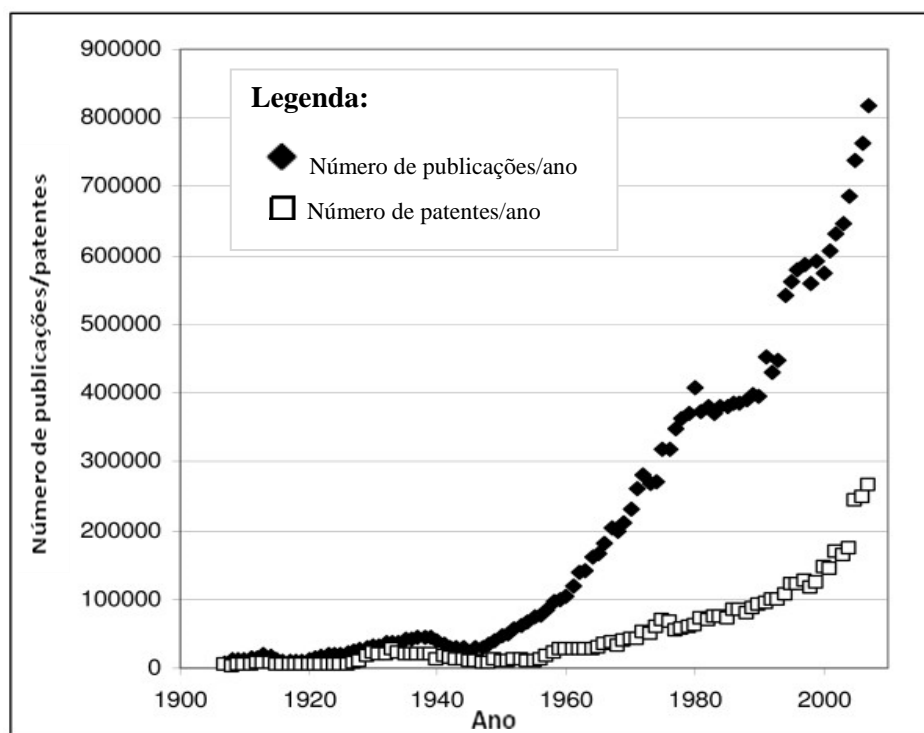


Figura 1 – Representação da curva de crescimento do número de artigos e patentes (registados pelo CAS) relativos à informação química publicada ao longo do tempo (adaptado de [2]).

² Consiste no uso de métodos informáticos para resolver problemas químicos [1,4].

1.2. Motivação

A investigação na área da Química requer, muitas vezes, a pesquisa de dados relativos a diversas propriedades dos compostos químicos. No entanto, a pesquisa dos mesmos em livros e artigos científicos é uma tarefa complexa e morosa. Para colmatar estes problemas é importante o desenvolvimento de bases de dados públicas que permitam compilar e realizar pesquisas rápidas e fáceis. Nos dias de hoje, a maioria das bases de dados termoquímicas existentes são de acesso privado, sendo a sua aquisição cara, não permitindo o desenvolvimento e aplicação de métodos estatísticos e/ou de previsão. Por sua vez, as bases de dados públicas contêm uma quantidade de informação reduzida, maioritariamente não criticamente avaliada e são muitas vezes termodinamicamente inconsistentes, envolvendo grande quantidade de recursos, o que as torna complexas e confusas de utilizar [4-6].

Para resolver estes problemas, o Grupo de Energética Molecular³ do Centro de Química e Bioquímica da Faculdade de Ciências da Universidade de Lisboa, que tem como principal objectivo o estudo das relações entre a estrutura, a energética, a reactividade e a função dos compostos químicos, compilou da literatura e avaliou criticamente os dados termoquímicos experimentais de diversos compostos químicos. Estes dados foram guardados em folhas de cálculo, mas este modo de armazenamento revela vários problemas, nomeadamente:

- **Integridade dos Dados** – para alterar ou apagar dados relativos a um ou a vários compostos em múltiplas linhas e/ou múltiplas folhas de cálculo é necessário que a mesma acção seja repetida várias vezes. Com este procedimento, se algum campo é esquecido, os dados relativos ao composto vão tornar-se ambíguos e perder a integridade.

³ Grupo de Energética Molecular: <http://cqb.fc.ul.pt/menergetics/>

- **Redundância dos Dados** – este tipo de armazenamento é muitas vezes ineficiente, exigindo a repetição dos mesmos dados várias vezes. Assim, o conjunto de dados cresce desnecessariamente, são necessários mais recursos computacionais, o acesso torna-se mais difícil e lento, e a integridade dos dados é mais complicada de manter. Um bom exemplo deste problema é a existência de 28 diferentes características atribuídas aos compostos orgânicos de acordo com os grupos funcionais presentes na estrutura, entre outras características físicas. Tendo em conta que cada composto pode ter mais que uma característica, teríamos três opções de armazenamento:
 - a. múltiplas características na mesma célula;
 - b. cada característica numa coluna;
 - c. cada coluna é uma característica e em cada célula é marcada a sua presença ou ausência.

Neste caso, foi adoptado o último processo de armazenamento, o que obriga à existência de 28 colunas para guardar informação sobre as características de um composto (Figura 2).

- **Validade e não-uniformidade dos dados** – a inserção de dados pelo Homem é falível e variável pelo que, sem restrições e meios de verificação, as folhas de cálculo não nos permitem detectar os erros, diferenças nos formatos e dados iguais escritos de diferentes formas. Isto causa problemas quando queremos, por exemplo, fazer estatísticas sobre os compostos químicos, como seja o número de compostos pertencentes a uma determinada classe ou família. Se as mesmas estiverem escritas de diferentes formas vão ser interpretadas como sendo classes ou famílias diferentes. Este problema também poderá por em causa a unicidade dos dados.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	
310				X	X	X																						
311				X	X	X																						
312				X	X	X																						
313				X	X	X																						
314				X	X	X																						
315													X															
316													X															X
317													X															X
318													X															
319													X															
320													X															
321													X															
322													X															
323													X															
324													X															
325														X														
326														X														
327														X														
328														X														
329														X														
330														X														
331														X														
332														X														
333														X														
334														X														
335														X														
336														X														
337														X														
338														X														
339														X														
340														X														
341														X														
342														X														
343														X														
344														X														
345														X														
346														X														
347														X														
348														X														
349														X														
350		X												X														
351		X												X														
352		X												X														
353																				X								
354																				X								
355																				X								
356																				X								
357																				X								
358																				X								
359																				X								
360																				X								
361																				X								
362																				X								
363																				X								

Figura 2 – Armazenamento dos dados relativos às características dos compostos nas folhas de cálculo. Cada uma das 28 características encontra-se numa coluna e a sua presença em cada um dos compostos é assinalada com um “X”.

- **Vista sobre o conjunto dos dados muito limitada** – as folhas de cálculo não permitem seleccionar apenas determinadas zonas do conjunto de dados ou fazer ordenações aos mesmos, o que torna a sua consulta muito complicada. Para visualizar dados mais abaixo na folha não conseguimos ver o cabeçalho das colunas e quando existem múltiplas folhas de cálculo temos de memorizar o termo que faz a correspondência entre ambas. Outro problema prende-se com o armazenamento das imagens dos compostos (por exemplo no formato JPEG), que não se encontram directamente nas folhas de cálculo, e mais uma vez tem de haver um código de ligação entre os dados dos composto e a sua imagem estrutural (Figura 3).

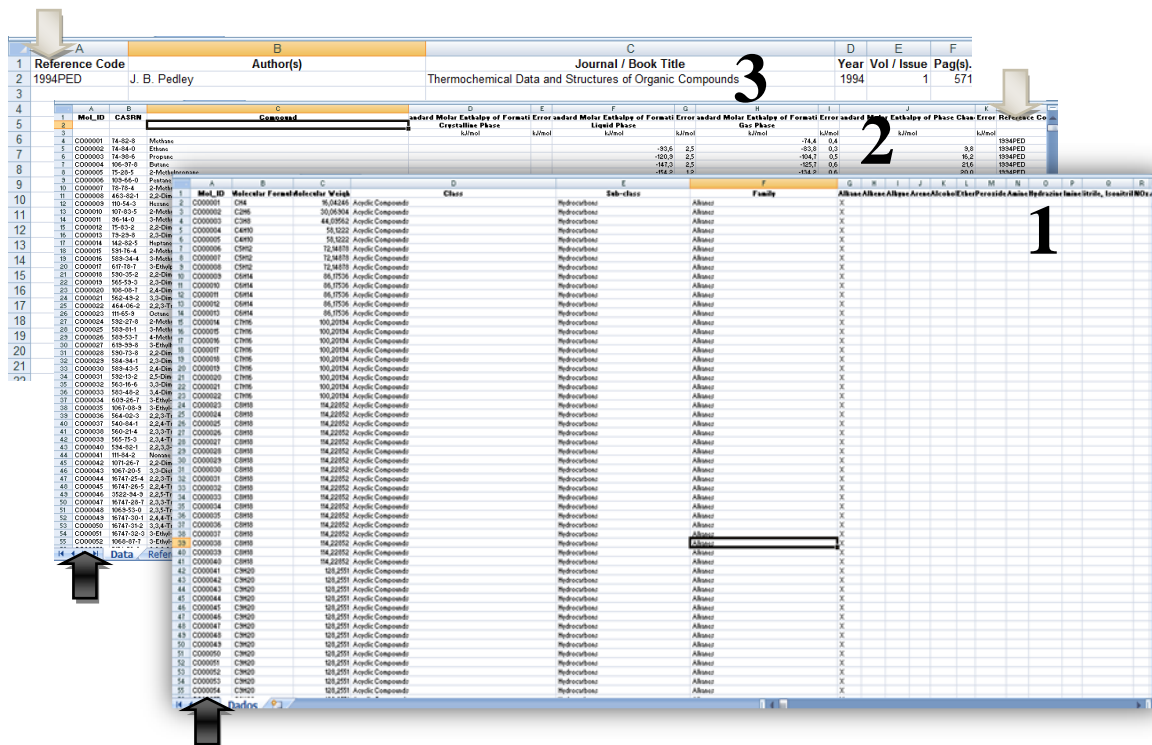


Figura 3 – Exemplo das folhas de cálculo a consultar para obter a informação completa relativamente a um composto. Pode verificar-se que não conseguimos ter uma vista completa de todas as linhas e colunas do conjunto de dados. As setas a preto indicam o código de ligação entre os dados estruturais (1) e os dados termoquímicos (2) e as setas a cinzento indicam o código de ligação entre os dados termoquímicos (2) e as referências (3).

- **Capacidade e performance** – à medida que o conjunto de dados aumenta, a eficiência do uso das folhas de cálculo vai decrescendo devido à quantidade de memória necessária. Dado o número de compostos químicos existentes é previsível que o conjunto de dados aumente muito, tornando-se incomportável o uso deste tipo de formato de armazenamento e consulta.
- **Multiuso e partilha de dados pela comunidade científica com diferentes privilégios** – o multiuso do conjunto de dados utilizando as folhas de cálculo requer muita disciplina e atenção, especialmente se o grupo de pessoas for extenso, podendo o mesmo ser facilmente corrompido. Não é possível restringir os utilizadores de aceder e/ou modificar determinadas informações.

- **Evolução** – a utilização das folhas de cálculo dificulta a evolução do conjunto de dados através de novas integrações e a aplicação de métodos de previsão de propriedades ou de outros tipos.

A partir destas premissas e com vista a encontrar uma solução para as mesmas, foi estruturada a proposta deste trabalho: a implementação de um Sistema de Informação para coligir e apresentar propriedades estruturais e termoquímicas de compostos orgânicos. Este Sistema de Informação foi baptizado de *ThermInfo*.

1.3. Objectivos

O Sistema de Informação que se preconiza, utilizará o conjunto de dados estruturais e termoquímicos de base, recolhido e criticamente avaliado pelo Grupo de Energética Molecular. Os principais objectivos são organizar e disponibilizar essa informação de forma estruturada, coerente e simples, permitindo a cooperação entre a comunidade científica e a evolução ao longo do tempo (com a possibilidade de integrar métodos estatísticos e de previsão de propriedades, assim como outros tipos de ferramentas).

Como objectivos específicos, este trabalho terá as seguintes fases:

- Compreender as necessidades dos utilizadores e de que forma este Sistema de Informação pode potenciar e facilitar as tarefas de pesquisa e organização de dados termoquímicos de compostos orgânicos;
- Analisar os requisitos desta ferramenta para compilar e apresentar dados termoquímicos e estruturais de forma adequada tendo em conta os utilizadores;
- Implementar e avaliar a ferramenta tendo em conta os aspectos mencionados acima.

1.4. Metodologia

Este projecto recorre às tecnologias de informação e procura responder simultaneamente às necessidades de armazenamento, manipulação e visualização de dados termoquímicos. Para isso, encontra-se dividido em 4 fases de desenvolvimento (Figura 4), interdependentes e complementares, para que as funcionalidades e a interface sejam adequadas às tarefas e às capacidades dos investigadores na área da termoquímica [7-8]:

- 1. Percepção** – É formulado e analisado o problema e são delineadas soluções com base em análises de requisitos dos utilizadores e do sistema. Esta análise inclui entrevistas a potenciais utilizadores, em que são discutidas as suas necessidades e problemas na manipulação de dados termoquímicos.
- 2. Concepção** – São geradas e avaliadas soluções, modelando o sistema, e são feitas as escolhas das soluções mais apropriadas. Para isso, são desenvolvidos e avaliados protótipos de baixa fidelidade e diagramas de classes, utilizando *Unified Modeling Language* (UML), que modelam as funcionalidades do sistema.
- 3. Implementação** – É implementada a solução escolhida na fase anterior. Os dados são armazenados numa base de dados *My Structured Query Language* (MySQL)⁴, com uma arquitectura adequada aos dados estruturais e termoquímicos. As funcionalidades de pesquisa, inserção e gestão dos dados, são realizadas através de uma interface *Web* implementada em *Hypertext Preprocessor* (PHP)⁵, *Java Script* (JS), *HyperText Markup Language* (HTML) e *Cascading Style Sheets* (CSS).

⁴ MySQL: <http://www.mysql.com/>

⁵ PHP: <http://www.php.net/>

4. **Avaliação** – A usabilidade e possíveis problemas da interface são testados com e sem utilizadores. Para isso são utilizadas uma série de heurísticas, estruturados questionários e um conjunto de tarefas a realizar através do qual é monitorizado o comportamento dos utilizadores (temporalmente e em termos do número de erros cometidos).

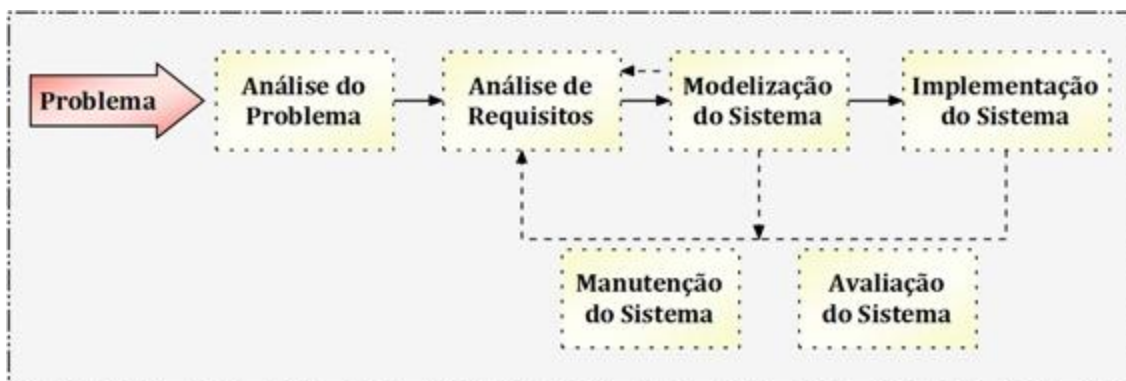


Figura 4 – Metodologia adoptada para o desenvolvimento do sistema *ThermInfo*, dividida em fase distintas que se complementam e interagem entre si.

1.5. Organização do Documento

Este documento está organizado da seguinte forma:

- Capítulo 1, enquadra e apresenta o problema em estudo assim como a motivação e os benefícios de encontrar e desenvolver uma solução para o mesmo.
- Capítulo 2, introduz alguns conceitos de Química necessários para contextualizar este trabalho, nomeadamente no que diz respeito à natureza dos dados.
- Capítulo 3, descreve funcionalmente a ferramenta desenvolvida e a metodologia utilizadas para a sua implementação.

- Capítulo 4, apresenta e discute os resultados alcançados que permitem avaliar a base de dados e a interface implementadas.
- Capítulo 5, são expostas as principais conclusões deste trabalho e direcções futuras.

2

Conceitos Básicos

“If we knew what it was we were doing, it would not be called research, would it?”

- Albert Einstein

No Capítulo 1 foi discutida a motivação para a construção de um Sistema de Informação para coligir e apresentar propriedades estruturais e termoquímicas de compostos orgânicos e as vantagens que este pode oferecer quando comparado com os métodos de armazenamento e consulta anteriormente utilizados. Neste Capítulo será dada uma visão geral sobre as propriedades dos compostos orgânicos presentes no sistema *ThermInfo* e a sua importância.

2.1. Compostos Orgânicos

Denominam-se por compostos orgânicos os compostos químicos que são constituídos por átomos de carbono, hidrogênio e por vezes também por outros elementos como o oxigênio, o azoto, o enxofre, os halogéneos e o fósforo [9-12].

A maioria destes compostos obtêm-se por síntese, embora a maneira mais fácil de os obter seja isolá-los a partir de produtos animais ou vegetais [9-12].

Estes compostos são de extraordinária importância: existem em elevado número devido às características especiais do átomo de carbono; as suas moléculas podem atingir grandes dimensões e complexidade; e são a base de todos os seres vivos e de um grande número de produtos [9-12].

Tal como já foi referido anteriormente, o sistema *ThermInfo* contém dados relativos a compostos orgânicos, que podem ser divididos em três categorias:

1. Dados Estruturais, constituídos por descritores que especificam a estrutura molecular dos compostos, mostrando a forma como os átomos estão ligados, o tamanho da molécula e outras propriedades. Os dados estruturais podem reunir e sistematizar um elevado número de características relativos a centenas de milhares de compostos diferentes [9-12].

- *Identity Descriptor (ID) Molecular*, é um identificador único para cada um dos compostos, atribuído pelo sistema *ThermInfo*. Este ID tem o formato CONNNNN (N é um dígito).
- *Nome do Composto*, é o nome atribuído a cada um dos compostos baseado numa nomenclatura sistemática de acordo com as recomendações da *International Union of Pure and Applied Chemistry (IUPAC)*⁶ [13].
- *CAS Registry Number (CASRN)*, é um identificador único, instituído e atribuído a cada substância química pelo CAS⁷. Este identificador não tem qualquer significado químico e é atribuído numa ordem sequencial, de forma a assegurar a unicidade. Tem o formato NNNNNNN-NN-N (1 a 7 dígitos, hífen, 2 dígitos, hífen, 1 dígito). O último dígito é um dígito de controlo para verificar a validade e unicidade do identificador e é calculado da seguinte forma: multiplicar o último dígito por 1, o dígito seguinte por 2 e assim sucessivamente; todos estes produtos são somados; finalmente é computado o módulo 10 da soma. Por exemplo, o CASRN do *methanol* é 67-56-1, o

⁶ IUPAC recommendations: <http://www.iupac.org/web/ins/2001-043-1-800>

⁷ CAS Registry Number [<http://www.cas.org/expertise/cascontent/registry/regsys.html>]

ThermInfo: Sistema de Informação para Coligir e Apresentar Propriedades Termoquímicas

dígito controlo 1 é calculado da seguinte forma: $(6*1 + 5*2 + 7*3 + 6*4) = 61$; $61 \bmod 10 = 1$.

- *Fórmula Molecular*, dá indicação dos elementos químicos presentes e do número de átomos de cada um desses elementos, representado em índice [9-12]. No sistema *ThermInfo*, os elementos encontram-se dispostos na seguinte ordem, CHXNOS (C = carbono, H = hidrogénio, X = halogéneo [flúor, cloro, bromo, iodo], N = azoto, O = oxigénio, S = enxofre).
- *Estrutura Química*, apresenta a geometria molecular, mostrando o arranjo espacial dos átomos numa molécula e as ligações químicas que os mantêm juntos [9-12]. A estrutura química dos compostos encontra-se representada no formato 2D, num ficheiro JPEG.
- *Peso Molecular*, é a soma dos pesos atómicos de todos os átomos que constituem a molécula. Indica quantas vezes uma molécula é mais pesada que a duodécima parte de um átomo de carbono-12 [9-12].
- *Estado Físico*, indica a situação em que o composto se encontra no que diz respeito às suas propriedades e ao movimento das partículas, dependendo da temperatura e pressão [9-12]. Pode dividir-se em:
 - Cristalino: os compostos possuem tamanho e forma definidos porque os seus átomos estão muito próximos, ligados por forças de coesão consideráveis e possuem ligeira vibração no que diz respeito à sua posição média.
 - Líquido: os compostos possuem propriedades intermédias entre os sólidos e os gases.
 - Gasoso: a principal característica dos compostos no estado gasoso é a mobilidade dos seus componentes, o que permite às substâncias ocuparem todo o volume dos recipientes que os contêm. O movimento é desordenado e as forças de interação entre as moléculas são fracas.

- SMILES (*Simplified Molecular Input Line Entry System*), é uma especificação para descrever a estrutura química das moléculas usando uma curta sequência de caracteres *American Standard Code for Information Interchange* (ASCII). Um ponto importante é que existem diferenças entre maiúsculas e minúsculas. Por exemplo, o *cyclohexane* tem o SMILES C1CCCCC1 enquanto o *benzene* tem o SMILES c1ccccc1 [14].
 - USMILES (*Unique Simplified Molecular Input Line Entry System*), é um tipo especial e único de SMILES entre todas as possibilidades válidas, para uma dada estrutura molecular [15].
 - Classe, divide os compostos nas grandes classes estruturais (por exemplo, cadeia aberta, cíclicos, aromáticos, poli-aromáticos, etc.) [9-12].
 - Subclasse, divide os compostos pelo tipo de átomos presentes (por exemplo, CH, CHO, CHN, etc.) ou pelo tamanho dos ciclos presentes nos compostos cíclicos (anéis de 3, 4, 5, etc. átomos) [9-12].
 - Família, separa os compostos em famílias químicas de acordo com os arranjos de átomos (grupos funcionais) mais relevantes existentes na molécula [9-12].
 - Características, são *tags* atribuídas aos compostos e estão relacionadas com a presença de um determinado grupo funcional, o qual é responsável pelas propriedades químicas ou por determinadas características físicas do composto [9-12].
- 2. Dados Termoquímicos.** A termoquímica estuda a energia associada a uma reacção química. A reacção é classificada como exotérmica se se realiza com libertação de energia ou endotérmica se se realiza com absorção de energia [16-17].

- Entalpia de Formação Molar Padrão, é a variação de entalpia⁸ para a reacção em que um composto químico é formado a partir dos seus elementos constituintes, cada um no seu estado de referência padrão [16-17]. No sistema *ThermInfo* são apresentados valores de entalpia de formação molar padrão e respectivos erros para os estados físicos cristalino, líquido e gasoso.
- Entalpia de Transição de Estado Molar Padrão, é a variação de entalpia associada aos processos físicos de transição de estado físico [16-17]:
 - Fusão, transição do estado sólido para o estado líquido;
 - Vaporização, transição do estado líquido para o estado gasoso;
 - Sublimação, transição do estado sólido para o estado gasoso.

3. Dados Bibliográficos, referências completas relativamente à origem dos dados termoquímicos dos compostos, incluindo: autor(es), revista científica/título do livro, ano, volume e página(s).

2.2. Importância do Conhecimento da Energética dos Compostos Químicos

O conhecimento da energética dos compostos químicos contribui para a compreensão da natureza das ligações químicas e, conseqüentemente, para o estudo das relações de interdependência/correlação entre energética, estrutura e reactividade [18].

O conhecimento da entalpia de formação de um composto químico no estado gasoso é um parâmetro termodinâmico muito importante para a avaliação da sua perigosidade energética potencial numa instalação industrial. Muitos dos critérios de avaliação da perigosidade dos compostos químicos propostos pela *American Society for*

⁸ É uma função termodinâmica, especialmente útil quando se lida com processos a pressão constante, definida por $H = E + PV$, onde E é a energia, P a pressão e V o volume.

Testing and Materials (ASTM)⁹ têm por base a utilização da entalpia de formação do composto no estado gasoso [19-20].

A entalpia de formação padrão é fundamental para a avaliação do potencial de decomposição exotérmica de um composto (geralmente, quanto mais positivo é o valor da entalpia maior será o potencial de decomposição exotérmica) [18].

Este parâmetro termoquímico pode ser obtido por via experimental (metodologias complexas e nem sempre aplicáveis), ou por métodos estatísticos de previsão (por extrapolação de dados já existentes). Assim, é evidente a grande necessidade de se organizar e disponibilizar os valores desta propriedade, de modo a permitir, por exemplo, a aplicação de métodos empíricos para a sua previsão [21].

⁹ ASTM: <http://www.astm.org/>

3

Metodologia e Implementação

“ We never do anything well till we cease to think about the manner of doing it.”

- William Hazlitt

Ao longo deste capítulo será apresentada a metodologia utilizada no desenvolvimento do sistema *ThermInfo* e os detalhes da sua implementação. Inicia-se com o levantamento do contexto do problema, e uma análise de requisitos do público-alvo e das funcionalidades do sistema. É também apresentada a descrição e implementação da base de dados, tal como da interface do sistema.

3.1. Análise do Problema

Nesta etapa, foram realizadas reuniões e entrevistas com potenciais utilizadores do sistema e analisada a documentação existente de forma a clarificar o problema e discutir propostas de resolução [22].

Tal como foi descrito no Capítulo 1, é necessário o desenvolvimento de modelos de dados adequados às necessidades do Grupo de Energética Molecular – Centro de

Química e Bioquímica e da comunidade termoquímica, de forma a armazenar e partilhar os dados descritos no Capítulo 2, processá-los e disponibilizar informação que possa ser visualizada em formato coerente. O sistema também deverá permitir uma gestão de dados adequada e a cooperação entre investigadores da área.

3.2. Análise de Requisitos

Esta etapa tem como principais objectivos validar e sintetizar os resultados obtidos na fase anterior. Seguidamente, são apresentados os requisitos do público-alvo e do sistema que se pretende desenvolver [22].

3.2.1. Público-alvo

O público-alvo do sistema *ThermInfo* são essencialmente pessoas com formação em Química e que trabalham na área da Energética Molecular. Assim, estamos perante utilizadores altamente qualificados para a análise destes dados, mas pouco especializados em Informática e Tecnologias de Informação. É, também, importante salientar que de uma forma geral estes utilizadores irão ter pouco tempo disponível para uso do sistema. Quando questionados acerca da utilização de um sistema deste tipo, os objectivos esperados são o acesso rápido e fácil à informação termoquímica.

3.2.2. Sistema

Dada a análise das características do público-alvo, na concepção do sistema temos de ter em conta:

- que os dados devem ser armazenados numa base de dados com arquitectura relacional adequada, de forma a que todas as operações se possam processar de forma rápida, eficiente e sem exigir elevados recursos computacionais [23];
- a integridade dos dados, ou seja, garantir que estes devem ser devidamente processados e administrados para evitar incorrecções e

incoerências que podem prejudicar o trabalho do público-alvo [23]. Posteriormente, devemos garantir que a informação correcta não seja alterada intencionalmente ou acidentalmente e dificultar o uso de *software* malicioso que tenha como objectivo fazer *spam* ou copiar o conjunto de dados. Só determinados utilizadores autorizados podem ter acesso aos dados/informação para os alterar/corrigir;

- que a interface deve ser adequada à tarefa, fácil de usar e oferecer ajuda aos utilizadores com tutoriais exemplificativos, de forma a minimizar os eventuais problemas com a utilização de Tecnologias de Informação e diminuir o tempo de aprendizagem necessário;
- que a interface deve também ser simples e focada no objectivo, já que os utilizadores não têm muito tempo disponível. A apresentação de muita informação, muitas vezes irrelevante, irá exigir mais tempo para desempenhar as tarefas;
- a extensibilidade, a adição ou modificação de funcionalidades não deve interferir com o funcionamento das restantes;
- a capacidade de aumentar o conjunto de dados armazenados através dos contributos de utilizadores registados no sistema, sujeitos a posterior validação por parte dos administradores;
- a manutenção, com a realização de *backups* automáticos do conjunto de dados e funcionalidades administrativas adequadas.

Assim, as principais funcionalidades (Figura 5) requeridas podem ser divididas em duas grandes classes:

1. Pesquisar compostos, utilizando um dos dois métodos disponíveis:

- Pesquisa Simples – pesquisa os compostos com base num termo de pesquisa e no tipo de pesquisa seleccionado. Dada a natureza dos dados, será relevante fazer uma pesquisa deste tipo com base no ID, nome, fórmula molecular, CASRN ou SMILES do composto;

- Pesquisa Estrutural – pesquisa os compostos com base na utilização de vários termos de pesquisa correspondentes à estrutura da molécula, restringindo assim os resultados que vão ser obtidos.

Uma pesquisa com sucesso, utilizando um destes dois métodos, deverá apresentar:

- Uma lista de resultados com o resumo da informação sobre o composto (ID, nome, fórmula molecular, CASRN e SMILES), ordenada de acordo com a relevância para o(s) termo(s) de pesquisa.
- Uma ficha por composto, com toda a informação organizada de uma forma estruturada e coerente, quando tal for desejado pelo utilizador através de uma hiperligação existente na lista de resultados.

Funcionalidades	Administradores	Utilizadores Registrados	Utilizadores
Pesquisa Simples	✓	✓	✓
Pesquisa Estrutural	✓	✓	✗
Reportar Erros nos Dados	✓	✓	✗
Inserção de Dados	✓	✓	✗
Actualização de dados	✓	✗	✗
Eliminação de dados	✓	✗	✗
Validação de dados inseridos	✓	✗	✗
Controlo do Sistema	✓	✗	✗

Figura 5 – Funcionalidades disponibilizadas pelo sistema *ThermInfo* aos diferentes tipos de utilizadores. O preenchimento da célula com um visto indica que a funcionalidade está disponível para esse tipo de utilizador.

2. Gerir os dados dos compostos:

- Reportar Erros nos Dados – disponível para todos os utilizadores, permite reportar aos administradores do sistema *ThermInfo* possíveis erros nos dados existentes na base de dados.
- Inserir dados - permite aos utilizadores registados a inserção de novos dados para compostos orgânicos no sistema *ThermInfo*. Esta funcionalidade pressupõe o registo e *login* no sistema, tal como a posterior validação dos dados inseridos por um administrador. O principal objectivo desta função é a de apoiar a expansão da base de dados por parte da comunidade termoquímica.
- Apagar/Actualizar dados – permite aos administradores pesquisar por um composto com dados suspeitos ou desactualizados e proceder à sua remoção ou actualização na base de dados do sistema *ThermInfo*.
- Validar dados - permite aos administradores validar os dados relativos a compostos inseridos por utilizadores registados. Esta funcionalidade permite manter controlada a qualidade e a consistência termodinâmica dos dados disponibilizados pelo sistema.
- Controlar o Sistema – permite aos administradores monitorizar a evolução e o uso da base de dados do sistema *ThermInfo*. Permite também adicionar novos utilizadores que poderão inserir novos compostos orgânicos no sistema.

Na Figura 6 estão representados os actores (Utilizadores e Administradores), que exercem o papel de utilizadores, e a sua interacção com os *Use-Cases*, ou seja, as funcionalidades que o sistema tem disponíveis [24-25]. O actor Administrador possui acesso a todos os *Use-Cases* disponíveis, podendo também fazer o papel de Utilizador. O actor Utilizador pode pesquisar e visualizar os dados dos compostos orgânicos, criar uma conta no sistema, reportar erros nos dados e, mediante *login*, inserir novos dados.

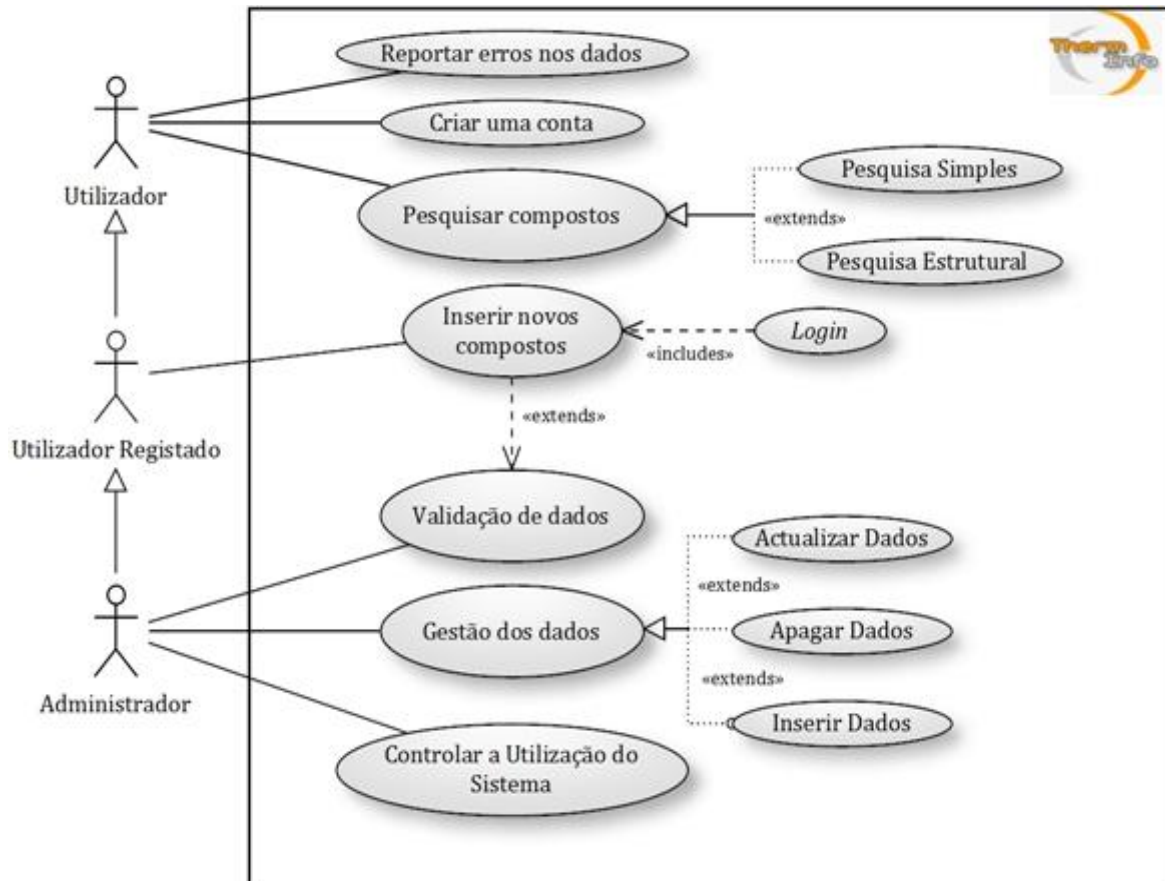


Figura 6 – Visão global das funcionalidades do *ThermInfo* em termos de actores e dependências entre *Use-cases*.

3.3. Base de Dados

A estrutura da base de dados foi concebida para permitir o acesso rápido, acomodar dados heterogêneos e manter a integridade referencial dos mesmos. O conjunto de dados é constituído pelas propriedades estruturais e termoquímicas de compostos orgânicos, termodinamicamente consistentes, seleccionados de literatura científica relevante e descritos no Capítulo anterior, assim como as respectivas referências bibliográficas. Foi desenvolvido um modelo relacional constituído por entidades que contêm os atributos que as caracterizam e pelas relações entre elas. Cada linha da entidade (tuplo) representa uma colecção de dados relacionados. O modelo relacional foi escolhido uma vez que tem um bom desempenho e facilidade em realizar consultas aos dados (utilizando uma linguagem de alto nível, *Structured Language Query* (SQL)), é fácil de administrar, tem ampla aceitação e está muito bem

documentado (o que facilita a utilização e desenvolvimento de aplicações que trabalhem sobre a base de dados). Para além destes motivos, uma abordagem orientada a objectos não era indicada para a estrutura deste conjunto de dados [23].

A Figura 7 mostra o diagrama de classes (UML) da estrutura da base de dados [23-24]. Podemos considerar quatro categorias: os dados estruturais, os dados termoquímicos, as referências bibliográficas e os dados dos utilizadores registados.

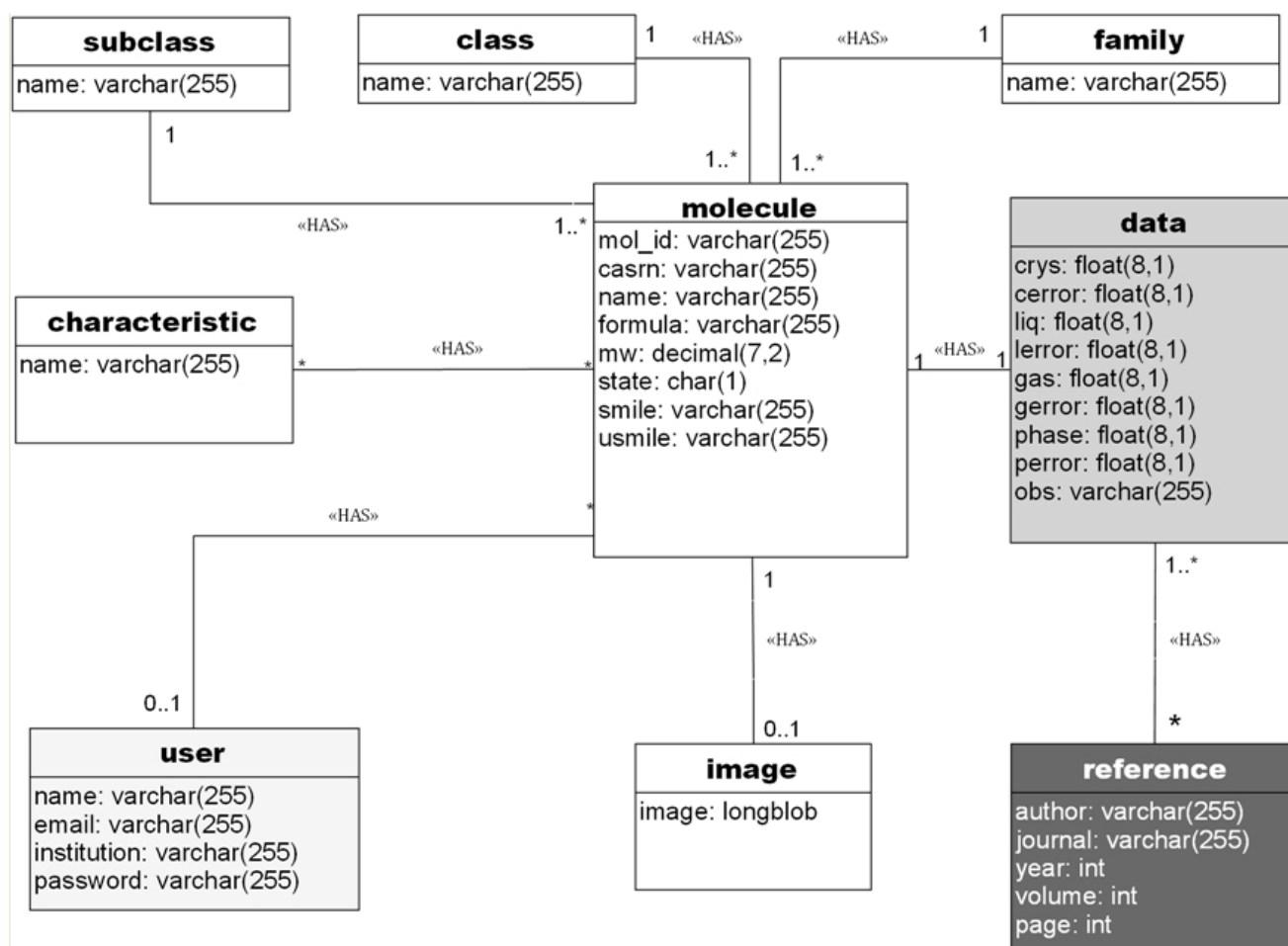


Figura 7 – Diagrama de classes UML da base de dados do *ThermInfo*. As tabelas estão dispostas segundo um código de cores que se refere a cada uma das categorias descritas no texto: branco para dados estruturais, cinzento-médio para dados termoquímicos, cinzento-escuro para dados bibliográficos e cinzento-claro para os utilizadores registados no sistema.

Dentro da categoria dos dados estruturais temos 6 entidades relacionadas entre si da seguinte forma:

- a entidade central *molecule* contém oito atributos (ID molecular, CASRN, nome, fórmula molecular, peso molecular, estado físico, SMILES e USMILES), dos quais quer o ID molecular quer o CASRN são únicos.
- as entidades *class*, *subclass* e *family* têm como atributo o respectivo nome. Cada composto terá uma *class*, *subclass* e *family*. Cada *class*, *subclass* e *family* podem ser atribuídas a vários compostos (*molecule*).
- a entidade *characteristic* tem como atributo o nome da característica e a chave primária é atribuída incrementalmente. Cada composto poderá ter várias *characteristic*, assim como cada *characteristic* pode pertencer a vários compostos. De forma a melhorar a performance das operações de selecção de dados nesta relação de “muitos para muitos” as colunas foram indexadas¹⁰.
- a entidade *image* tem como atributo a própria imagem do composto. Cada composto apenas poderá ter uma imagem e cada imagem corresponde a um único composto.

Dentro da categoria dos dados termoquímicos temos uma entidade, *data*, que tem 9 atributos (os valores das entalpias e respectivos erros descritas no Capítulo 2). Cada composto terá apenas um valor de cada um dos atributos desta entidade. Esta categoria também se relaciona com os dados bibliográficos (*reference*), sendo que os dados termoquímicos poderão ter várias referências bibliográficas. Mais uma vez, de forma a melhorar a performance das operações de selecção de dados nesta relação de “muitos para muitos” as colunas foram indexadas.

Dentro da categoria dos dados bibliográficos temos uma entidade, *reference*, que tem 5 atributos (autor, nome da revista/título do livro, ano, volume e página). Esta

¹⁰ Índices de colunas: <http://dev.mysql.com/doc/refman/5.0/en/mysql-indexes.html>

entidade relaciona-se com a entidade que descreve os dados termoquímicos (*data*), tal como já foi explicado anteriormente.

Dentro da categoria dos dados dos utilizadores registados temos uma entidade, *user*, que tem 4 atributos (nome, e-mail, instituição e palavra-chave). O atributo e-mail é único, de forma a garantir que cada utilizador terá apenas uma conta. Cada utilizador poderá inserir vários compostos, mas cada composto é inserido por um só utilizador. Esta relação entre a entidade *molecule* e *user* permite visualizar a contribuição do utilizador com novos dados.

3.4. Arquitectura do Sistema

A base de dados foi implementada utilizando o SGBD MySQL e foi utilizado o *phpMyAdmin*¹¹ para gestão do seu conteúdo a partir de uma interface *Web*. O *phpMyAdmin* permite criar, remover e alterar bases de dados e tabelas, inserir, remover e editar dados, executar *queries* SQL, exportar e importar a base de dados, entre outros processos administrativos (Figura 8 e Figura 9).

A base de dados foi populada com dados recolhidos pelo Grupo de Energética Molecular – Centro de Química e Bioquímica e armazenados em folhas de cálculo. Para importar estes dados, foi desenvolvido um *script* em Practical Extraction and Report Language (Perl)¹² que se liga à base de dados MySQL, abre as folhas de cálculo e vai inserindo cada uma das linhas de dados num *array*. Seguidamente, divide cada linha pelo carácter delimitador (espaço em branco ou *tab*) e realiza *queries* SQL para inserir esses dados na entidade/atributo correspondente. Futuramente, os dados serão inseridos a partir da interface *Web*.

¹¹ *phpMyAdmin*: <http://www.phpmyadmin.net/>

¹² Perl: <http://www.perl.org/>

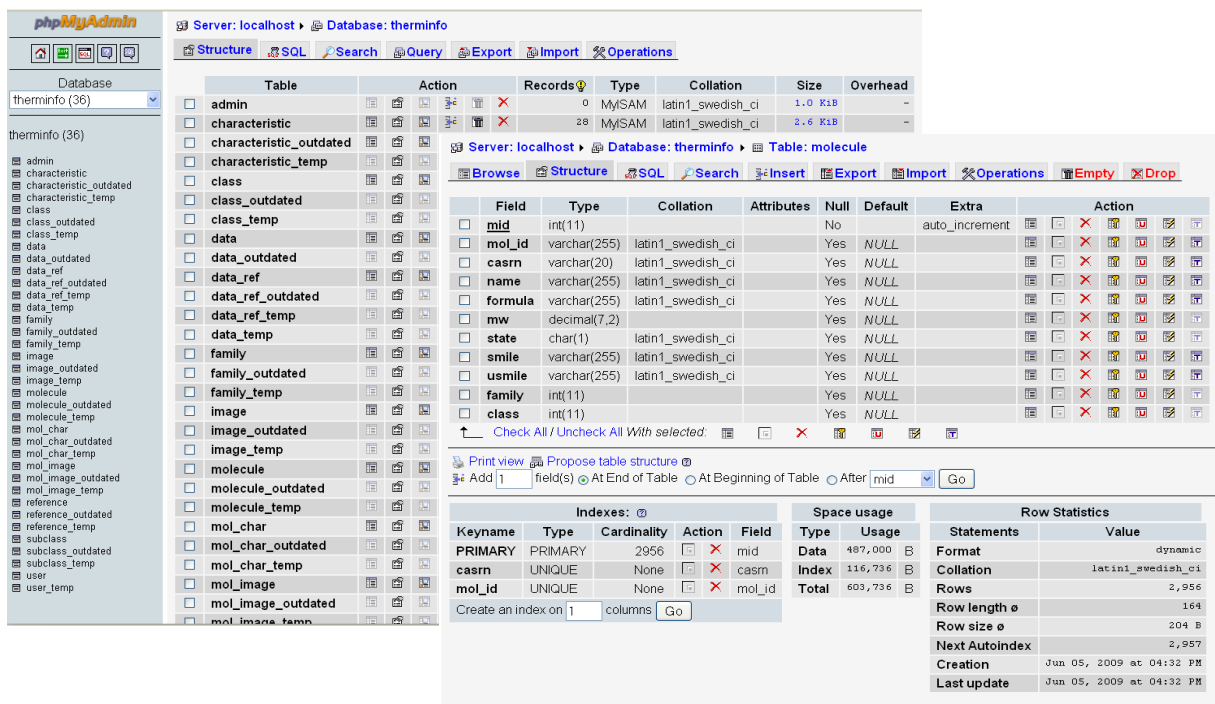


Figura 8 – Administração das tabelas do sistema *ThermInfo*, utilizando o phpMyAdmin.

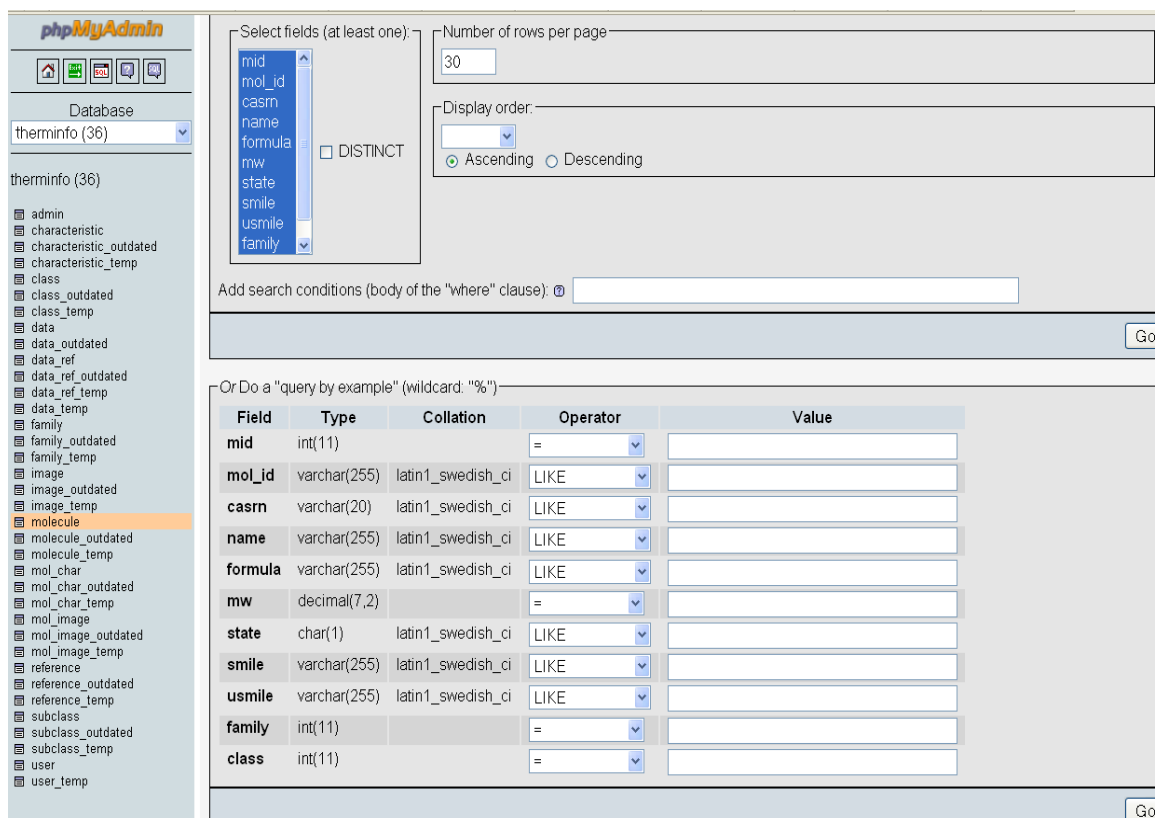


Figura 9 – Utilização do phpMyAdmin para realizar *queries* SQL à base de dados do sistema *ThermInfo*.

A interface foi desenvolvida utilizando PHP, que é uma linguagem de programação *server-side* (processada no servidor) desenvolvida especialmente para a *Web*, com a possibilidade de ser embebida em código HTML.

O PHP está muito bem documentado, suporta transacções intensivas, executa rápido, funciona bem com outros *softwares* escolhidos para o desenvolvimento deste projecto, o MySQL e *Javascript* (JS). A escolha do PHP permitiu também utilizar *Dynamic Hyper-Text Markup Language* (DHTML) para fornecer e actualizar na interface informação armazenada na base de dados [26].

O controlo de acesso às funções administrativas é realizado utilizando a funcionalidade *hypertext access* (.htaccess) do servidor Apache¹⁴.

O *ThermInfo* encontra-se oficialmente disponível no *Website* <http://www.therminfo.com>. Um esquema resumo da arquitectura do sistema *ThermInfo* é apresentado na Figura 10.

¹⁴ Apache: <http://www.apache.org/>

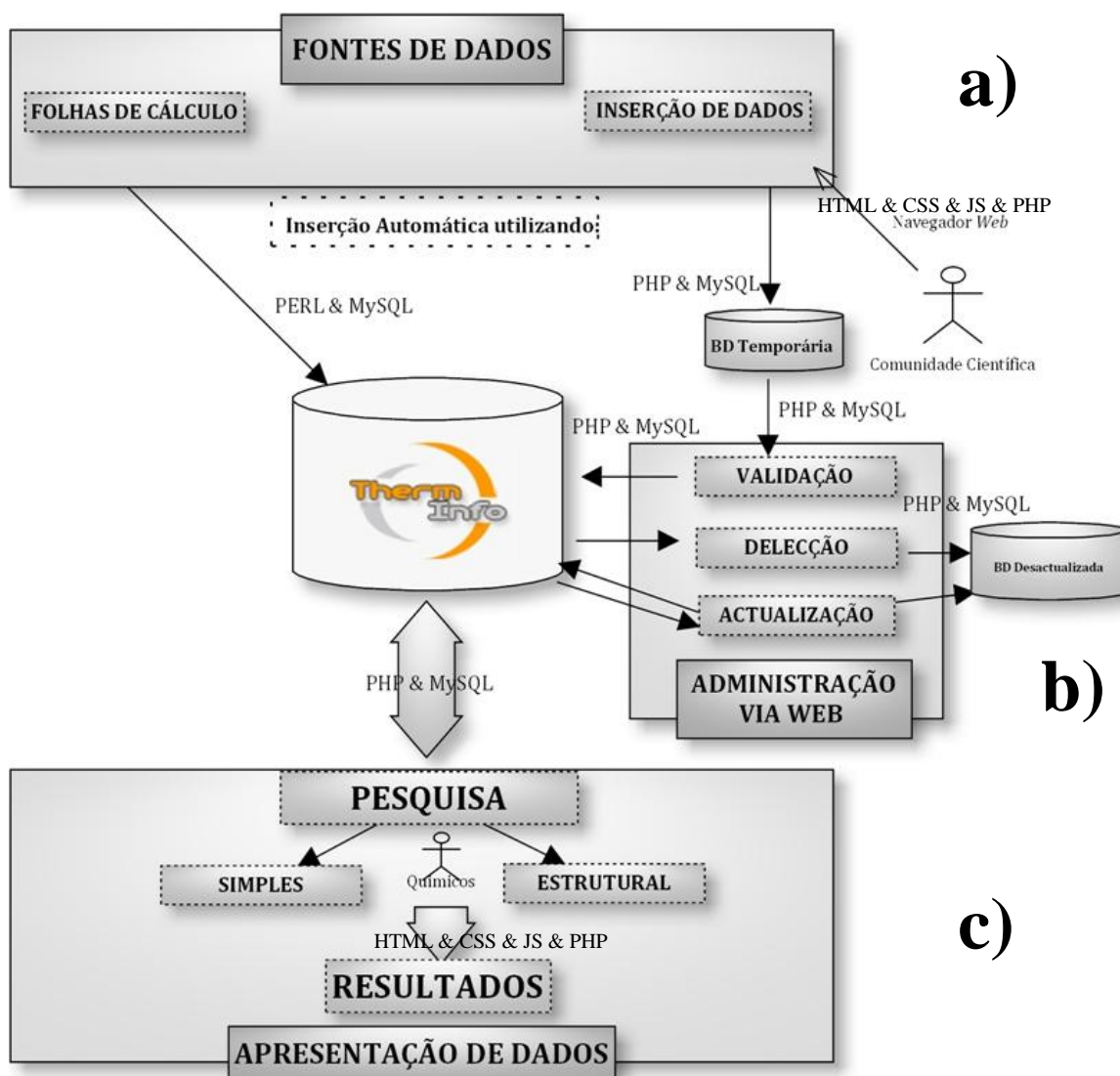


Figura 10 – Modelo simplificado do funcionamento do sistema *ThermInfo*. Neste esquema estão representadas as fontes de dados, as funcionalidades administrativas, a pesquisa e apresentação de dados e a forma como estas se relacionam com a base de dados. a) Fontes de dados: (1) as folhas de cálculo, nas quais os dados são automaticamente inseridos na base de dados utilizando *scripts* Perl e MySQL; (2) a inserção de dados, por utilizadores, via navegador *Web* utilizando JS, HTML, PHP e MySQL estes ficam em espera pelo processo de validação por um administrador, numa base de dados temporários. b) Funcionalidades administrativas – todas interagem com bases de dados e por isso utilizam as tecnologias PHP e MySQL (1) a validação de dados permite passar os dados inseridos pelos utilizadores (armazenados na base de dados temporária) para a base de dados que disponibiliza os dados publicamente. (2) a remoção/actualização dos dados, transfere os dados originais para uma base de dados desactualizados e no caso da actualização realiza as modificações necessárias aos dados na base de dados principal. c) A apresentação de informação pode ser obtida com base em 2 tipos de pesquisa à base de dados: simples e estrutural, utilizando as tecnologias HTML, CSS, JS, PHP e MySQL.

3.5. Interface

Como já foi descrito anteriormente, a interface do sistema *ThermInfo* tem como principais objectivos disponibilizar as funcionalidades do sistema via *Web* de forma simples e adequada às tarefas e utilizadores em questão, ou seja, cumprir os requisitos delineados.

3.5.1. Protótipos de baixa-fidelidade

A primeira abordagem ao *design* da interface foi a realização de protótipos de baixa-fidelidade, para serem testados com utilizadores. Esta abordagem tem como objectivos: explorar o universo de possibilidades de uma forma rápida e a baixo custo; colmatar a análise de requisitos realizada com utilizadores; e conceber um fio condutor ao design do protótipo de alta-fidelidade [22, 25, 27, 28].

- **Página de Entrada**

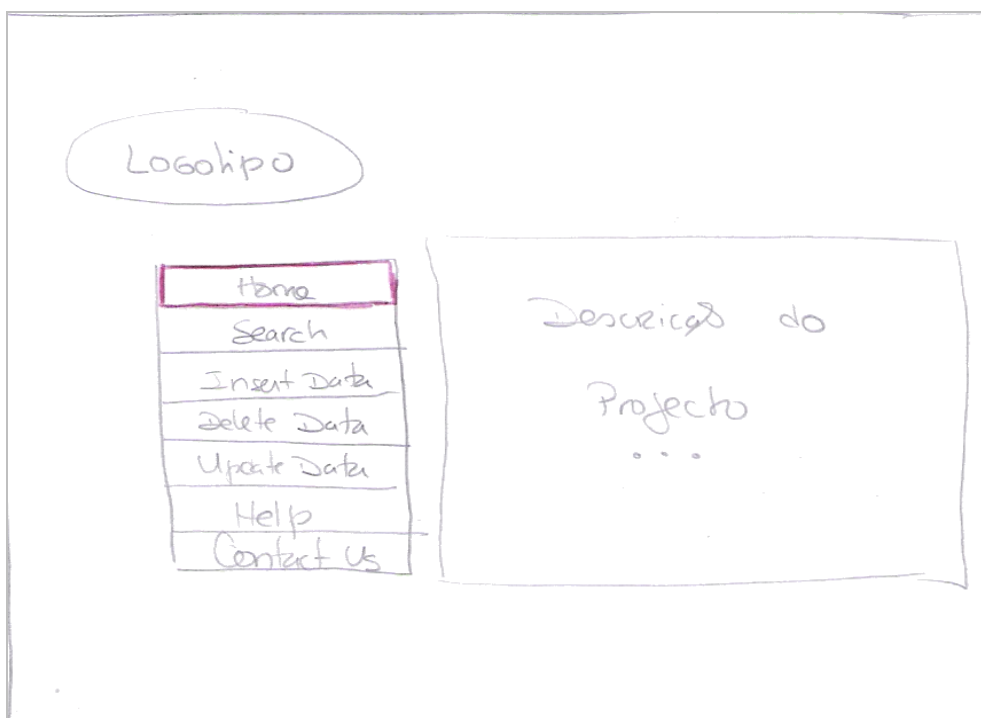


Figura 11 – Protótipo de baixa-fidelidade para a página de entrada no sistema.

A página de entrada (Figura 11) apresenta um menu lateral com as diversas funcionalidades disponíveis (que se mantém nas restantes interfaces) e uma apresentação do sistema *ThermInfo*.

• Página de Pesquisa

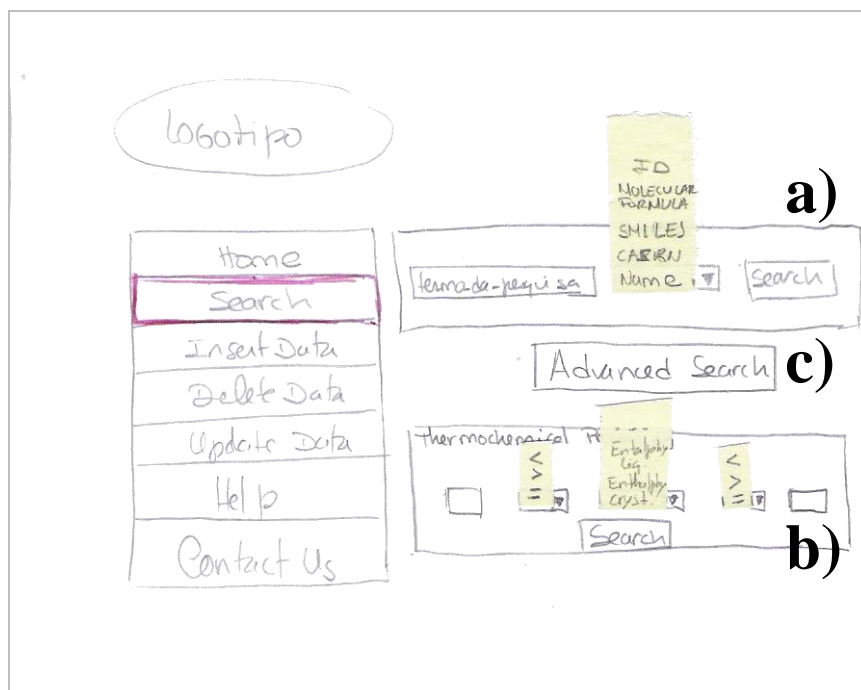


Figura 12 – Protótipo de baixa-fidelidade para a página de pesquisa de compostos no sistema.

A página de pesquisa apresenta um menu lateral com as diversas funcionalidades disponíveis e três possibilidades de pesquisa:

1. Pesquisa de um composto com base num termo de pesquisa e do parâmetro de pesquisa seleccionado de uma listagem (Figura 12 a)).
2. Pesquisa com base num intervalo de valores relativos aos dados termoquímicos (entalpias dos diversos estados físicos) (Figura 12 b)). Este tipo de pesquisa foi considerado, pelo público-alvo, sem interesse para o seu trabalho.

3. Pesquisa avançada (Figura 13) que permite especificar diversas características estruturais do composto. Foi denotado que a sua ausência no menu de funcionalidades poderia causar alguma entropia, para além de sobrecarregar a página de pesquisas simples (Figura 12 c)).


Figura 13 – Protótipo de baixa-fidelidade para a página de pesquisa avançada de compostos no sistema.

Ambas as pesquisas (simples e estrutural) obtêm um conjunto de resultados (Figura 14), sob a forma de listagem com um resumo da informação e a possibilidade de consultar mais informação (visualização da ficha do composto – Figura 15).

Name	CASRN	Molecular Formula	SMILES	
Methanol	67-56-1	CH ₄ O	CO	More Info
...	More Info
..	
...	

Figura 14 – Protótipo de baixa-fidelidade para a listagem de resultados obtida quando é realizada uma pesquisa de um composto (simples ou estrutural).

Nome do Composto



Structural Data

Molecular Id :	Compound Name :
CASRN :	Molecular Formula :
Molecular weight :	Physical State :
SMILES :	
USMILES :	

Thermochemical Data

Enthalpy - Crystalline : ...	Error : ...
" - liquid : ...	Error : ...
...	...

Observations :

References

1. author
Journal / book , chapter , year , page(s)
2.

Figura 15 – Protótipo de baixa-fidelidade para a ficha de um composto.

• Página de Inserção de dados

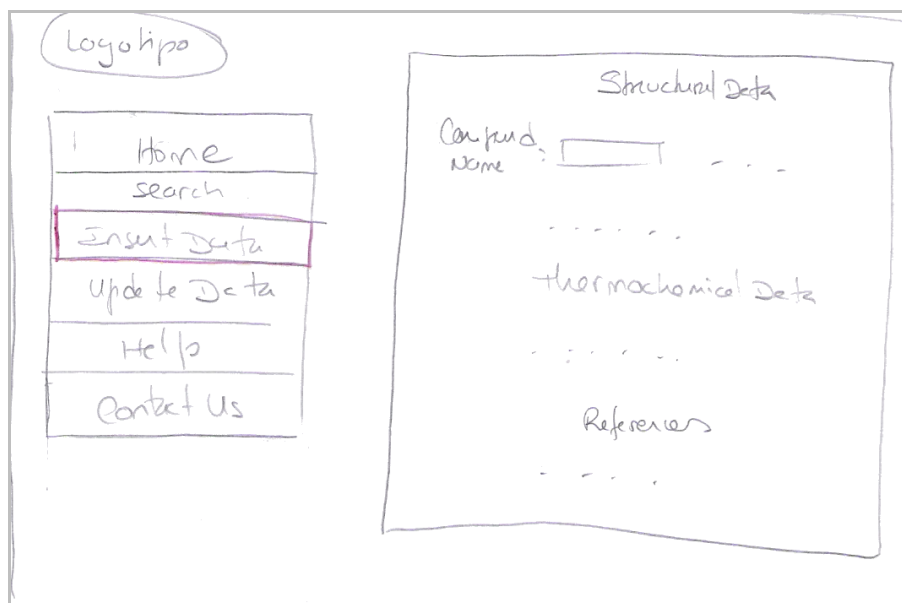


Figura 16 – Protótipo de baixa-fidelidade para a página de inserção de dados no sistema.

Nesta página é apresentado o formulário com os campos de inserção de todos os dados do composto (Figura 16).

• Página de Actualização/Eliminação de dados

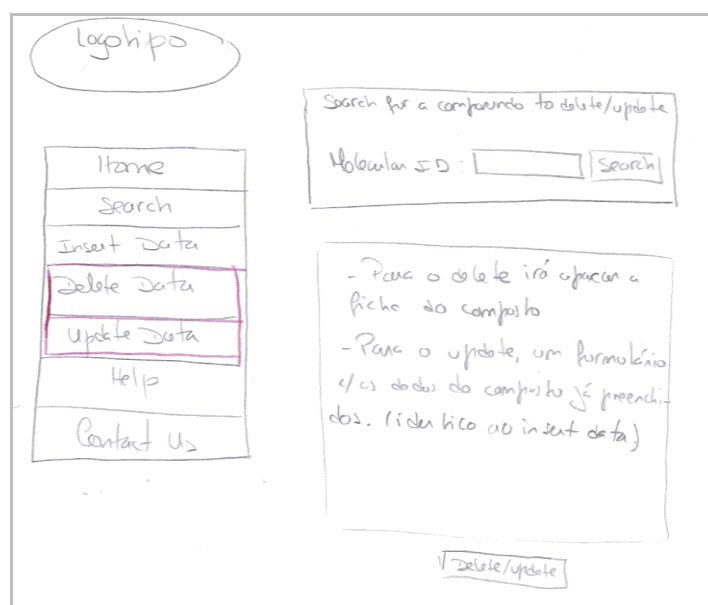


Figura 17 – Protótipo de baixa-fidelidade para a página de actualização/eliminação de dados no sistema.

Esta página (Figura 17) apresenta um campo para especificar o ID do composto a ser eliminado ou actualizado. No caso do objectivo ser a eliminação, a ficha do composto é apresentada e utilizador terá de confirmar a sua eliminação. Se o objectivo for a actualização, será apresentado um formulário idêntico ao da inserção dos dados, no entanto, são carregados os dados presentes na base de dados, relativos ao composto. O utilizador poderá modificá-los, acrescentá-los ou apagá-los e confirmar a actualização.

No desenrolar do projecto surgiram novas funcionalidades e o design da interface foi alterado em vários aspectos, relativamente a estes protótipos. No entanto, estes foram uma boa base de partida para estruturar possíveis soluções.

3.5.2.Implementação e Protótipos de alta-fidelidade

Após termos definido os objectivos, os requisitos e as linhas de design do sistema, segue-se a implementação dos protótipos de alta-fidelidade. Estes protótipos utilizam as tecnologias do sistema final, mas com algumas funcionalidades limitadas [25].

- **Página Principal da interface**

Esta página contém o logótipo e uma pequena apresentação do *ThermInfo* (Figura 18). As suas funcionalidades encontram-se destacadas num menu lateral à esquerda, em cabeçalho e em rodapé. O esquema da página irá manter-se para todas as funcionalidades de modo a manter a coerência e consequentemente reduzir a carga cognitiva necessária para lidar com o sistema.

The screenshot shows the homepage of the ThermInfo system. At the top left, the date is 'Tuesday, July 7, 2009'. A navigation bar at the top right contains links: Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us. The main content area is divided into three sections: a) a vertical menu on the left with options: Home, About Us, Search, Structural Search, Login, Insert Data, Administration, Help, and Contact Us; b) a secondary navigation bar at the bottom with the same links as the top bar; and c) the main text area, which includes a 'Welcome to ThermInfo!' heading, a sub-heading 'Collecting and presenting thermochemical properties', and a detailed paragraph about the project's goals and methods. At the bottom of the main text area, there are two logos: 'W3C XHTML 1.0' and 'W3C CSS', both with red checkmarks.

Figura 18 – Página Principal do sistema *ThermInfo*. a) Menu de funcionalidades disponíveis, para utilizadores e administradores. b) Menus de funcionalidades disponíveis, para utilizadores. c) Apresentação do *ThermInfo*.

- **Página de apresentação da equipa do sistema *ThermInfo* (*About us*)**

Esta página (Figura 19) apresenta os membros da equipa do sistema *ThermInfo* e respectivos contactos assim como as publicações relacionadas. Esta equipa envolve uma colaboração entre pessoas das áreas da Química e da Informática.

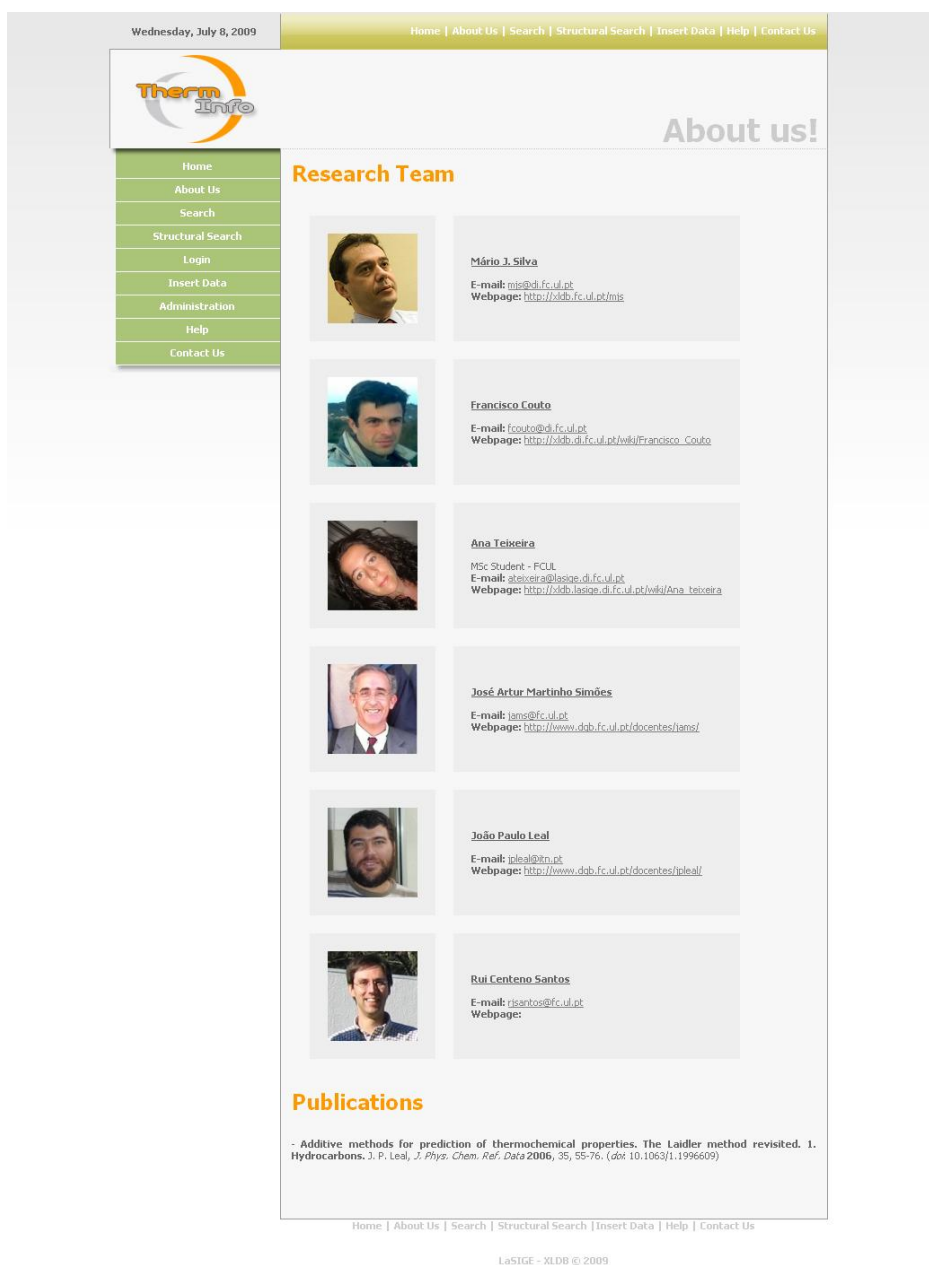


Figura 19 – Página de apresentação da equipa do sistema *ThermInfo*.

- **Código de Segurança**

Todas as funcionalidades públicas disponibilizadas pelo *ThermInfo* utilizam um “*Completely Automated Public Turing Test to Tell Computers and Humans Apart*” (CAPTCHA)¹⁵ para diferenciar humanos e máquinas. O

¹⁵ CAPTCHA: <http://captcha.net/>

principal objectivo deste código de segurança no sistema *ThermInfo* é evitar *spam* e impedir que software malicioso automatizado acesse inadvertidamente a base de dados. Este CAPTCHA (Figura 20) requer que o utilizador identifique apenas os números num código aleatório de seis caracteres composto por letras, números e/ou caracteres especiais. Se o utilizador não inserir ou inserir um código de segurança errado, será retornada uma mensagem adequada (Figura 21 e 22).



Figura 20 – Código de segurança (CAPTCHA) utilizado no sistema *ThermInfo* em todas as funcionalidades disponíveis publicamente.

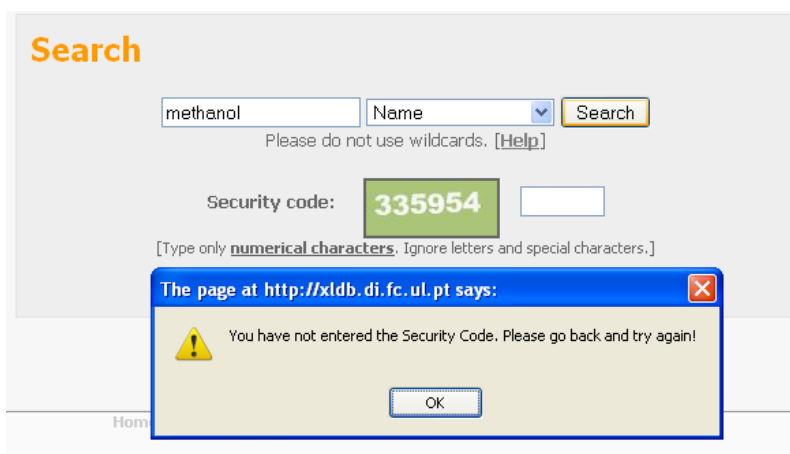


Figura 21 – Aviso indicador de que o código de segurança não foi digitado.

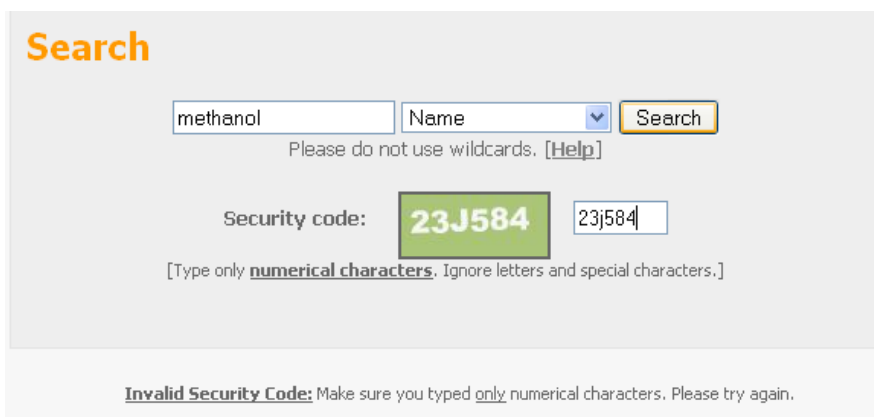


Figura 22 – Mensagem de erro quando o código inserido não está correcto.

• Pesquisa de compostos

Para pesquisar um composto no sistema *ThermInfo*, o utilizador especifica, num formulário, o(s) termo(s) e os parâmetros de pesquisa que pretende assim como o código de segurança. Estes dados são recebidos pelo servidor que por sua vez realiza um pré-processamento aos mesmos, envia uma *query* à base de dados e devolve o seu resultado ao utilizador (Figura 23).

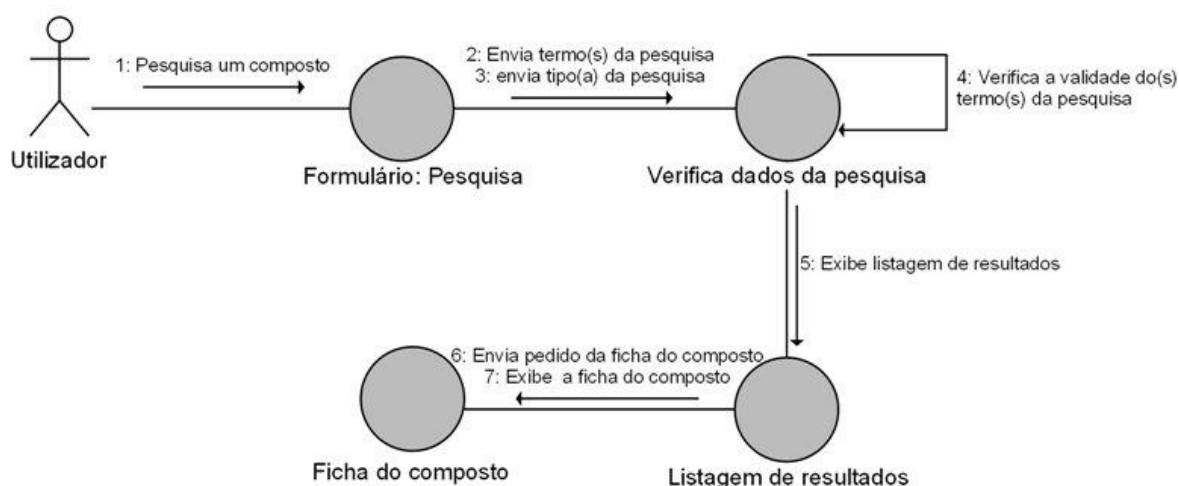


Figura 23 – Diagrama de Colaboração do Caso de Uso: Pesquisar um Composto. Representa os eventos que o actor utilizador realiza para visualizar a ficha do composto, descrevendo as entradas e saídas de dados e as mensagens do sistema para cada evento.

Para este efeito, podem ser realizados dois tipos de pesquisa:

- 1. Pesquisa simples**, baseada apenas num parâmetro que poderá ser: Nome do composto, fórmula molecular, ID molecular, CASRN ou SMILES (Figura 24). Se o termo da pesquisa não for inserido, o utilizador é alertado com uma mensagem que o avisa do sucedido (Figura 25).

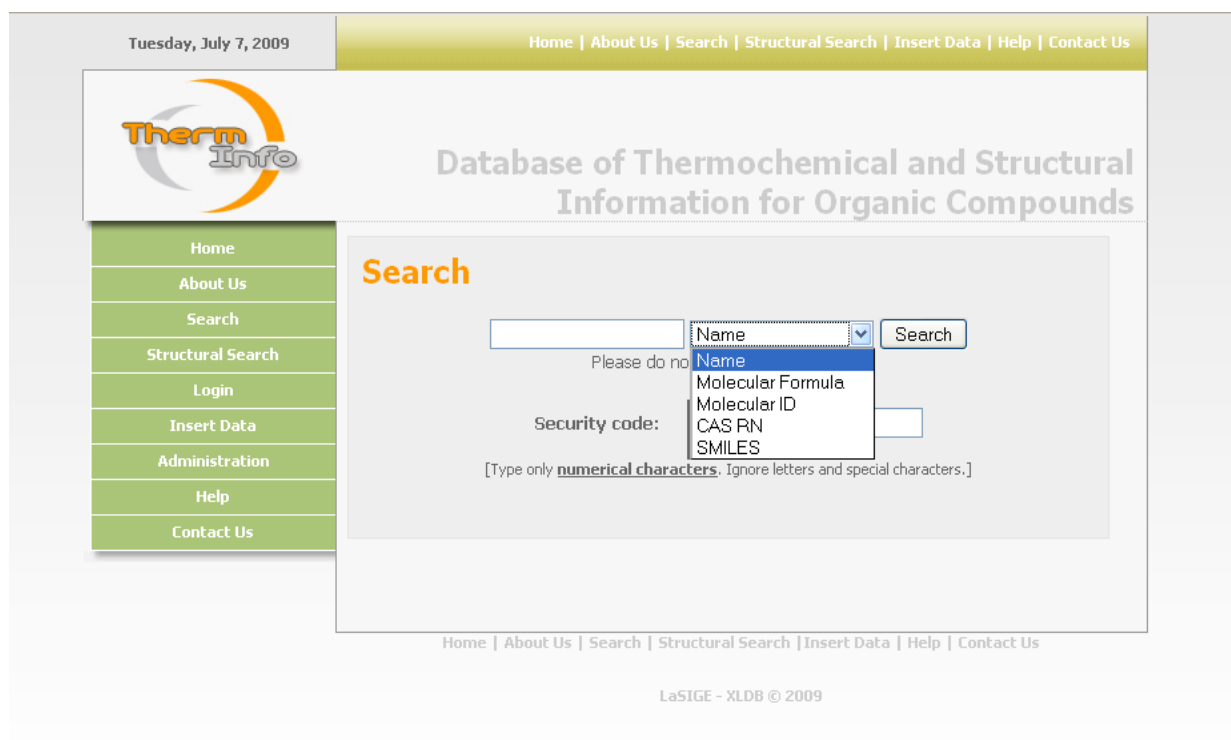


Figura 24 – Página de ‘Pesquisa Simples’, na qual pode ser especificado um termo de pesquisa e o tipo de pesquisa a efectuar.

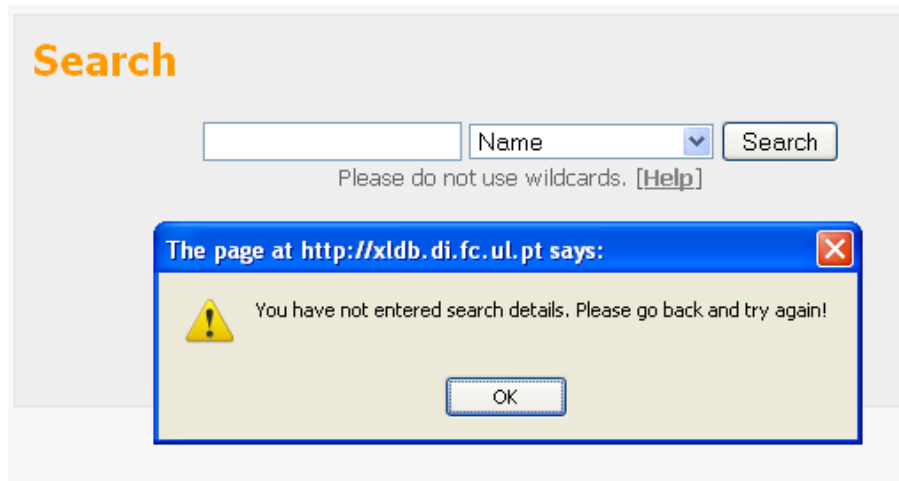


Figura 25 – Mensagem de aviso de que o utilizador não inseriu o termo da pesquisa.

Na pesquisa por nome, o utilizador deverá inserir o nome do composto de acordo com as regras IUPAC. Não será necessário utilizar *wildcards*, já que a comparação se baseia na procura do termo inserido nos nomes contidos na base de dados, utilizando a função LIKE ‘%termo%’ (‘%’ representa zero ou mais caracteres)

do MySQL. Em termos de interface, quando o tipo de pesquisa “Name” é seleccionado, é apresentada uma instrução de pesquisa que indica ao utilizador que não deve utilizar *wildcards*, assim como uma ligação à secção de ajuda deste tipo de pesquisa (Figura 26).

The image shows a search interface with a text input field, a dropdown menu set to 'Name', and a 'Search' button. Below the input field, there is a message: 'Please do not use wildcards. [Help]'.

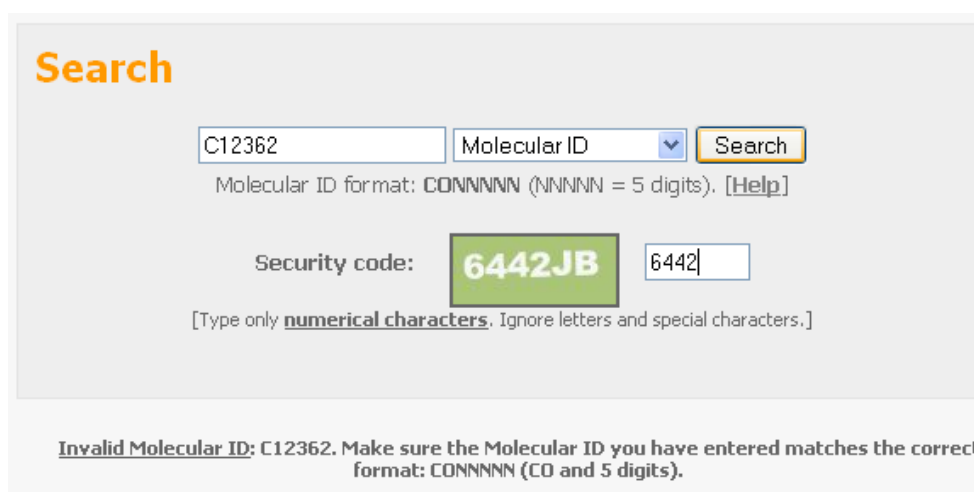
Figura 26 – Formulário da ‘Pesquisa Simples’ por nome (particularidade da Figura 24).

Na pesquisa por fórmula molecular o utilizador deverá inserir a fórmula que pretende encontrar com os elementos químicos na ordem CHXNOS. Poderá utilizar o wildcard ‘?’, para substituir o número de átomos de um determinado elemento (por exemplo, a pesquisa por ‘C?H11’ irá retornar todos os compostos que têm ? = [2, 9]). Mais uma vez o modo de pesquisa não é estrito, retornando as fórmulas existentes na base de dados que contêm o termo da pesquisa. Em termos de interface, quando o tipo de pesquisa “Molecular Formula” é seleccionado são apresentadas duas instruções de pesquisa que indicam ao utilizador a ordem em que os elementos se devem encontrar e a possibilidade de utilizar o *wildcard* ‘?’, assim como uma ligação à secção de ajuda deste tipo de pesquisa (Figura 27).

The image shows a search interface with a text input field, a dropdown menu set to 'Molecular Formula', and a 'Search' button. Below the input field, there is a message: 'Please type Molecular Formula with atoms in CHXNOS (X = halogen) order. Wildcard: ? represents one character [Help]'.

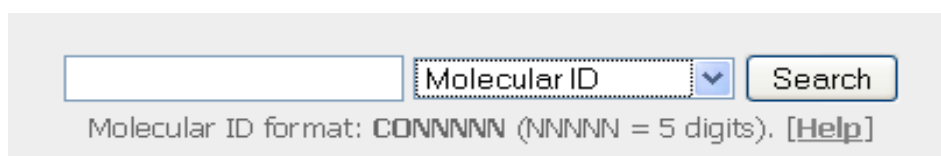
Figura 27 – Formulário da ‘Pesquisa Simples’ por fórmula molecular (particularidade da Figura 24).

Na pesquisa por ID molecular o utilizador deverá inserir ID do composto de acordo com o seu formato: CONNNNN (N = dígito). Quando a pesquisa é submetida, antes da procura do termo inserido na base de dados, é feita uma verificação a este formato. Caso o formato não seja o correcto, será retornada uma mensagem que informa o utilizador (Figura 28). Em termos de interface, quando o tipo de pesquisa “Molecular ID” é seleccionado, é apresentada uma instrução de pesquisa que indica o formato do ID, assim como uma ligação à secção de ajuda deste tipo de pesquisa (Figura 29).



The screenshot shows a search interface with the title "Search" in orange. It features a text input field containing "C12362", a dropdown menu set to "Molecular ID", and a "Search" button. Below the input field, a message reads: "Molecular ID format: CONNNNN (NNNNN = 5 digits). [Help]". A "Security code" section displays "6442JB" in a green box, with a corresponding input field containing "6442". A note below states: "[Type only numerical characters. Ignore letters and special characters.]". At the bottom, an error message is displayed: "Invalid Molecular ID: C12362. Make sure the Molecular ID you have entered matches the correct format: CONNNNN (CO and 5 digits)."

Figura 28 – Mensagem de erro retornada numa pesquisa com o formato do ID molecular errado.

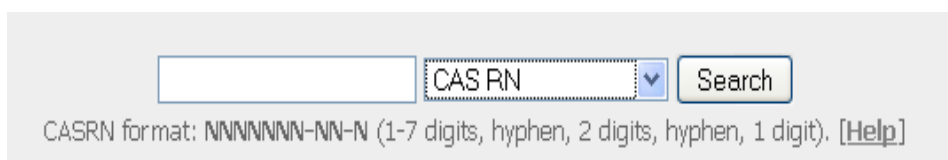


The screenshot shows a search interface with an empty text input field, a dropdown menu set to "Molecular ID", and a "Search" button. Below the input field, a message reads: "Molecular ID format: CONNNNN (NNNNN = 5 digits). [Help]".

Figura 29 – Formulário da ‘Pesquisa Simples’ por ID molecular (particularidade da Figura 24).

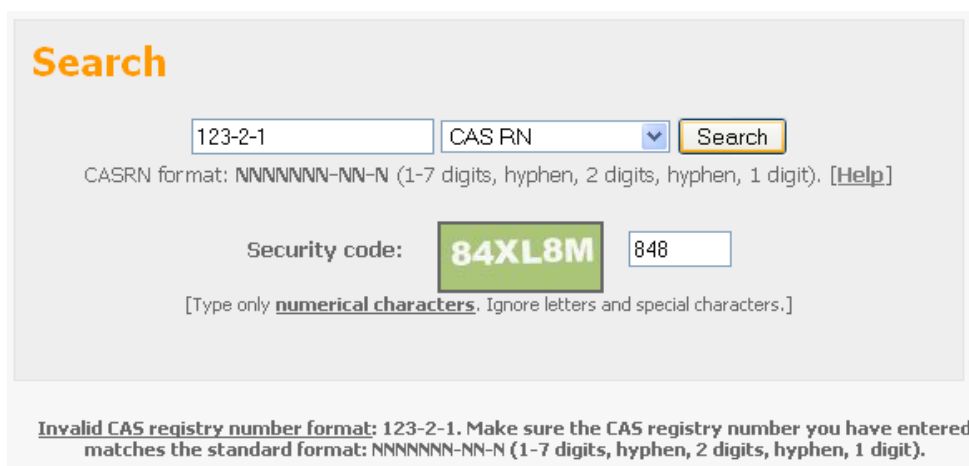
Na pesquisa por CASRN o utilizador deverá inserir o CASRN do composto de acordo com o seu formato: NNNNNNN-NN-N (1-7 dígitos, hífen, 2 dígitos, hífen, 1 dígito). Quando a pesquisa é submetida, antes da procura do termo inserido na base de dados, é feita não só uma verificação a este formato como ao dígito de

controlo (Capítulo 2). Caso o formato não seja o correcto ou não se verifique o dígito de controlo correcto, será retornada uma mensagem que informa o utilizador (Figura 31 e 32). Em termos de interface, quando o tipo de pesquisa “CASRN” é seleccionado, é apresentada uma instrução de pesquisa que indica o formato do CASRN, assim como uma ligação à secção de ajuda deste tipo de pesquisa (Figura 30).



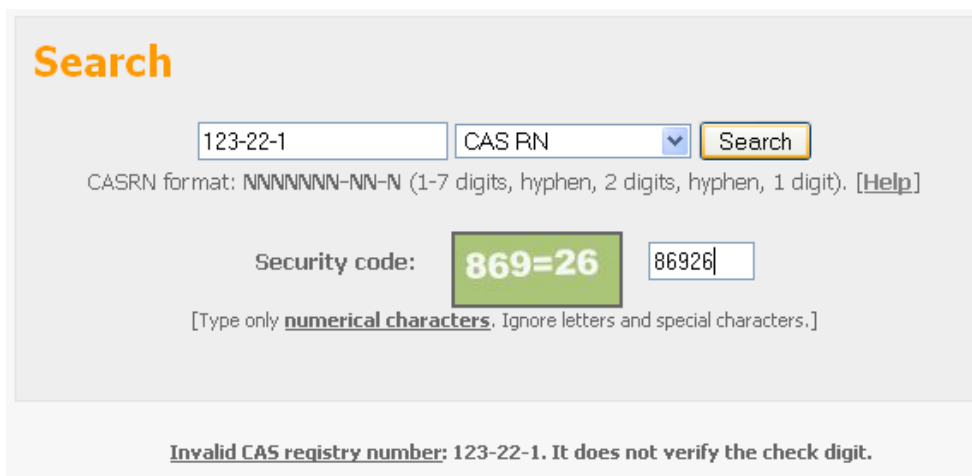
The image shows a search interface with a text input field, a dropdown menu set to 'CAS RN', and a 'Search' button. Below the input field, the text reads: 'CASRN format: NNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit). [Help]'

Figura 30 - Formulário da ‘Pesquisa Simples’ por CASRN [particularidade da Figura 24].



The image shows a search interface with the title 'Search' in orange. The search input field contains '123-2-1', the dropdown menu is set to 'CAS RN', and the 'Search' button is highlighted. Below the input field, the text reads: 'CASRN format: NNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit). [Help]'. Below this, there is a 'Security code:' label, a green box containing '84XL8M', and a text input field containing '848'. Below the security code, the text reads: '[Type only **numerical characters**. Ignore letters and special characters.]'. At the bottom, the error message reads: 'Invalid CAS registry number format: 123-2-1. Make sure the CAS registry number you have entered matches the standard format: NNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit).'

Figura 31 – Mensagem de erro retornada numa pesquisa com o formato do CASRN errado.



The screenshot shows a search interface with the following elements:

- Search** (orange heading)
- Input field: 123-22-1
- Dropdown menu: CAS RN
- Search button: Search
- Text: CASRN format: NNNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit). [Help]
- Security code: 869=26 (highlighted in green)
- Input field: 86926
- Text: [Type only **numerical characters**. Ignore letters and special characters.]
- Error message: **Invalid CAS registry number: 123-22-1. It does not verify the check digit.**

Figura 32 – Mensagem de erro retornada numa pesquisa com um CASRN errado. A verificação da validade do CASRN é feita baseada no dígito de controlo.

Na pesquisa por SMILES o utilizador deverá inserir o SMILES do composto tendo em conta que existe uma diferença entre letras maiúsculas e minúsculas, devido às regras de concepção de um SMILES [14]. Para realizar esta pesquisa *case sensitive* é utilizado o *COLLATE latin1_bin* do MySQL. Em termos de interface, quando o tipo de pesquisa “SMILES” é seleccionado, é apresentada uma instrução de pesquisa que indica que esta pesquisa é *case sensitive*, assim como uma ligação à secção de ajuda deste tipo de pesquisa (Figura 33).



The screenshot shows a search interface with the following elements:

- Input field (empty)
- Dropdown menu: SMILES
- Search button: Search
- Text: SMILES: the search is **case sensitive**. [Help]

Figura 33 – Formulário da ‘Pesquisa Simples’ por SMILES (particularidade da Figura 24).

Todos os tipos de pesquisa disponíveis têm hiperligações para a respectiva secção de *Help*. Os resultados da pesquisa são apresentados ao utilizador sob a forma de listagem, ordenada por relevância, isto é, de acordo com a distância do termo de pesquisa ao início da *string* dos termos encontrados. Esta listagem inclui a

informação relativa ao nome do composto, fórmula molecular, ID molecular, CASRN e SMILES para cada dos compostos encontrados. São apresentados no máximo 100 compostos, contudo, o número real de compostos que satisfazem as condições da pesquisa é apresentado (Figura 34).

The screenshot displays the search interface of the ThermInfo system. On the left is a vertical navigation menu with options: Home, About Us, Search, Structural Search, Login, Insert Data, Administration, Help, and Contact Us. The main content area is titled 'Search [Help]' and features a search input field with a dropdown menu set to 'Name' and a 'Search' button. Below the search field is a note: 'Please do not use wildcards. [Help]'. A 'Security code' field contains the value '8TB12F' and is followed by an empty input box. A note below the security code reads: '[Type only numerical characters. Ignore letters and special characters.]'. The search results section indicates 'You are searching for: methanol ... Number of compounds found: 16'. It lists five results, each with a 'View' button:

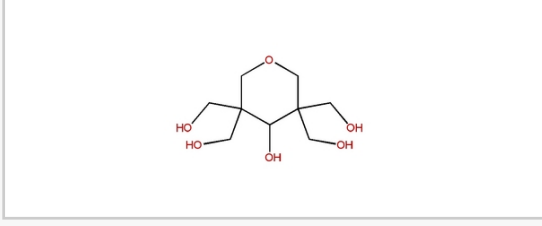
- Molecular ID:** C000164
Compound Name: Methanol
Molecular Formula: CH₄O
CAS registry number: 67-56-1
SMILES: CO
More info: [View](#)
- Molecular ID:** C001398
Compound Name: 2-Furanmethanol
Molecular Formula: C₅H₆O₂
CAS registry number: 98-00-0
SMILES: OCC1CCCC1
More info: [View](#)
- Molecular ID:** C002022
Compound Name: Benzenemethanol
Molecular Formula: C₇H₈O
CAS registry number: 100-51-6
SMILES: OCC1CCCC1
More info: [View](#)
- Molecular ID:** C002024
Compound Name: Biphenylmethanol
Molecular Formula: C₁₃H₁₂O
CAS registry number: 41376-19-6
SMILES: OCC1CCCC1C2CCCC2
More info: [View](#)
- Molecular ID:** C000307
Compound Name: Dioxybismethanol

Figura 34 – Parte da listagem de resultados obtida para a pesquisa de compostos com o nome *methanol*.

A totalidade da informação sobre cada um dos compostos pode ser visualizada clicando em *View*, abrindo-se uma nova janela (Figura 35) de forma a manter visível a página com a listagem de resultados.

Tuesday, July 21, 2009

4-Hydroxy-2H-pyran-3,3,5,5,(4H,6H)-tetramethanol



Structural Data

Molecular ID:	CO01689	Compound:	4-Hydroxy-2H-pyran-3,3,5,5,(4H,6H)-tetramethanol
CASRN:	4744-47-2	Molecular Formula:	C ₉ H ₁₈ O ₆
Molecular Weight:	222.24	Physical State:	Crystal
SMILES:	OCC1(CO)COC(CO)(CO)C1O		
Unique SMILES:	[CH2]1[C]([C]([C]([CH2][O])1)[CH2][O])][CH2][O])1([C][O])1[CH2][O]		

Thermochemical Data

Standard Molar Enthalpy of Formation - Crystalline Phase [kJ/mol]:	-1267.7	Error:	+/- 5.0
Standard Molar Enthalpy of Formation - Liquid Phase [kJ/mol]:	n.a.	Error:	n.a.
Standard Molar Enthalpy of Formation - Gas Phase [kJ/mol]:	n.a.	Error:	n.a.
Standard Molar Enthalpy of Phase Change [kJ/mol]:			
- Solid - Liquid	n.a.	Error:	n.a.
- Liquid - Gas	n.a.	Error:	n.a.
- Solid - Gas	n.a.	Error:	n.a.
Observations:	n.a.		

References

- J. B. Pedley
Thermochemical Data and Structures of Organic Compounds, 1, 1994, 571

Figura 35 – Ficha de um composto contendo toda a informação relativa às propriedades estruturais e termoquímicas assim como a(s) respectiva(s) referência(s).

- A Pesquisa estrutural**, permite especificar em maior detalhe a estrutura do composto, de forma a limitar o número de resultados a obter. Para isso é disponibilizado um formulário com os seguintes campos: nome do composto, CASRN, fórmula molecular, peso molecular (sob a forma de intervalo: maior, menor ou igual ao valor especificado), estado físico, SMILES, USMILES, classe, subclasse, família e características (Figura 36).

Nenhum dos campos é de preenchimento obrigatório, isto é, o utilizador poderá fazer as combinações de campos que entender. De

forma a facilitar e a diminuir o número de erros na entrada dos termos a pesquisar, o estado físico, a classe, a subclasse e a família aparecem sobre a forma de *drop down menus* com todos os valores contidos na base de dados para esse atributo, e as características sobre a forma de *checkboxes* (podendo ser seleccionadas várias características em simultâneo). Os restantes campos de preenchimento manual, pelo utilizador, obedecem às mesmas regras e processamento enunciados na pesquisa simples.

Figura 36 – Formulário de ‘Pesquisa Estrutural’, no qual podem ser especificados os termos de pesquisa.

Como este tipo de pesquisa também está disponível publicamente, pelas razões enunciadas anteriormente, o utilizador terá de digitar o código de segurança. A listagem de resultados obtida é idêntica à da pesquisa simples, exceptuando a ordem pela qual os compostos são dispostos.

• Inserção de compostos

A funcionalidade que permite aumentar o conjunto de dados disponibilizado pelo sistema *ThermInfo* está limitado a utilizadores registados, de forma a evitar inserções inadequadas de dados.

Para suportar o aumento da comunidade que deposita dados, existe uma funcionalidade de registo no sistema. Esta apresenta um formulário que requer o preenchimento dos seguintes dados: nome, e-mail e a instituição a que o utilizador está afiliado (Figura 37). O formato do e-mail é verificado após a submissão, se este for válido o utilizador é informado de que o seu registo foi efectuado com sucesso. No entanto, este tem de esperar pela validação da conta por um administrador (que é informado da existência de um novo pedido de registo por e-mail). Após a validação da conta, o utilizador recebe um e-mail com os dados para entrar no sistema: e-mail e uma palavra-chave gerada aleatoriamente (Figura 38).

Saturday July 25, 2009 Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us

ThermInfo Register!

Home
About Us
Search
Structural Search
Login
Insert Data
Administration
Help
Contact Us

Register an account on ThermInfo

On this page you can register an account for inserting new compounds on ThermInfo. Your name, e-mail address and Institution are required. Afterwards, you have to wait for approval by an administrator when that happens you will receive an e-mail containing the password to Login. Thus if you fail to provide your e-mail address, you will not be able to use this service.

Name:

E-mail address:

Institution:

Security code: **5E7K49**

[Type only numerical characters. Ignore letters and special characters.]

Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us

Figura 37 – Formulário de registo no sistema *ThermInfo*.

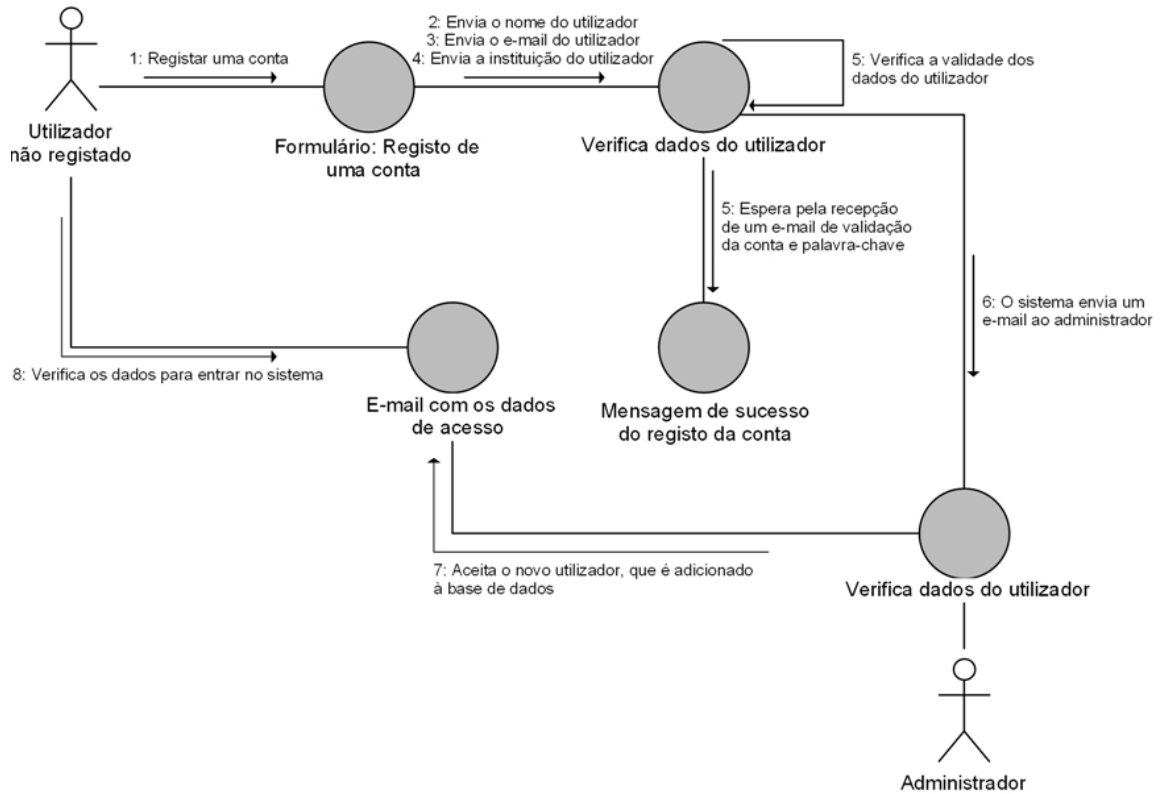


Figura 38 – Diagrama de Colaboração do Caso de Uso: Registrar uma conta. Representa os eventos que o actor utilizador realiza para registar uma conta no sistema, descrevendo as entradas e saídas de dados e as mensagens do sistema para cada evento.

Para inserir novos dados o utilizador terá de se autenticar no sistema, através do formulário apresentado na Figura 39, com os dados que recebeu no seu e-mail.

Log in to ThermInfo!

Your e-mail address and your password are required. Note that password is case sensitive. You do not have an account? [Register Here](#). [Forgot your password? Click Here](#)

E-mail address:

Password:

Security code: 71)%6{

[Type only numerical characters. Ignore letters and special characters.]

Figura 39 – Formulário que permite a entrada no sistema para inserção de novos compostos orgânicos.

A inserção de dados é realizada a partir do formulário apresentado na Figura 40, que disponibiliza diversos campos para as propriedades estruturais e termoquímicas, e para as respectivas referências bibliográficas, descritas no Capítulo 2. Os campos de preenchimento obrigatório são o nome do composto e pelo menos uma referência bibliográfica. Os campos de preenchimento: numéricos (as entalpias e erros associados, o peso molecular e o ano de publicação) são verificados se efectivamente só contêm números; CASRN é verificado o formato e o dígito controlo; estado físico, classe, subclasse e família podem ser seleccionadas de uma lista para evitar erros de digitação, no entanto, existe a opção de adicionar uma família não existente na lista; a estrutura química pode ser adicionada sob a forma de imagem JPEG; as características da molécula são seleccionadas em *checkboxes*; e as referências bibliográficas podem ser estendidas até seis.

Figura 40 – Formulário que permite a inserção de novos compostos orgânicos no sistema.

Se os dados forem validados, é atribuído ao novo composto um ID molecular, ficando em espera pelo processo de validação a executar por um administrador (Figura 41). Apenas os administradores podem inserir directamente dados na base de dados sem que passem pelo processo de validação. Para isso, existe uma funcionalidade semelhante à descrita no menu de administração, em que a única diferença é inserir directamente o composto na base de dados.

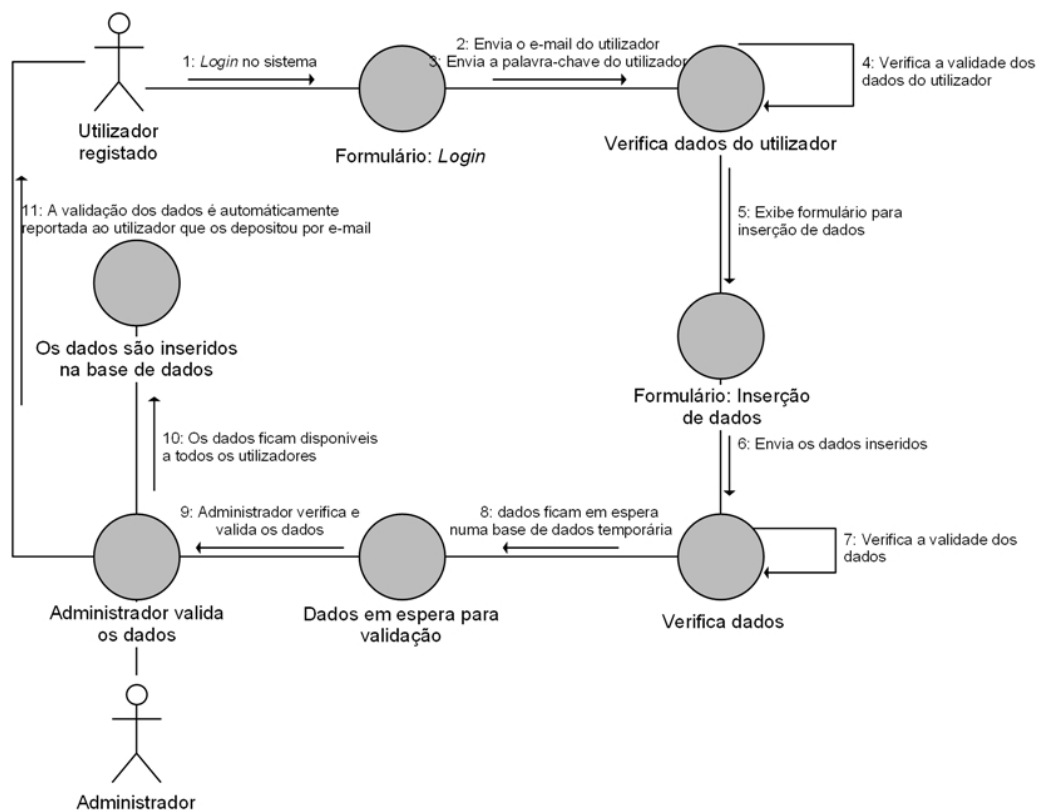


Figura 41 – Diagrama de Colaboração do Caso de Uso: Inserir um novo composto orgânico. Representa os eventos que o actor utilizador realiza para inserir um novo composto na base de dados, descrevendo as entradas e saídas de dados e as mensagens do sistema para cada evento.

- **Validar dados inseridos por utilizadores**

Apesar dos dados serem automaticamente validados aquando da inserção pelos utilizadores (em termos de formato, tipo de dados, dígito controlo, entre outros), um dos principais objectivos do sistema *ThermInfo* é conter

dados de qualidade. Para isso, todos os dados inseridos têm de ser validados por um administrador experiente em termoquímica. Esta validação consiste numa avaliação crítica da consistência dos dados através de análise bibliográfica, por comparação com valores experimentais de compostos da mesma família, por aplicação de métodos empíricos de aditividade de propriedades [29-30] e/ou por utilização de métodos de química computacional [31]. Uma vez que na inserção dos dados pelos utilizadores os únicos campos obrigatórios são o nome e uma referência bibliográfica, o administrador para poder validar os dados tem de garantir a inserção de pelo menos todas as propriedades estruturais do composto.

Esta funcionalidade ainda não se encontra completamente implementada, no entanto, o plano é mostrar ao administrador a listagem dos nomes dos compostos inseridos pelos utilizadores e que se encontram em espera para validação. Este poderá seleccionar um composto, sendo-lhe apresentado um formulário idêntico ao da inserção de dados, mas os campos submetidos pelo utilizador aquando da inserção são automaticamente carregados. Assim, o administrador poderá adicionar, remover ou actualizar esses dados e posteriormente guardá-los (se o processo de validação ainda não estiver completo), apagá-los (se considera que os dados não têm interesse ou não são consistentes com a bibliografia apresentada), ou então inseri-los na base de dados. Quando o processo de validação estiver concluído, a decisão do administrador é automaticamente reportada por e-mail ao utilizador que depositou os dados.

- **Actualizar/apagar dados dos compostos**

As funcionalidades para remover e actualizar os dados dos compostos orgânicos existentes na base de dados estão disponíveis apenas para administradores. Para encontrar o composto a actualizar ou remover, o administrador terá de o pesquisar utilizando o ID molecular numa caixa de pesquisa (Figura 42). Como este se trata de um identificador único, esta

pesquisa só retorna um composto (ou nenhum, se este não existir na base de dados).

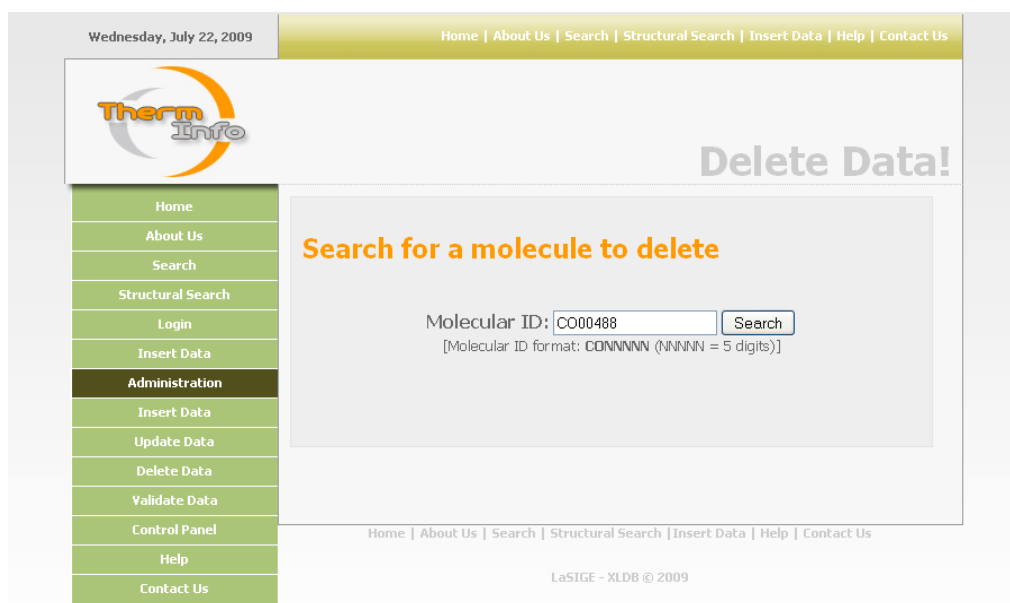


Figura 42 – Caixa de pesquisa por ID molecular de um composto a apagar.

Se a funcionalidade pretendida for a actualização dos dados do composto, é apresentado um formulário idêntico ao da funcionalidade de inserção de dados. No entanto, os dados disponíveis para esse composto são carregados nos respectivos campos do formulário. Assim, podem ser selectivamente alterados, adicionados ou removidos os valores dos campos (Figura 43).

Thursday, July 23, 2009 Home | About Us | Search | Insert Data | Update Data | Help | Contact Us

Update Data!

Search

Molecular ID: Search

You are searching for: C000500 :: Number of compounds found: 1

[Please Check the information below and change any incorrect/incomplete data]

Structural Data

Compound: CASRN:
 Molecular Formula: Molecular Weight:
 Chemical structure: Browse... Physical State:
 SMILES:
 Unique SMILES:
 Class:
 Sub-Class:
 Family:

Characteristic:

Alkane Alkene Alkyne Arene Alcohol Ether
 Peroxide Aldehyde Ketone Carboxylic Acid Ester Amine
 Hydrazine Imine Nitrile/Isonitrile NOx Amide Thiol
 Thioether Polysulphide Thiocarbonyl SOx Halogen Radical
 Charges Ionic Solvation Polymer

Thermochemical Data

Standard Molar Enthalpy of Formation - Crystalline Phase [kJ/mol]: Error: +/-
 Standard Molar Enthalpy of Formation - Liquid Phase [kJ/mol]: Error: +/-3.3
 Standard Molar Enthalpy of Formation - Gas Phase [kJ/mol]: Error: +/-3.6
 Standard Molar Enthalpy of Phase Change [kJ/mol]: Error: +/-1.4

Observations:

References

Author:
 Journal:
 Year: Volume: Page:
 More References:

Author:
 Journal:
 Year: Volume: Page:

Home | About Us | Search | Insert Data | Update Data | Help | Contact Us

LaSIGE - XLDB © 2009

Figura 43 – Formulário de actualização dos dados dos compostos existentes na base de dados do sistema ThermInfo.

Se a funcionalidade pretendida for a remoção do composto, é apresentada a ficha que especifica toda a informação contida na base de dados, relativo ao mesmo (Figura 44). Após visualizar os dados do composto o administrador poderá optar por removê-lo carregando em “Delete”. Ser-lhe-á solicitada a confirmação de que realmente quer realizar esta operação

(Figura 44 a)) e em caso afirmativo o sucesso da mesma é explicitado com uma mensagem adequada.

Wednesday, July 22, 2009 Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us

ThermInfo

Home
 About Us
 Search
 Structural Search
 Login
 Insert Data
 Administration
 Help
 Contact Us

Delete Data!

Search for a molecule to delete

Molecular ID:
[Molecular ID format: C000000 (NNNNN = 5 digits)]

(Z)-3-Pentenoic acid ethyl ester

CCOC(=O)C=CC

Structural Data

Molecular ID:	C000488	Compound:	(Z)-3-Pentenoic acid ethyl ester
CASRN:	27829-70-5	Molecular Formula:	C ₇ H ₁₂ O ₂
Molecular Weight:	128.17	Physical State:	Liquid
SMILES:	CC=C(CC(=O)O)O		
Unique SMILES:	[C]([CH2][CH]=[CH])([CH3])(=O)[O][CH2][CH3]		

Thermochemical Data

Standard Molar Enthalpy of Formation - Crystalline Phase [kJ/mol]:	n.a.	Error:	n.a.
Standard Molar Enthalpy of Formation - Liquid Phase [kJ/mol]:	-432.4	Error:	+/- 3.3
Standard Molar Enthalpy of Formation - Gas Phase [kJ/mol]:	-387.6	Error:	+/- 3.5
Standard Molar Enthalpy of Phase Change [kJ/mol]:	44.8	Error:	+/- 1.2
Observations:	n.a.		

References

1. J. B. Pedley
Thermochemical Data and Structures of Organic Compounds, 1, 1994, 571

The page at <http://xldb.di.fc.ul.pt> says:
 Do you really want to DELETE the compound?
 a)

Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us

LaSIGE - XLDB © 2009

Figura 44 – Informação sobre as propriedades estruturais e termoquímicas de um composto a remover da base de dados. a) Pedido de confirmação da operação de remoção do composto da base de dados.

É importante salientar que em ambas as funcionalidades (actualizar/apagar) estes dados não são apagados do sistema, mas sim movidos para uma base de dados paralela, não disponível para as pesquisas dos utilizadores, que alberga dados desactualizados. Antes de cada operação é também realizado um *backup* automático do conteúdo da base de dados (utilizando a função *mysqldump* do MySQL), isto é, é criado um ficheiro com todos os comandos SQL que permitem restaurar a base de dados caso seja necessário.

- **Painel de Controlo**

A funcionalidade “Painel de Controlo” está disponível apenas para os administradores do sistema *ThermInfo* e permite monitorizar os utilizadores, a evolução da base de dados e a utilização das funcionalidades de pesquisa.

Relativamente aos utilizadores do sistema, esta é a funcionalidade que permite adicionar ou eliminar utilizadores que se registaram. Assim, é apresentado um quadro que contém o nome do utilizador, o seu e-mail e a instituição a que pertence (Figura 45 a)). Se o administrador aceitar o utilizador, este é adicionado à base de dados e recebe um e-mail com os dados necessários para o *login* (o e-mail e uma palavra-chave gerada aleatoriamente) e assim posteriormente ter acesso à funcionalidade de inserção de dados. Todos os utilizadores que têm acesso à inserção de dados são apresentados ao administrador num quadro idêntico ao descrito anteriormente (Figura 45 b)). O administrador poderá em qualquer momento remover o acesso a qualquer um destes utilizadores registados.

Wednesday, July 22, 2009 Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us

ThermInfo

Control Panel!

Add new users to the Database! a)

ID	Name	E-mail	Institution	Add?	Delete?
6	Miguel Teixeira	jmiguel@gmail.com	IST	<input type="checkbox"/>	<input type="checkbox"/>
5	Ana Teixeira	anaisabel_teixeira@hotmail.com	FCUL	<input type="checkbox"/>	<input type="checkbox"/>

Add/Delete Users

Database Users: b)

ID	Name	E-mail	Institution	Delete?
1	Ana Teixeira	analnoteixeira@gmail.com	FCUL	<input type="checkbox"/>
2	Rui Santos	rjcs@fc.ul.pt	FCUL	<input type="checkbox"/>
3	Ana	ateixeira@lasige.di.fc.ul.pt	FCUL	<input type="checkbox"/>

Add/Delete Users

Figura 45 – Funcionalidade “Painel de Controlo” para a monitorização dos utilizadores que podem inserir dados de compostos orgânicos. a) Utilizadores que se registaram e se encontram em lista de espera para inserção na base de dados. b) Utilizadores registados com acesso à funcionalidade de inserção de dados.

De forma a monitorizar a evolução da base de dados, o administrador tem a possibilidade de visualizar um quadro que apresenta no momento qual é o número total de compostos existentes na base de dados, assim como o número de propriedades estruturais e termoquímicas existentes (Figura 46).

Database Statistics:

Total	SMILES	Chemical Structure	Classes	Sub-Classes	Families	Enthalpies			
						Crys	Liq	Gas	Liq-Gas
2956	2956	2956	6	12	489	1383	1481	1740	1648

Figura 46 – Estatísticas da base de dados: número total de compostos (sublinhado a cor de laranja), número de SMILES, estruturas químicas e entalpias presentes, assim como o número de classes, subclasses e famílias diferentes existentes na base de dados.

Para monitorizar o uso das funcionalidades de pesquisa, é apresentado um quadro com as estatísticas para o número de pesquisas realizadas (Figura 47 a)). Este quadro apresenta o número total de pesquisas realizadas, assim como os subtotais para cada ano (linhas), para os doze meses do ano (colunas) e para o cruzamento destes dois factores (célula de cruzamento dos dois factores). É também apresentada uma listagem dos cem últimos endereços *Internet Protocol* (IP) dos utilizadores que realizaram pesquisas no sistema (Figura 47 b)).

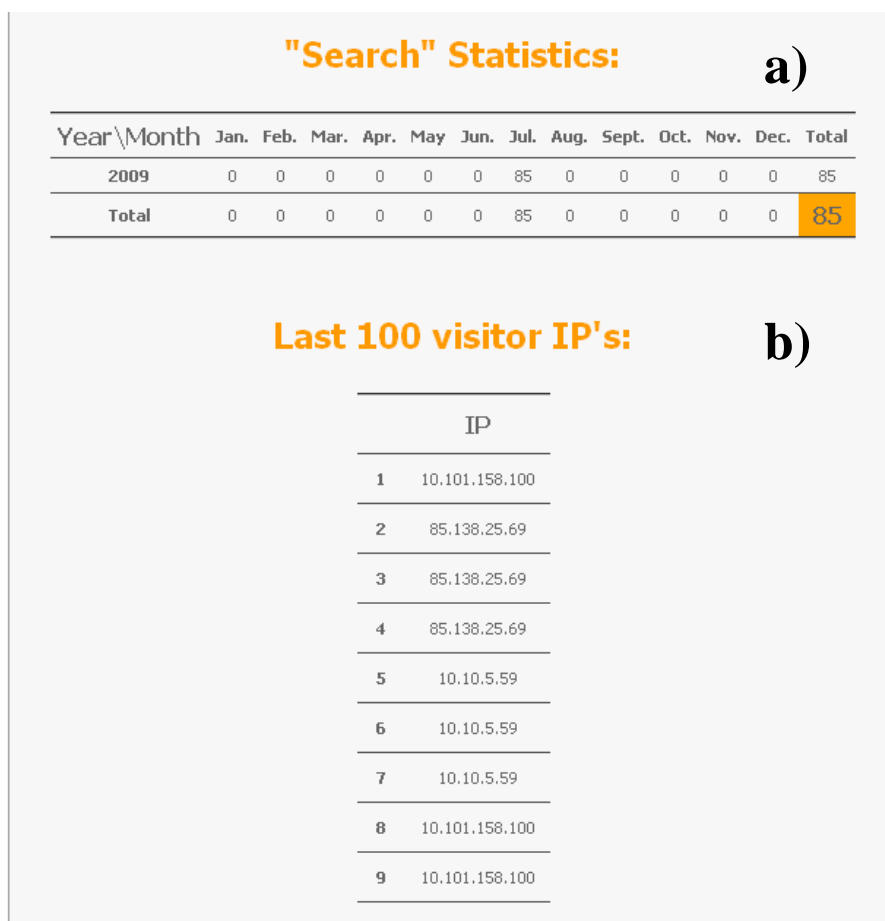


Figura 47 – Interface que permite a monitorização do uso das funcionalidades de pesquisa. a) Estatísticas do número de pesquisas realizadas e sua localização temporal. b) Listagem de endereços IP dos cem últimos utilizadores que realizaram pesquisas.

- **Ajuda**

Para além das informações de utilização que vão aparecendo nas diversas funcionalidades e que já foram explicitadas anteriormente, o sistema *ThermInfo* contém um manual de instruções (“*Help*”). Este manual contém informação sobre o que são e como se utiliza correctamente cada uma das funcionalidades disponíveis para os utilizadores. Para além da descrição textual, existem tutoriais em vídeo e imagens alusivas a essa componente escrita (Figura 48 e 49). Todas as funcionalidades que se encontram mencionadas neste manual contêm hiperligações directas à secção correspondente.

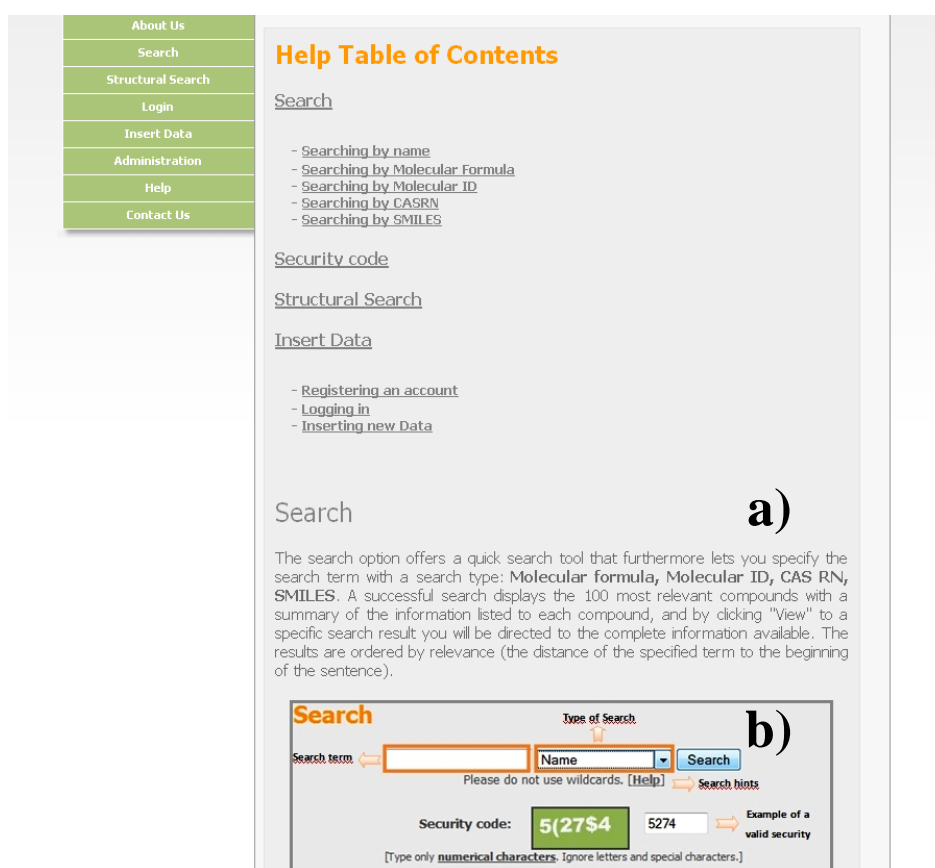


Figura 48 – Conteúdos do manual de instruções do sistema *ThermInfo*. a) Descrição textual do que é a funcionalidade e quais os passos a seguir para realizar uma pesquisa. b) Imagem alusiva à descrição em a).

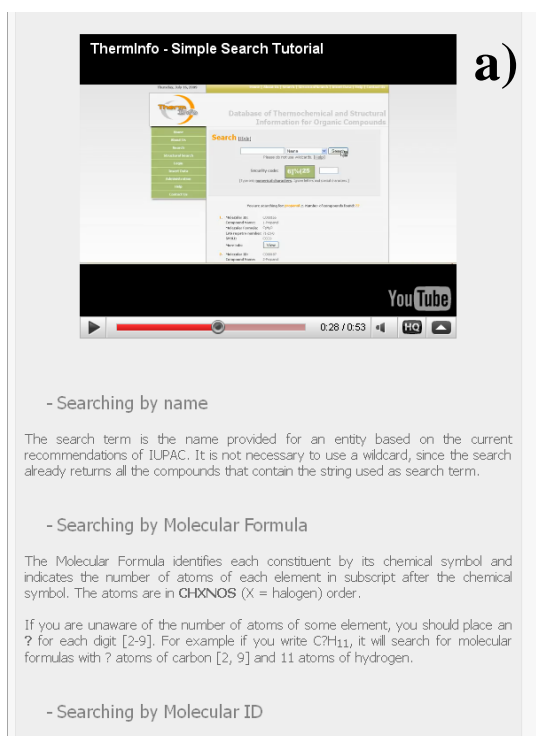


Figura 49 – Conteúdos do manual de instruções do sistema *ThermInfo*. a) Tutoriais em vídeo alusivos à descrição textual de como são utilizadas as funcionalidades.

- **Comunicação dos utilizadores com os Administradores**

Esta funcionalidade permite aos utilizadores,, a partir da interface, o envio de mensagens aos administradores do sistema *ThermInfo* (Figura 50). Para isso o utilizador terá de preencher um formulário que contém campos para: nome, e-mail, assunto da mensagem (*drop down menu* com as seguintes opções: “Sugestão”, “Comentário”, “Dúvida”, “Reportar dados errados/desactualizados”, “Reportar dados incompletos”, “Outro”) e a mensagem propriamente dita. Como se trata de uma funcionalidade acessível publicamente, o utilizador terá de inserir o código de segurança. Antes da mensagem ser submetida, é verificado o preenchimento de todos estes campos e a validade do formato do e-mail. Caso a submissão da mensagem falhe devido a algum erro nos pontos anteriormente referidos, o utilizador visualizará uma mensagem a esse respeito, caso contrário o sucesso do envio

da mensagem é confirmado. A mensagem submetida com sucesso é enviada para o e-mail da administração do sistema *ThermInfo*.

The image shows a web interface for sending feedback. On the left is a vertical navigation menu with items: Home, About Us, Search, Structural Search, Login, Insert Data, Administration, Help, and Contact Us (highlighted). The main content area is titled 'Leave us a comment!' and contains a form with the following fields: Name (text input), E-mail address (text input), Subject (dropdown menu with options: Suggestion, Comment, Question, Erroneous Data, Incomplete Data, Other), and Message (text area). Below the form is a security code '164879' in a green box, followed by a small input field and two buttons: 'Submit Message' and 'Erase Message'. A note below the security code reads: '[Type only numerical characters. Ignore letters and special characters.]'. At the bottom of the page, there is a footer with navigation links: Home | About Us | Search | Structural Search | Insert Data | Help | Contact Us, and a copyright notice: LaSIGE - XLDB © 2009.

Figura 50 – Formulário para o envio de mensagens aos administradores do *ThermInfo*.

Esta funcionalidade é muito importante, não só para receber questões/comentários dos utilizadores, mas também porque permite que a comunidade científica participe no processo de avaliação da qualidade dos dados disponibilizados pelo sistema *ThermInfo*.

3.5.3. Avaliação de Usabilidade

As várias técnicas de avaliação de usabilidade, assim como os resultados obtidos serão apresentados no Capítulo seguinte. No entanto, é importante referir que ao longo da implementação da interface apresentada neste Capítulo foram tidos em consideração princípios e testes de usabilidade de uma forma iterativa [25, 27, 28].

4

Avaliação

“Make everything as simple as possible, but not simpler.”

- Albert Einstein

O Sistema de Informação *ThermInfo* foi desenvolvido e os dados estruturais e termoquímicos de compostos orgânicos foram inseridos e categorizados na base de dados. Estes dados foram compilados da literatura e avaliados criticamente pelo Grupo de Energética Molecular – Centro de Química e Bioquímica. Actualmente, o sistema já se encontra apto para receber novos dados de compostos orgânicos inseridos por investigadores na área da termoquímica que se encontrem registados no sistema.

4.1. Base de Dados

As estatísticas da base de dados, em Julho de 2009, podem ser visualizadas na Figura 51. A sua observação permite verificar a representatividade do conjunto de dados. Presentemente, existem cerca de 3000 compostos orgânicos únicos e não redundantes com todos os dados descritos anteriormente disponíveis. A análise do histograma mostra que os 2956 compostos estão divididos por 6 classes, 12 subclasses e 489 famílias; estão disponíveis mais de 1300 valores de entalpias para cada um dos

estados físicos/transição de estado, sendo que a entalpia de formação no estado gasoso contém o maior número de dados (1740).

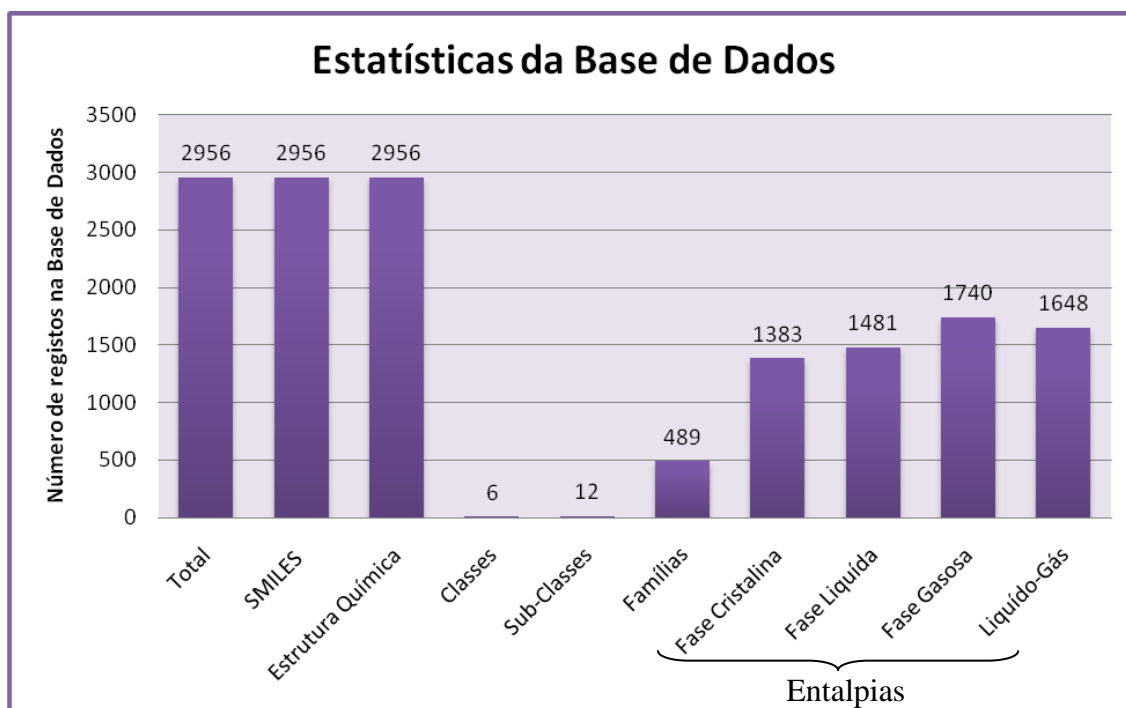


Figura 51 – Representação gráfica das estatísticas da base de dados: número de compostos na base de dados que contêm as diferentes categorias.

Foi efectuada uma avaliação ao desempenho das consultas/inserções realizadas sobre a base de dados para as funcionalidades disponíveis aos utilizadores em geral: pesquisa simples; pesquisa estrutural; e inserção de dados. O desempenho foi medido a partir da recolha do tempo necessário para a execução de cada *query* (média de 5 execuções) à base de dados e encontram-se especificados na Tabela 1.

Tabela 1 – Tempo, em segundos, utilizado para as *queries* realizadas para pesquisar um composto e inserir dados, no conjunto de dados actual.

Funcionalidade	Queries	Tempo (segundos)	Tuplos avaliados	Tuplos retornados
Pesquisa simples por nome (por exemplo: <i>methanol</i>)	SELECT mid, mol_id, casrn, name, formula, smile, instr(name,'methanol') as fpos FROM molecule WHERE molecule.name LIKE '%methanol%' order by fpos, name LIMIT 100;	0.01	2956	16
Pesquisa estrutural com base em: - Compound: <i>propane</i> - Molecular weight: > 50 - Physical state: Liquid - Class: Ring Systems Containing Isolated Benzenoid and Non-Benzenoid Rings - Characteristics: Alkane, Arene	1) SELECT molecule FROM mol_char WHERE charact IN(1, 4) GROUP BY molecule HAVING COUNT(*) >=2; 2) SELECT molecule.mid, molecule.mol_id, molecule.casrn, molecule.name, molecule.formula, molecule.smile FROM molecule, class WHERE molecule.mid IN(2531, 2532, 2533, 2534, 2535, 2536, 2584, 2585, 2586, 2587, 2840, 2841, 2842, 2843, 2844, 2845, 2846, 2847, 2848, 2850, 2883, 2884, 2886, 2887, 2888, 2889, 2890, 2892, 2893, 2895, 2896, 2897, 2947, 2948, 2949, 2950, 2951, 2952, 2953) AND molecule.name LIKE '%propane%' AND molecule.mw > '50' AND molecule.state = 'l' AND molecule.class = class.cid AND class.name = 'Ring Systems Containing Isolated Benzenoid and Non-Benzenoid Rings';	0.01 + 0.00 = 0.01	5831 + 40 = 5871	39 + 5
Inserir um novo composto com todos os parâmetros relativos a propriedades estruturais e termoquímicas preenchidos e uma referência.	1)SELECT casrn FROM molecule WHERE casrn = '10025-87-3'; 2) SELECT mol_id FROM molecule order by mid DESC LIMIT 1; 3) INSERT INTO molecule (mol_id) VALUES ('CO02957'); 4) UPDATE molecule SET casrn = '10025-87-3', name = 'meth', formula = 'C5H6', mw = '40', state = 'l', smile = 'CCCC', usmile = 'CCCC' WHERE mol_id = 'CO02957'; 5) SELECT mid FROM molecule order by mid DESC LIMIT 1; 6) SELECT cid FROM characteristic WHERE name = 'Alkane'; 7) INSERT INTO mol_char (molecule, charact) VALUES ('2957', '1'); 8) INSERT INTO data (idmol, crys, cerror, liq, lerror, gas, gerror, phase, perror, obs) VALUES (2957, 40, 0, 40, 0, 40, 0, 40, 0, 'exp values');	0.01+0.00+ 0.02+0.00+ 0.00+0.00+ 0.00+0.04+ 0.02 + 0.04 = 0.13	1+ 2956 + + 0 + 1 + 2957 + 28 + 0 + 0 + 0 + 0	0 + 1 + 0 + 0 + 1 + 1 + 0 + 0 + 0 + 0

	<p>9) INSERT INTO reference (author, journal, year, volume, page) VALUES ('JPL', 'Nature', '2009', '50', '70');</p> <p>10) INSERT INTO data_ref (idmol, refid) VALUES (2957, 3);</p>			
--	--	--	--	--

Analisando os dados da Tabela 1 verifica-se que, em geral, as *queries* executadas sobre a base de dados obtiveram um tempo de resposta pequeno. O tempo de execução é prioritário para a pesquisa de dados (uma vez que o utilizador espera uma resposta imediata), tendo-se o sistema revelado eficiente (0.01 segundos). Dado o bom desempenho na execução de *queries*, para este conjunto de dados, não foram realizados estudos de escalabilidade no âmbito deste projecto. Uma vez que para isso seria necessário o acréscimo de dados fictícios de forma a averiguar o tipo de função matemática que modela o tempo de execução. No entanto, este será um estudo a realizar quando o conjunto de dados ganhar maior proporção.

4.2. Usabilidade da Interface

Uma vez que um dos principais objectivos do desenvolvimento do *ThermInfo* é a apresentação de informação e a cooperação da comunidade científica com a inserção de novos dados, tornou-se evidente a necessidade da realização de avaliações à usabilidade da sua interface *Web*.

Os principais objectivos da avaliação foram [25, 27, 28]:

- a. a obtenção de uma interface com funcionalidades adequadas ao problema e um design adequado à realização das mesmas pelos utilizadores, de modo fácil e eficiente.
- b. detectar problemas de usabilidade do sistema em fases precoces do seu desenvolvimento, em termos de:
 - i. facilidade de aprendizagem;
 - ii. rapidez no desenvolvimento de tarefas;
 - iii. número de erros cometidos na interacção;

- iv. facilidade de memorização do modo como as tarefas são realizadas;
- v. satisfação do utilizador.

Foram realizados dois tipos principais de avaliação durante o desenvolvimento do sistema *ThermInfo*, especialmente para as funcionalidades que vão estar disponíveis aos utilizadores em geral, a pesquisa e a inserção de compostos.

1. Inspeção de Usabilidade (sem utilizadores)

A inspeção de usabilidade foi realizada por alguns membros da equipa do sistema *ThermInfo* e baseou-se em:

- a) Avaliação dos protótipos de baixa-fidelidade: esta avaliação ocorreu numa fase inicial do desenvolvimento da interface e teve como principal objectivo verificar se as funcionalidades requeridas foram correctamente compreendidas, isto é, serviu como complemento às entrevistas realizadas para elaborar a análise de requisitos (Figura 52). Esta avaliação serviu também para testar possibilidades de *design* e *layout* a serem utilizadas em fases posteriores do projecto.

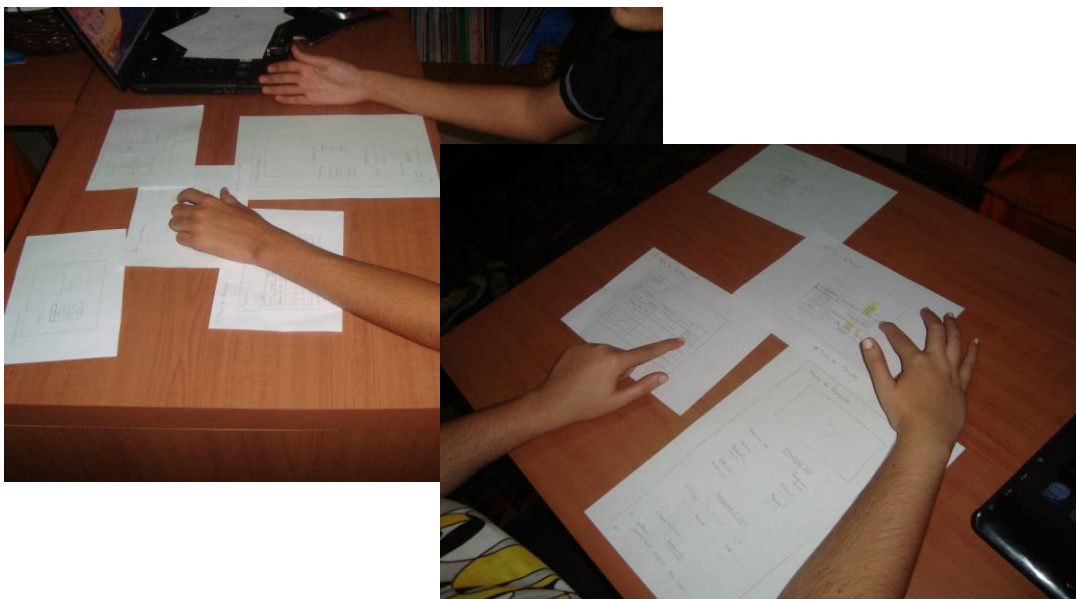


Figura 52 – Avaliação dos protótipos de baixa fidelidade. Um dos analistas apresenta os protótipos de papel de acordo com as escolhas do utilizador durante a realização das tarefas, simulando o fluxo de interacção.

- b) Análises Heurísticas: com base nos princípios gerais de usabilidade tais como recomendações e *guidelines* para o desenvolvimento de interfaces *Web*. Algumas das recomendações utilizadas foram: as do *World Wide Web Consortium (W3C)*¹⁶, Nielsen¹⁷ e validação de HTML¹⁸ (Figura 53). Ao longo do desenvolvimento da interface *Web* foram efectuadas listagens de problemas de acordo com violações às recomendações mencionadas anteriormente. Assim, a interface final é produto de uma aplicação iterativa destas recomendações [22, 26, 32-35].

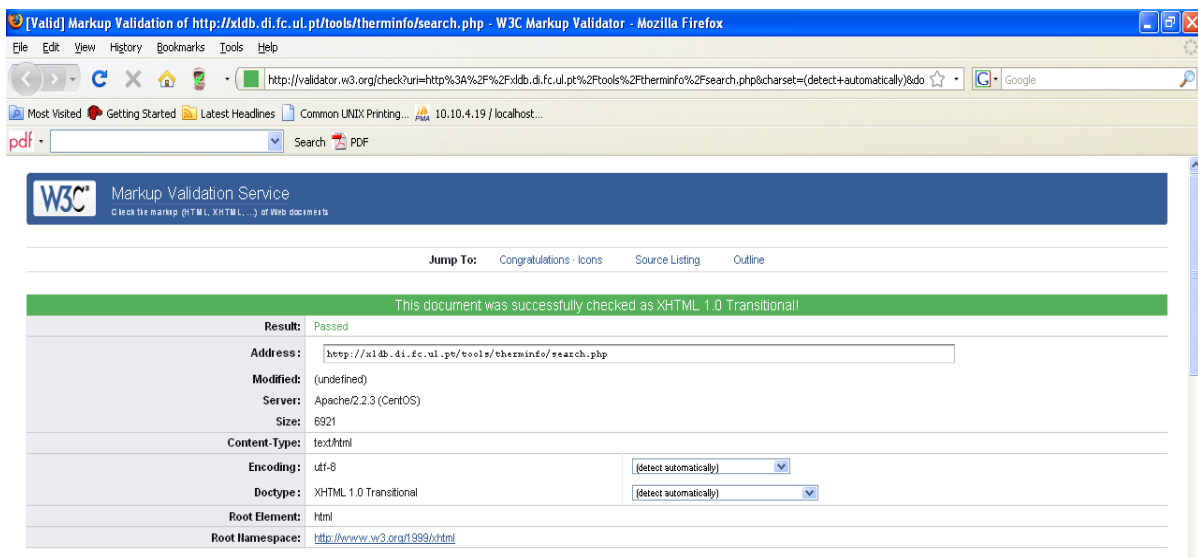


Figura 53 – Verificação do código HTML num sistema de validação do W3C.

De todas as recomendações existentes, foram tidas em especial atenção as seguintes:

- Visibilidade do sistema, utilizando mensagens adequadas para transmitir os erros (Figura 54).

¹⁶ W3C: <http://www.w3.org/>

¹⁷ Guidelines para desenvolvimento de interface Web de Nielsen: <http://www.useit.com/alertbox/>

¹⁸ Validação de HTML (W3C): <http://validator.w3.org/>

Search

123-2-1 CAS RN Search

CASRN format: NNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit). [Help]

Security code: 77#*[7 777]

[Type only **numerical characters**. Ignore letters and special characters.]

Invalid CAS registry number format: 123-2-1. Make sure the CAS registry number you have entered matches the standard format: NNNNNNN-NN-N (1-7 digits, hyphen, 2 digits, hyphen, 1 digit).

Figura 54 – Apresentação de mensagens adequadas para transmitir erros.

- Permitir a navegação dentro e entre páginas, deixando sempre visível o menu lateral que faz ligações a outras funcionalidades do *ThermInfo*.
- Consistência ao longo da interface, utilizando sempre o mesmo design para todas as funcionalidades e formulários.
- Prevenção de erros, utilizando informações de preenchimento nos formulários e documentação de ajuda (Figura 55).

Search

Molecular Formula Search

Please type Molecular Formula with atoms in CHXNOS (X = halogen) order.
Wildcard: ? represents one character [Help]

Figura 55 – Prevenção de erros durante a pesquisa de um composto, utilizando informações do formato do termo de pesquisa.

- Estética e design minimalista, de forma a centrar a atenção do utilizador na tarefa que está a desempenhar e não na interface.
- c) Percurso Cognitivo: análise do caminho percorrido para realizar cada tarefa quer em termos de número de passos (GOMS – *Goals, Operators,*

Methods, Selection) quer em termos de tempo de execução de cada um dos passos (KLM – *Keystroke Level Model*) (Tabela 2) [25].

Tabela 2 – Tempo de execução para cada um dos comandos utilizados durante a realização de tarefas segundo o modelo KLM.

Comando	Abreviatura do Comando	Tempo (t) (segundos)
Keying – tempo que o utilizador demora a carregar numa tecla	K	0,28
Tempo que o utilizador demora a pressionar o botão do rato	B	0,2
Pointing – tempo que o utilizador demora a apontar uma posição no ecrã	P	1,1
Homing - tempo que o utilizador demora a mover a mão do rato para o teclado ou vice-versa.	H	0,4
Mentally Preparing – tempo que o utilizador leva a preparar mentalmente a próxima acção	M	1,35
Responding – tempo que o utilizador deve esperar pela resposta do sistema	R	? (variável)

Seguidamente foram calculados os tempos de execução esperados para a realização de três tarefas relacionadas com as funcionalidades disponíveis para os utilizadores. Mais adiante estes tempos serão utilizados como referência, uma vez que os testes de usabilidade com utilizadores se basearam na realização das mesmas tarefas.

Objectivo: Pesquisar o composto com SMILES: CCCCCO.

Acções:

- 1 – Mover a mão para o rato. **H**
- 2 – Preparação Mental. **M**
- 3 – Mover o cursor do rato até “Search”. **P**
- 4 – Carregar no botão do rato. **B**
- 5 – Preparação Mental. **M**
- 6 – Mover o cursor até à caixa do termo da pesquisa. **P**
- 7 – Mover a mão para o teclado. **H**
- 8 – Preparação Mental. **M**

- 9 – Digitar o termo da pesquisa CCCCCO. **6K**
- 10 – Mover a mão para o rato. **H**
- 11 – Preparação Mental. **M**
- 12 – Mover o cursor até à caixa da escolha do tipo de pesquisa. **P**
- 13 – Clicar no tipo de pesquisa pretendido: “SMILES”. **B**
- 14 – Preparação Mental. **M**
- 15 – Mover o cursor até à caixa do código de segurança. **P**
- 16 – Clicar na caixa do código de segurança. **B**
- 17 – Mover a mão para o teclado. **H**
- 18 – Preparação Mental. **M**
- 19 – Digitar o código de segurança. **6K**
- 20 – Mover a mão para o rato. **H**
- 21 – Preparação Mental. **M**
- 22 – Mover o cursor até ao botão “Search”. **P**
- 23 – Clicar no botão “Search”. **B**
- 24 – Resposta do sistema. **R** (0,18 segundos)
- 25 – Preparação Mental. **M**
- 26 – Mover o cursor até ao composto pretendido. **P**
- 27 – Clicar no botão “View”. **B**
- 28 – Resposta do sistema. **R** (0,14 segundos)

HMPBMPHMKKKKKKHMPBMPBHMKKKKKKHMPBRMPBR

Para calcular o tempo de execução esperado, foram substituídos os comandos pelos tempos estimados para a sua execução (Tabela 2).

Obteve-se:

$$t_{\text{execução}} = 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,4 + 1,35 + (6 * 0,28) + 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 0,4 + 1,35 + (6 * 0,28) + 0,4 + 1,35 + 1,1 + 0,2 + 0,18 + 1,35 + 1,1 + 0,2 + 0,14 = \mathbf{24,08}$$
 segundos

Objectivo: Realizar uma pesquisa estrutural para encontrar o composto com:

- Nome do composto: *methylene*.
- Peso molecular: *superior a 30*;
- Estado físico: *crystal*;
- Característica: *polymer*.

Acções:

- 1 – Mover a mão para o rato. **H**
- 2 – Preparação Mental. **M**
- 3 – Mover o cursor do rato até “*Structural Search*”. **P**
- 4 – Carregar no botão do rato. **B**
- 5 – Preparação Mental. **M**
- 6 – Mover o cursor até à caixa do nome do composto. **P**
- 4 – Carregar no botão do rato. **B**
- 7 – Mover a mão para o teclado. **H**
- 8 – Preparação Mental. **M**
- 9 – Digitar o nome do composto *methylene*. **9K**
- 10 – Mover a mão para o rato. **H**
- 11 – Preparação Mental. **M**
- 12 – Mover o cursor até à caixa da escolha do tipo de intervalo para o peso molecular. **P**
- 13 – Clicar no tipo de intervalo pretendido: “>”. **B**
- 14 – Preparação Mental. **M**
- 15 – Mover o cursor até à caixa do peso molecular. **P**
- 16 – Mover a mão para o teclado. **H**
- 17 – Preparação Mental. **M**
- 18 – Digitar o peso molecular “30”. **2K**
- 19 – Mover a mão para o rato. **H**
- 20 – Preparação Mental. **M**
- 21 – Mover o cursor até à caixa da escolha do estado físico. **P**
- 22 – Clicar no estado físico pretendido: “crystal”. **B**
- 23 – Preparação Mental. **M**
- 24 – Mover o cursor até à checkbox correspondente à característica polymer. **P**
- 25 – Clicar na checkbox **B**
- 26 – Preparação Mental. **M**
- 27 – Mover o cursor até à caixa do código de segurança. **P**
- 28 – Clicar na caixa do código de segurança. **B**
- 29 – Mover a mão para o teclado. **H**
- 30 – Preparação Mental. **M**
- 31 – Digitar o código de segurança. **6K**
- 32 – Mover a mão para o rato. **H**
- 33 – Preparação Mental. **M**
- 34 – Mover o cursor até ao botão “Search”. **P**
- 35 – Clicar no botão “Search”. **B**
- 36 – Resposta do sistema. **R**
- 37 – Preparação Mental. **M**

38 – Mover o cursor até ao composto pretendido. **P**

39 – Clicar no botão “View”. **B**

40 – Resposta do sistema. **R** (0,12 segundos)

HMPBMPHMKKKKKKKKKHMPBMPHMKKHMPBMPBMPBHMKKKKKKKHMPBRMPBR

Para calcular o tempo de execução esperado, foram substituídos os comandos pelos tempos estimados para a sua execução (Tabela 2).

Obteve-se:

$$t_{\text{execução}} = 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,4 + 1,35 + (9 \cdot 0,28) + 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,4 + 1,35 + 0,28 + 0,28 + 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 0,4 + 1,35 + (6 \cdot 0,28) + 0,4 + 1,35 + 1,1 + 0,2 + 0,13 + 1,35 + 1,1 + 0,2 + 0,12 = 35,31 \text{ segundos}$$

Objectivo: Inserção de um composto de *CASRN 10025-87-3* e quatro campos:

- Nome do composto: *methylene*;
- Peso molecular: 30;
- Estado físico: *crystal*;
- Característica: *polymer*.

Acções:

- 1 – Mover a mão para o rato. **H**
- 2 – Preparação Mental. **M**
- 3 – Mover o cursor do rato até “Insert Data”. **P**
- 4 – Carregar no botão do rato. **B**
- 5 – Preparação Mental. **M**
- 6 – Mover o cursor até à caixa do CASRN. **P**
- 7 – Carregar no botão do rato. **B**
- 8 – Mover a mão para o teclado. **H**
- 9 – Preparação Mental. **M**
- 10 – Digitar o CASRN do composto 10025-87-3. **10K**
- 11 – Mover a mão para o rato. **H**
- 12 – Preparação Mental. **M**
- 13 – Mover o cursor até à caixa do nome do composto. **P**
- 14 – Carregar no botão do rato. **B**
- 15 – Mover a mão para o teclado. **H**
- 16 – Preparação Mental. **M**

- 17 – Digitar o nome do composto methylene. **9K**
 19 – Mover a mão para o rato. **H**
 20 – Preparação Mental. **M**
 21 – Mover o cursor até à caixa do peso molecular. **P**
 22 – Carregar no botão do rato. **B**
 23 – Mover a mão para o teclado. **H**
 24 – Preparação Mental. **M**
 25 – Digitar o peso molecular “30”. **2K**
 26 – Mover a mão para o rato. **H**
 27 – Preparação Mental. **M**
 28 – Mover o cursor até à caixa da escolha do estado físico. **P**
 29 – Clicar no estado físico pretendido: “crystal”. **B**
 30 – Preparação Mental. **M**
 31 – Mover o cursor até à checkbox correspondente à característica polymer. **P**
 32 – Clicar na checkbox **B**
 33 – Preparação Mental. **M**
 34 – Mover o cursor até ao botão “Add Data”. **P**
 35 – Clicar no botão “Add Data”. **B**
 36 – Resposta do sistema. **R**

HMPBMPBHMKKKKKKKKKKHMPBHMKKKKKKKKKKHMPBHMKKHMPBMPBMPBR

Para calcular o tempo de execução esperado, foram substituídos os comandos pelos tempos estimados para a sua execução (Tabela 2).

Obteve-se:

$$t_{\text{execução}} = 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 0,4 + 1,35 + (10 \cdot 0,28) + 0,4 + 1,35 + 1,1 + 0,2 + 0,4 + (9 \cdot 0,28) + 1,35 + 0,28 + 0,28 + 0,4 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 1,35 + 1,1 + 0,2 + 0,37 = \mathbf{26,85 \text{ segundos}}$$

Tabela 3 – Quadro resumo dos tempos de execução totais, em segundos, para cada uma das tarefas delimitadas com base no modelo KLM-GOMS.

Tarefa	Tempo de execução total (segundos)
Pesquisa Simples	24,08
Pesquisa Estrutural com base em 4 parâmetros	35,31
Inserção de um composto com 5 parâmetros	26,85

A análise destes resultados (Tabela 3) permite-nos verificar que a execução das três tarefas requer pouco tempo de execução. Obviamente que este tempo vai variar com diversos factores como, por exemplo, a experiência e destreza do utilizador, o conhecimento da interface, entre outros.

2. Testes de usabilidade (com utilizadores)

A avaliação de usabilidade com a participação de utilizadores deu-se em fases finais (depois de se ter uma interface funcional) e teve como principais objectivos melhorar a ferramenta e identificar dificuldades na realização das tarefas, por parte dos utilizadores.

Foram testadas três funcionalidades fulcrais e destinadas a utilizadores: a pesquisa simples, a pesquisa estrutural e a inserção de dados. Com este propósito foram delineadas e apresentadas três tarefas aos utilizadores:

1 – “Encontre a ficha do composto com SMILES: CCCCCO”

Procedimento Esperado: O utilizador clica em “Search” no menu lateral, preenche a caixa de pesquisa com o termo “CCCCCO”, selecciona o tipo de pesquisa “SMILES”, preenche o código de segurança (apenas os dígitos), clica no botão “Search” e no primeiro composto da lista de resultados clica em “View”, para visualizar a ficha do composto.

2 – “Encontre a ficha do composto com:

- Peso molecular: *superior a 30*;
- Estado físico: *crystal*;
- Característica: *polymer*;
- Nome do composto: *methylene*.”

Procedimento Esperado: O utilizador clica em “Structural Search” no menu lateral, preenche/selecciona no formulário os quatro parâmetros especificados acima, preenche o código de segurança (apenas os dígitos), clica no botão “Search” e no primeiro composto da lista de resultados clica em “View”, para visualizar a ficha do composto.

3 – Insira um composto de *CASRN 10025-87-3* e quatro campos à escolha.

Procedimento Esperado: O utilizador clica em “*Insert Data*” (o sistema de *Login* ainda não estava disponível, à data de realização dos testes de usabilidade) no menu lateral, preenche/selecciona no formulário o *CASRN* e quatro parâmetros e clica no botão “*Add Data*” .

Nestes testes foram registadas medidas de performance, com base na velocidade de execução de cada uma das tarefas (tempo dispendido a realizar a tarefa) e no número de erros cometidos, relativamente ao procedimento esperado (Figura 56).



Figura 56 – Avaliação dos protótipos de alta-fidelidade. O utilizador desempenha as tarefas apresentadas enquanto o analista monitoriza o tempo dispendido e o número de erros cometidos durante a realização das tarefas.

Foi recolhida a opinião dos utilizadores com base em comentários ao longo da utilização da interface e posteriormente num questionário. Este questionário centrou-se em três questões categóricas subjectivas de escala entre 0 e 5:

1 – Como classifica a facilidade/tempo de aprendizagem necessário?

2 – Como classifica a facilidade em memorizar a forma como a tarefa é executada?

3 – Como classifica a sua satisfação em utilizar esta interface?

Este teste de avaliação de usabilidade foi realizado para um total de 43 participantes, dos quais 18 eram profissionais da área da Química e 25 tinham elevados conhecimentos e experiência na utilização das Tecnologias de Informação. Os resultados dos testes de usabilidade e questionários encontram-se resumidos na Tabela 4, em termos de média e desvio padrão para cada uma das tarefas e para cada um dos grupos. Toda a análise estatística foi realizada utilizando o SPSS 16.0. (ANEXO I – Tabela de dados utilizada para realizar a avaliação de usabilidade).

Tabela 4 – Resultados dos testes de usabilidade para todos os participantes e para cada um dos grupos em específico.

Grupo de Utilizadores	N	Tarefa	Média do tempo (segundos)	Desvio Padrão	Nº de erros	Desvio Padrão	Facilidade (0-5)	Memorização (0-5)	Satisfação (0-5)
All	43	1	65.8	21.4	0.4	0.7			
		2	68.3	20.1	0.3	0.6	4.1	4.4	4.1
		3	64.8	14.5	0.3	0.4			
I	25	1	66.6	21.8	0.3	0.5			
		2	68.0	18.6	0.3	0.5	4.0	4,5	4.1
		3	64.4	13.3	0.2	0.4			
Q	18	1	64.8	21.5	0.6	1.0			
		2	68.6	22.3	0.3	0.6	4.3	4.2	4.0
		3	65.3	16.2	0.4	0.5			

I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: relativa à pesquisa simples. Tarefa 2: relativa à pesquisa estrutural. Tarefa 3: relativa à inserção de dados.

Seguidamente, apresentam-se um conjunto de *Box Plots* (Figura 57) que nos permitem relacionar visualmente, os valores de tempo que os utilizadores de cada um dos grupos demoram a realizar cada uma das tarefas, com várias medidas estatísticas, nomeadamente, no que diz respeito à dispersão e distribuição do conjunto de dados dentro e entre grupos.

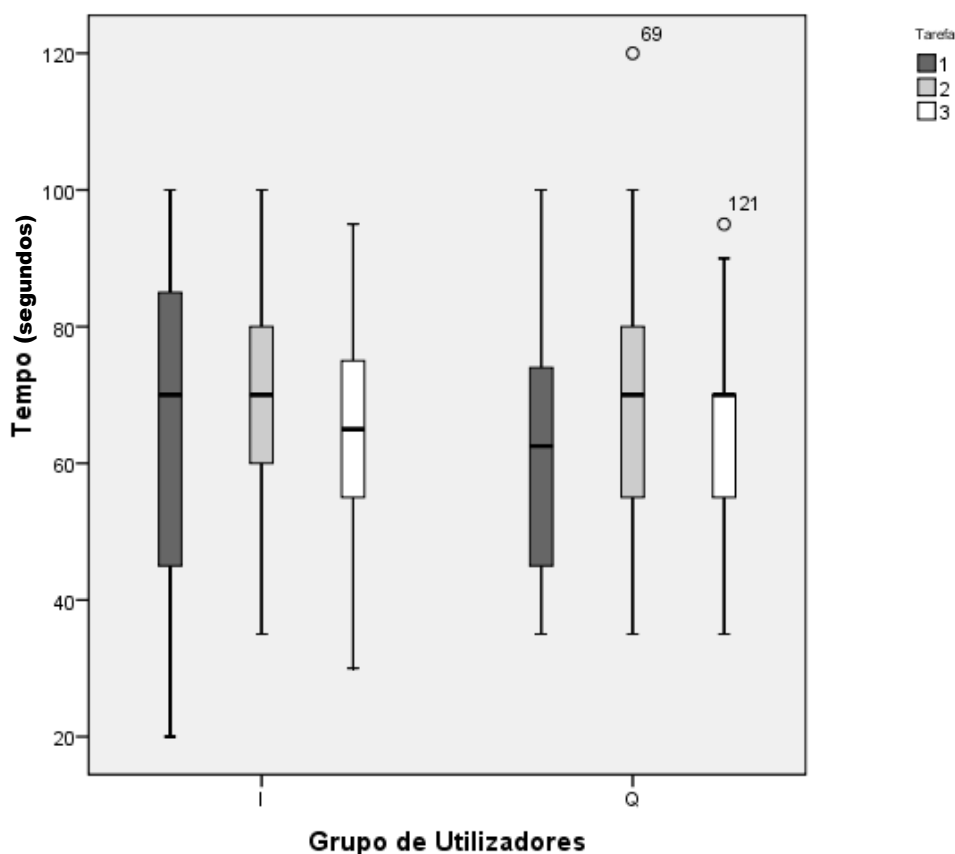


Figura 57 – Representação gráfica, utilizando *Box Plots*, da variável tempo necessário para realizar cada tarefa, para cada um dos grupos de utilizadores. I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

A análise do gráfico para o grupo de utilizadores experientes na utilização de Tecnologias de Informação (I) revela que:

- a tarefa 1 (pesquisa simples) tem uma amplitude superior às restantes tarefas e uma distribuição ligeiramente assimétrica;
- as tarefas 1 e 2 (pesquisa estrutural) apresentam uma amplitude e mediana idênticas, exceptuando para o limite inferior que é mais elevado para a tarefa 2;

- a tarefa 3 (inserção de dados) tem uma mediana inferior às restantes;
- a distribuição do tempo que os utilizadores demoram a realizar cada uma das tarefas 2 e 3 é aproximadamente simétrica.

A análise do gráfico para o grupo de utilizadores experientes na área da Química (Q) revela que:

- a tarefa 1 e 2 têm uma amplitude superior à tarefa 3;
- a tarefa 1 apresenta o valor de mediana mais baixo e uma distribuição assimétrica.
- a tarefa 3 apresenta uma distribuição assimétrica e 50% dos valores centrais da distribuição (zona interquartil) encontram-se abaixo do valor da mediana;
- existem 2 *outliers* (valores fora da amplitude da amostra) para as tarefas 2 e 3, com tempos de execução muito elevados.

Comparando os dois grupos de utilizadores relativamente a cada uma das tarefas, podemos verificar que:

- para a tarefa 1 o grupo Q apresenta um valor de mediana e uma amplitude inferior ao do grupo I;
- para a tarefa 2 os grupos têm um comportamento muito semelhante;
- para a tarefa 3 o grupo I tem um valor de mediana inferior ao do grupo Q, no entanto apresenta uma dispersão superior.

A análise deste gráfico e dos valores das médias e desvios padrão apresentados na Tabela 4 permitem verificar que os grupos apresentam um comportamento semelhante relativamente às três tarefas. No entanto, não podemos desprezar as diferenças encontradas sem confirmar que estatisticamente estas se devem ao acaso e que efectivamente os grupos são iguais em termos de valor médio. Isto é, queremos testar se a formação profissional interfere com o desempenho com que os utilizadores

realizam as tarefas na interface. Assim, será efectuado o teste estatístico *t-Student* para duas amostras independentes com as seguintes hipóteses [25]:

H_0 : não existe diferença entre as médias do desempenho em cada grupo. ($\mu_I = \mu_Q$)

versus

H_1 : existe diferença entre as médias do desempenho em cada grupo. ($\mu_I \neq \mu_Q$)

No entanto este teste tem como pressupostos que os grupos em análise sejam provenientes de uma população com distribuição Normal e igualdade das variâncias. De modo a avaliar o primeiro pressuposto, foi realizado o teste de *Kolmogorov – Smirnov* (KS) [25]. Este permite testar as seguintes hipóteses:

H_0 : a população tem uma distribuição Normal.

versus

H_1 : a população não tem uma distribuição Normal.

O *p-value* de cada um dos testes encontra-se especificado na Tabela 5.

Tabela 5 – Valores de *p-value* obtido para cada um dos testes de KS realizados.

<i>Grupo</i>	<i>Tarefa</i>	<i>p-value</i> (teste KS)
I	1	0,090
	2	0,188
	3	0,2
Q	1	0,2
	2	0,072
	3	0,027

I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

A um nível de significância de 5% não rejeitamos a hipótese nula para nenhum dos testes realizados à exceção do teste para o Grupo Q - Tarefa 3 ($p\text{-value} = 0,027 < \text{nível de significância}$). Isto é, todas as amostras provem de populações com distribuição Normal, à exceção do Grupo Q – Tarefa 3. No entanto vamos assumir que uma amostra de 18 indivíduos é suficientemente elevada para que possamos aproximar esta amostra como proveniente de uma população com distribuição Normal.

No que diz respeito ao pressuposto de igualdade de variâncias, foi realizado o teste de *Levene*, no qual as hipóteses são:

H_0 : existe homogeneidade entre as variâncias ($\sigma_I = \sigma_T$).

versus

H_1 : não existe homogeneidade entre as variâncias ($\sigma_I \neq \sigma_T$).

Tabela 6 – Valores de $p\text{-value}$ obtidos para os testes de *Levene* e *t-Student*.

<i>Tarefa</i>	<i>p - value (teste de Levene)</i>	<i>p-value (teste t-Student)</i>
1	0,772	0,787
2	0,662	0,923
3	0,567	0,848

Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

Para o teste de *Levene* os $p\text{-values}$ obtidos (Tabela 6) são todos muito elevados, pelo que para qualquer nível de significância não se rejeita a hipótese de que as variâncias são iguais. Assim, considera-se que as variâncias populacionais, embora desconhecidas, são iguais.

Assim, podemos considerar válida a aplicação do teste *t-Student* para comparar os dois grupos de amostras. Os resultados deste teste encontram-se especificados na Tabela 6. Todos os $p\text{-values}$ obtidos são muito elevados, pelo que para qualquer nível de significância não se rejeita a hipótese de que não existem diferenças entre as médias

do desempenho em cada grupo. Logo, considera-se que a formação profissional não interfere com o desempenho com que os utilizadores realizam as tarefas na interface.

Tabela 7 – Valores de tempo para desempenhar uma tarefa esperados segundo o modelo KLM-GOMS e a média de tempo realmente obtidos em cada um dos grupos de utilizadores e tarefas a desempenhar.

<i>Grupo</i>	<i>Tarefa</i>	<i>Tempo esperado - KLM-GOMS - (segundos)</i>	<i>Tempo obtido - média - (segundos)</i>
All	1	24,08	65,8
I	1		66,6
Q	1		64,8
All	2	35,31	68,3
I	2		68,0
Q	2		68,6
All	3	26,85	64,8
I	3		64,4
Q	3		65,3

I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

A análise da Tabela 7 permite-nos verificar que as médias dos tempos obtidos são superiores aos tempos que seriam esperados segundo o modelo KLM – GOMS, no entanto, seguem a mesma tendência. Este modelo serviu, em fases iniciais, para fazer previsões sobre o comportamento dos utilizadores. No entanto, este modelo é concebido para utilizadores experientes, conhecedores da interface e numa interacção sem cometer

qualquer tipo de erros. O que não foi o caso dos tempos obtidos na avaliação de usabilidade com os utilizadores. Estes utilizaram a interface pela primeira vez, onde foram cometidos alguns erros. Obviamente, o tempo de preparação mental e de recuperação de erros é muito superior numa primeira utilização da interface relativamente aos tempos utilizados no cálculo do tempo esperado segundo o modelo KLM-GOMS. No entanto, ambos os tempos permitem considerar a interface eficiente [25].

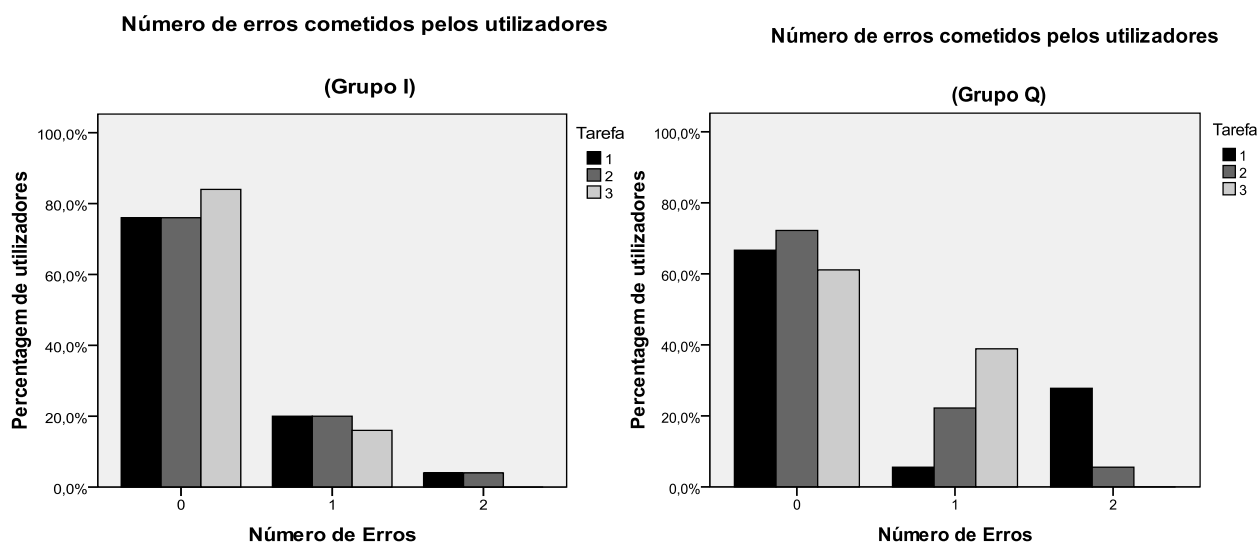


Figura 58 – Distribuição da percentagem de utilizadores pelo número de erros cometidos em cada uma das tarefas. I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

A análise da Tabela 4 permite verificar que a média do número de erros cometidos é muito baixa, uma média inferior a 0,6 erros. A Figura 58 permite visualizar a distribuição da percentagem de utilizadores a cometer um determinado número de erros. Em ambos os grupos a maior parte dos utilizadores não cometeram erros durante a utilização. As diferentes tarefas não apresentam grandes diferenças em termos do número de erros cometidos tal como as diferenças entre os grupos não parecem significativas. No entanto, esta análise de resultados deveria ser confirmada recorrendo ao teste estatístico *Qui-quadrado*, para verificar se o número de erros cometidos durante a realização das tarefas depende da formação profissional do utilizador. Os principais

erros cometidos foram: não ter em conta que a pesquisa por SMILES é *case sensitive*, apesar do aviso presente na página; não inserir o código de segurança ou inseri-lo com os números e letras na realização de pesquisas; e tentativa da realização da pesquisa estrutural na funcionalidade de pesquisa simples. Devido ao facto do número de erros cometidos ter sido muito baixo e eles serem de fácil recuperação/percepção, estes testes não foram realizados.

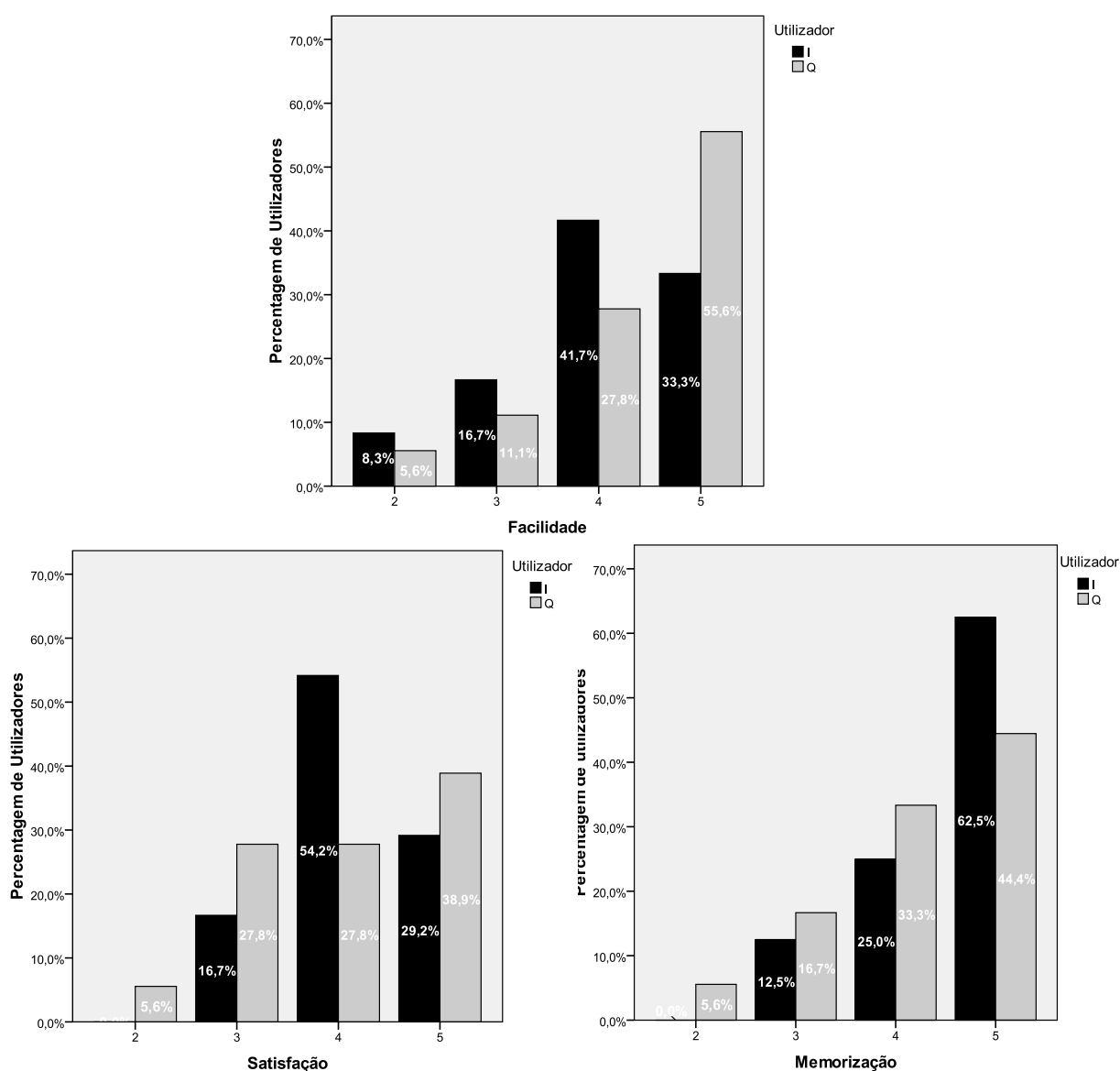


Figura 59 – Distribuição da percentagem de utilizadores pelas pontuações (escala 0-5) atribuídas aos três atributos de usabilidade: facilidade, satisfação no uso e facilidade de memorização. I – utilizadores experientes no uso de Tecnologias de Informação e Q – utilizadores experientes na área da Química. Tarefa 1: pesquisa simples. Tarefa 2: pesquisa estrutural. Tarefa 3: inserção de dados.

As medidas subjectivas da opinião dos utilizadores relativamente aos três atributos de usabilidade: facilidade de aprendizagem; memorização da forma como as tarefas são realizadas; e satisfação do uso, obtiveram resultados bastante bons. Em ambos os grupos a pontuação média foi superior a 4 (Tabela 4) e a distribuição das pontuações pelos utilizadores de ambos os grupos encontra-se representada na Figura 59.

De forma geral os utilizadores comentaram que numa utilização posterior do sistema não cometeriam os mesmos erros e que a similaridade e simplicidade dos formulários facilitavam bastante o uso e a memorização necessária de como são realizadas as tarefas. As principais sugestões realizadas pelos utilizadores foram: os avisos deveriam ser mais visíveis utilizando uma letra de tamanho superior ou cores diferentes; deveria ser implementado o uso de *wildcards* na pesquisa por fórmula química; e as características deveriam estar ordenadas de uma forma mais intuitiva.

Todos os resultados, comentários e sugestões obtidos neste processo de avaliação foram tidos em conta para aperfeiçoar as funcionalidades já implementadas e a implementar futuramente no sistema *ThermInfo*.

5

Conclusões

“What you get by achieving your goals is not as important as what you become by achieving your goals.”

- Zig Ziglar

Desde cedo que a necessidade de compreender os fenómenos da Natureza e o gosto pelo estudo das Ciências Físico-Químicas e Naturais despertou em mim o desejo de vir a enveredar por uma carreira de investigação nesta área. Razão pela qual ingressei no Curso de Biologia da Faculdade de Ciências da Universidade de Lisboa. Posteriormente, optei pelo ramo de especialização em Biologia Funcional e de Sistemas, no qual foram leccionadas cadeiras que servem de base para uma abordagem integrativa na resolução de problemas a vários níveis de organização, isto é, focando a atenção em todos os componentes do sistema e na forma como estes interagem enquanto parte do mesmo. A cadeira de Fundamentos de Bioinformática, leccionada pelo Professor Doutor Francisco Couto e os vastos conhecimentos do Professor Doutor José Alberto Feijó na área da Biologia de Sistemas, os quais proporcionaram diversas discussões, contribuíram inegavelmente para perceber o quão importante e desafiante é a aplicação da informática ao estudo destes sistemas complexos e ao desenvolvimento de modelos de uma forma integrativa [36]. O interesse em alargar os meus conhecimentos nesta área conduziu à minha inscrição no Mestrado em Tecnologias de Informação Aplicadas às Ciências Biológicas e Médicas.

De entre as várias cadeiras de Mestrado que frequentei no primeiro ano do curso, algumas mereceram uma atenção especial da minha parte, nomeadamente, Introdução às Bases de Dados, Aplicações na *Web* e Sistemas Interactivos. A escolha do tema de Dissertação de Mestrado, que apresentei neste documento, em muito se deveu aos conceitos introduzidos nas cadeiras mencionadas e na vontade de desenvolver um sistema que facilita o dia-a-dia de investigação de vários cientistas.

Um dos principais objectivos deste projecto era eliminar os problemas que advêm dos processos de armazenamento, organização, visualização e partilha de informação relativa às propriedades estruturais e termoquímicas de compostos orgânicos. Para isso desenvolvi um Sistema de Informação – *ThermInfo* – baseado numa base de dados para armazenar e organizar os dados, com uma interface pública de fácil utilização para inserção e consulta de informação e uma interface para administração do sistema. Este Sistema de Informação encontra-se disponível na Internet através do endereço <http://www.therminfo.com> e está funcional para os principais navegadores *Web* (*Firefox*, *Internet Explorer*, *Opera*, *Safari* e *Google Chrome*).

A base de dados do sistema *ThermInfo* ainda é pequena quando comparada com algumas das base de dados comerciais existentes, no entanto, contém cerca de 3000 compostos orgânicos únicos, criticamente avaliados, termodinamicamente consistentes e disponíveis publicamente.

A interface desenvolvida possibilita o uso das seguintes funcionalidades por parte dos:

- utilizadores: pesquisa simples, pesquisa estrutural e inserção de dados (mediante registo e autenticação no sistema).
- administradores: remoção/actualização de dados, validação de dados inseridos pelos utilizadores e controlo do uso do sistema.

De acordo com a avaliação realizada, a interface mostrou-se flexível, fácil de usar, reduzida necessidade de carga cognitiva por parte do utilizador e com tempos necessários para realizar as tarefas curtos.

O desenvolvimento deste Sistema de Informação não foi uma tarefa trivial e representou um desafio para todos os intervenientes, tanto para a equipa do

desenvolvimento informático, como para a equipa de investigadores da área da Termoquímica que compilaram e avaliaram criticamente os dados e que também participaram na modelação e avaliação do sistema *ThermInfo*.

A concepção deste sistema não se centrou apenas na implementação, mas sim num conjunto de tarefas integrativas, nomeadamente: na modelação; na análise de requisitos a vários níveis; nos utilizadores que o vão utilizar; no tipo de funcionalidades que se pretendeu disponibilizar; e nos resultados de avaliação do sistema. A metodologia adoptada permitiu a implementação deste Sistema de Informação para coligir e apresentar propriedades estruturais e termoquímicas de compostos orgânicos de acordo com as expectativas funcionais esperadas e com aplicabilidade na actividade de investigação de forma a agilizar este processo, especialmente quando estão envolvidas grandes quantidades de informação.

As contribuições deste projecto são o ponto de partida para a expansão, pela comunidade científica, do conjunto de dados disponibilizado, tal como a incorporação de novas funcionalidades que poderão conduzir ao aperfeiçoamento do sistema.

Ao longo do desenvolvimento deste projecto, novas frentes foram abertas, que não tendo sido possível incluir neste trabalho poderão constituir direcções futuras que irão ampliar as capacidades do sistema *ThermInfo*. Destacam-se como exemplos:

- estender a compilação de dados de compostos orgânicos a inorgânicos, organometálicos e radicais;
- a incorporação de ferramentas quimio-informáticas, tais como o *JChemPaint*¹⁹ (para desenhar estruturas químicas e combiná-las com termos de consulta textuais de forma a restringir as pesquisas) e o *OpenBabel*²⁰ (para converter uma *string* SMILES em outros formatos de estrutura química);
- a incorporação de outros métodos de pesquisa, tal como por semelhança entre SMILES (utilizando *fingerprints*);

¹⁹ JChemPaint: <http://apps.sourceforge.net/mediawiki/cdk/index.php?title=JChemPaint>

²⁰ OpenBabel: <http://openbabel.org/wiki/>

- a implementação e integração de métodos de previsão de propriedades, para se obterem estimativas de dados termoquímicos para compostos orgânicos de interesse que ainda não tenham sido analisados experimentalmente [29-30].

“The future of chemistry depends on the automated analysis of chemical knowledge, combining disparate data sources in a single resource which can be analyzed using computational techniques to assess and build on these data.”

Townsend et al. (2004)

Bibliografia

- [1] Chen, W. L. *Chemoinformatics: Past, Present, and Future*. J. Chem. Inf. Model. **2006**, *46*, 2230–2255.
- [2] Bachrach, S. M. *Chemistry publication – making the revolution*. J. Cheminformatics **2009**, *1* (doi:10.1186/1758-2946-1-2.).
- [3] Sousa, J. A. *Químio-informática. Conteúdos que urge ensinar*. BSP Química **2002**, *84*, 55–59.
- [4] Engel, T. *Basic Overview of Chemoinformatics*. J. Chem. Inf. Model. **2006**, *46*, 2267–2277.
- [5] Gasteiger, J. E. *Handbook of Chemoinformatics - From Data to Knowledge*; Wiley-VCH: Weinheim, 2003.
- [6] Linstrom, P.J.; Mallard, W.G. (Eds.) *NIST Chemistry WebBook*, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, <<http://webbook.nist.gov>> (accessed in May 2009).
- [7] Modha, J.; Gwinnett, A.; Bruce, M. *A Review of Information Systems Development Methodology (ISDM) Selection Techniques*. Omega **1990**, *18*, 473–490.
- [8] Blanchard, B. S.; Fabrycky, W. J. *Systems engineering and analysis*; 4th ed.; Prentice Hall: New Jersey, 2006.
- [9] Daley, R. F.; Daley, S. J. *Organic Chemistry*; Online organic chemistry textbook, <<http://www.ochem4free.info>> (accessed in May 2009).
- [10] Vollhardt, K.; Schore, N. *Organic Chemistry Structure and Function*; 4th ed.; W. H. Freeman and Company: New York, 2003.
- [11] Chang, R. *Química*; 8ª ed.; McGraw Hill: Lisboa, 2005.
- [12] Morrison, R. T.; Boyd, R. N., *Química Orgânica*; 13ª ed.; Fundação Calouste Gulbenkian: Lisboa, 1996.

- [13] McNaught, A.; Wilkinson, A. *Compendium of Chemical Terminology: Recommendations*; IUPAC International Union of Pure and Applied Chemistry; Blackwell Science, 1997 <<http://goldbook.iupac.org/>> (accessed in May 2009).
- [14] Weininger, D. *SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules*. *J. Chem. Inf. and Comp. Sciences* **1988**, 28, 31–36.
- [15] Weininger, D.; Weininger, A.; Weininger, J. L. *SMILES 2. Algorithm for Generation of Unique SMILES Notation*. *J. Chem. Inf. and Comp. Sciences* **1989**, 29, 97–101.
- [16] Pedley, J. B.; Naylor, R. D.; Kirby, S. P. *Thermochemical Data of Organic Compounds*; 2nd ed., Chapman and Hall: London, 1986.
- [17] Pedley, J. B. *Thermochemical Data and Structures of Organic Compounds*; TRC Data Series; vol. 1; College Station: TX, 1994.
- [18] Martinho Simões, J. A.; Minas da Piedade, M. E. *Molecular Energetics. Condensed-Phase Thermochemical Techniques*; Oxford University Press: New York, 2008.
- [19] Frurip, D.; Britton, L.; Fenlon, W.; Going, J.; Harrison, B. K.; Niemeier, J.; Ural, E. A. *The Role of ASTM E27 Methods in Hazard Assessment: Part I. Thermal Stability, Compatibility, and Energy Release Estimation Methods*. *Process Saf. Progr.* **2004**, 23, 266–278.
- [20] Lourenço, J. A. A.; Diogo, H. P.; Minas da Piedade, M. E. *Safety in the chemical industry: desk screening, thermoanalytical and calorimetric methods for thermal hazard evaluation*. *Engenharia Química* **2006**, 3, 27–35.
- [21] Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *The Properties of Gases and Liquids*; 5th ed.; McGraw-Hill: Singapore, 2001.
- [22] Nielsen, J. *Usability Engineering*; AP Professional; Mountain View EUA, 1993.

- [23] Ramakrishnan, R.; Gehrke, J. *Data Management Systems*; 3rd ed.; McGraw-Hill, 2003.
- [24] Booch, G.; Rumbaugh, J.; Jacobson, I. *The Unified Modeling Language User Guide*; Addison Wesley Longman: MA, 1999.
- [25] Dix, A.; Finlay, J.; Abowd, G. D.; Beale, R. *Human Computer Interaction*; 3rd ed.; Prentice Hall, 2003.
- [26] Williams, H. E.; Lane, D. *Web Database Applications with PHP and MySQL*; 2nd ed.; O'Reilly, 2004.
- [27] Shortliffe, E. H.; Climino, J. J. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*; 3rd ed.; Springer, 2006.
- [28] Holzinger, A. *HCI and Usability for Medicine and Health Care*; Proceedings of the Third Symposium of the Workgroup HCI and Usability Engineering of the Austrian Computer Society; USAB 2007; Graz, Austria; LNCS, Springer, 2007.
- [29] Leal, J. P. *Additive Methods for Prediction of Thermochemical Properties. The Laidler Method Revisited. 1. Hydrocarbons*. *J. Phys. Chem. Ref. Data* **2006**, *35*, 55–76.
- [30] Santos, R. C.; Leal, J. P.; Martinho Simões, J. A. *Additivity Methods for Prediction of Thermochemical Properties. The Laidler Method Revisited. 2. Hydrocarbons Including Substituted Cyclic Compounds*. *J. Chem. Thermodyn.* **2009**, doi:10.1016/j.jct.2009.06.013.
- [31] Montgomery, Jr., J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *A complete basis set model chemistry. VI. Use of density functional geometries and frequencies*. *J. Chem. Phys.* **1999**, *110*, 2822–2827.
- [32] Sebesta, R. *Programming the World Wide Web*; 4th ed.; Addison Wesley, 2007.
- [33] Nielsen, J. *Discount Usability - Counterpoint: Something Is Better than Nothing*. *IEEE Software* **1997**, *14*, 27–28.
- [34] Nielsen, J. *Designing web usability*; New Riders Publishing: Indianapolis, 2000.

[35] Shneiderman, B. *Designing the user interface: strategies for effective human-computer interaction*; Addison-Wesley: New York, 1997.

[36] Trewavas, A. *A Brief History of Systems Biology*. *The Plant Cell*, **2006**, *18*, 2420–2430.

ANEXO I – Tabela de dados utilizada para realizar a avaliação de usabilidade.

Utilizador	tarefa	Tempo (seg)	Facilidade/tempo de aprendizagem	Memorização	Satisfação
			[0-5]	[0-5]	[0-5]
I1	1	20	3	5	3
	2	40			
	3	30			
I2	1	70	4	5	5
	2	90			
	3	60			
I3	1	50	5	5	5
	2	60			
	3	50			
I4	1	40	4	5	4
	2	70			
	3	60			
I5	1	60	4	5	5
	2	70			
	3	75			
I6	1	70	4	5	5
	2	90			
	3	60			
I7	1	80	5	5	4
	2	100			
	3	70			
I8	1	75	4	5	4
	2	95			
	3	60			
I9	1	45	5	5	5
	2	75			
	3	65			
I10	1	40	5	5	4
	2	35			
	3	50			
I11	1	85	3	4	3
	2	60			
	3	55			
I12	1	70	4	5	4

ThermInfo: Sistema de Informação para Coligir e Apresentar Propriedades Termoquímicas

	2	45			
	3	60			
I13	1	35	4	4	4
	2	50			
	3	65			
I14	1	85	3	5	4
	2	60			
	3	65			
I15	1	80	2	3	3
	2	45			
	3	50			
I16	1	70	3	4	3
	2	95			
	3	75			
I17	1	45	5	5	4
	2	60			
	3	55			
I18	1	85	5	5	3
	2	60			
	3	75			
I19	1	90	4	4	4
	2	70			
	3	55			
I20	1	45	3	3	4
	2	60			
	3	75			
I21	1	100	5	5	5
	2	80			
	3	65			
I22	1	70	5	3	4
	2	90			
	3	75			
I23	1	95	4	4	5
	2	75			
	3	80			
I24	1	60	2	4	4
	2	45			
	3	85			
I25	1	100	4	5	4
	2	80			
	3	95			
Q1	1	97	5	5	4

ThermInfo: Sistema de Informação para Coligir e Apresentar Propriedades Termoquímicas

	2	120			
	3	75			
Q2	1	74	5	5	5
	2	100			
	3	70			
Q3	1	100	4	4	3
	2	80			
	3	80			
Q4	1	70	5	5	5
	2	80			
	3	70			
Q5	1	60	5	4	5
	2	70			
	3	70			
Q6	1	65	5	5	4
	2	80			
	3	70			
Q7	1	40	4	4	4
	2	45			
	3	40			
Q8	1	100	5	5	5
	2	80			
	3	65			
Q9	1	35	3	3	3
	2	40			
	3	35			
Q10	1	70	5	4	5
	2	80			
	3	95			
Q11	1	55	4	4	4
	2	60			
	3	70			
Q12	1	45	2	3	3
	2	35			
	3	40			
Q13	1	50	5	5	4
	2	55			
	3	65			
Q14	1	40	5	5	5
	2	35			
	3	50			

ThermInfo: Sistema de Informação para Coligir e Apresentar Propriedades Termoquímicas

Q15	1	95	3	4	3
	2	70			
	3	55			
Q16	1	65	4	5	5
	2	80			
	3	90			
Q17	1	60	4	2	2
	2	55			
	3	70			
Q18	1	45	5	3	3
	2	70			
	3	65			