



INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

To be or not to be: that is the question

Art.Ask
Perguntas e respostas no domínio da arte

Fábio José Santos Ferreira

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática e de Computadores

Júri

Presidente:	Doutora Maria dos Remédios Vaz Pereira Lopes Cravo
Orientador:	Doutora Maria Luísa Torres Ribeiro Marques da Silva Coheur
Co-orientador:	Doutor João Paulo da Silva Neto
Vogal:	Doutor Bruno Emanuel da Graça Martins

Outubro 2010

Agradecimentos

A todos aqueles que me apoiaram, das mais diversas formas, ao longo deste último ano.

Lisboa, 3 de Novembro de 2010

Fábio Ferreira

Resumo

Este trabalho tem como base o Art.Ask, um sistema de pergunta e resposta anteriormente desenvolvido no L2F, e cujo objectivo é responder a perguntas formuladas em língua natural sobre arte.

O foco deste trabalho vai incidir sobre a interpretação da pergunta.

Este trabalho envolve duas tarefas: a definição de uma taxonomia adaptada ao domínio do Art.Ask, através da introdução de classes mais específicas numa outra taxonomia, definida para problemas de domínio aberto. No contexto desta primeira tarefa foram desenvolvidos dois *corpora* anotados com esta taxonomia.

A segunda tarefa envolve a alteração do método utilizado para a interpretação da pergunta: pretende-se que esta fase passe a utilizar classificadores em vez do método actual (distância de *Levenshtein*), sendo os classificadores considerados neste trabalho o *Support Vector Machines* e o *Voting Model*, tendo este último sido implementado especificamente para este trabalho. Para além disto, são também apresentadas algumas extensões a estes métodos, que podem ser utilizadas para melhorar os resultados obtidos através do recurso a informação sobre o domínio.

Finalmente, são apresentados os resultados da avaliação utilizando os novos métodos, com recurso a vários *corpora*: dois em português (um sobre pintura e outro sobre cinema) e um em inglês (sobre arte), sendo que no caso em que é utilizado o *corpus* sobre cinema (aquele que permite obter os melhores resultados) é obtida uma precisão de 95% para o *Support Vector Machines* e 73% para o *Voting Model*.

Abstract

The baseline for this thesis is Art.Ask, a question answering system previously developed at L2F, designed to answer questions in natural language about arts.

The main goal for this thesis is the interpretation step of Art.Ask.

This thesis comprises two tasks: the definition of a new taxonomy, adapted to the domain of Art.Ask: This taxonomy will be defined by introducing more specific classes in another taxonomy, designed for open domain problems. For this first task, two corpora were developed and annotated using this new taxonomy.

The second task is the improvement of the interpretation algorithm: we will test the system's performance using statistical methods instead of the current algorithm (Levenshtein's distance), being the two considered methods Support Vector Machines and Voting Model. Also, we will present a few extensions that can use information about the domain in order to improve the results obtained by using these methods.

Finally, the results for the evaluation when using the new proposed methods are presented, using several corpora for evaluation: two in portuguese (one in the painting domain and one in the cinema domain) and one in english (in the arts domain), being that when the one about cinema is used (the one that obtains the best results), the precision value for Support Vector Machines is 95% and for Voting Model 73%.

Palavras Chave Keywords

Palavras Chave

Classificação

Língua Natural

Métodos Estatísticos

Pergunta e Resposta

Support Vector Machines

Taxonomia

Voting Model

Keywords

Classification

Natural Language

Question Answering

Statistical Methods

Support Vector Machines

Taxonomy

Voting Model

Índice

1	Introdução	1
1.1	Motivação	1
1.2	Objectivos	2
1.3	Problemática	2
1.3.1	A taxonomia do Art.Ask	2
1.3.2	A interpretação da pergunta no Art.Ask	3
1.4	Organização do documento	3
2	Estado da arte	5
2.1	Art.Ask – Versão inicial	5
2.2	Interpretação de perguntas baseada em métodos probabilísticos e em classificadores	6
2.3	Aplicações de classificadores	8
2.3.1	Análise semântica superficial em árvores sintácticas	8
2.3.1.1	Descrição	8
2.3.1.2	Avaliação	9
2.3.2	Tina	10
2.3.2.1	Descrição	10
2.3.2.2	Avaliação	11
2.3.3	Phoenix	11
2.3.3.1	Descrição	11
2.3.3.2	Avaliação	12
2.3.4	Sergeant Blackwell	13

2.3.4.1	Descrição	13
2.3.4.2	Avaliação	13
2.3.5	Comparação	14
3	Classificação de perguntas no Art.Ask	17
3.1	Introdução	17
3.2	Classificação inicial do Art.Ask	18
3.3	Definição da nova taxonomia	19
3.4	Descrição da taxonomia	20
3.4.1	Subclasses de DESCRIPTION	20
3.4.2	Subclasses de ENTITITY	21
3.4.3	Subclasses de HUMAN	24
3.4.4	Subclasses de LOCATION	25
3.4.5	Subclasses de NUMERIC	27
3.5	Questões em aberto	28
3.6	Criação de corpus	30
4	Interpretação de perguntas no Art.Ask	31
4.1	Classificadores considerados	31
4.1.1	Support Vector Machines	31
4.1.2	Voting Model	31
4.2	Extensões aos classificadores	31
4.2.1	Restrições no valor de atributos	31
4.2.2	Restrições baseadas em keywords	32
4.2.3	Utilização	33
5	Avaliação	35
5.1	Medidas utilizadas	35

5.1.1	Precisão	35
5.1.2	Cobertura	35
5.1.3	Medida-F	36
5.2	Corpora utilizados	36
5.3	Avaliação da versão inicial	36
5.4	Método utilizado	37
5.5	Resultados	38
5.5.1	Corpus sobre arte	38
5.5.1.1	Determinação do limiar e da medida-F	38
5.5.1.2	Análise de erros	38
5.5.1.3	Comparação com Support Vector Machines	38
5.5.2	Corpus sobre cinema	39
5.5.2.1	Avaliação inicial	39
5.5.2.2	Determinação das keywords	40
5.5.2.3	Análise de erros	41
5.5.3	Corpus sobre pintura	41
5.5.3.1	Determinação do limiar e da medida-F	41
5.5.3.2	Análise de erros	42
6	Conclusões e trabalho futuro	47
6.1	Contribuições	47
6.2	Conclusões	47
6.3	Trabalho futuro	48
6.3.1	Extracção de informação	48
6.3.2	Interactividade	48
6.3.3	Classificação por níveis	49

I Apêndices	53
A tabelas	55

Lista de Figuras

2.1	Arquitectura da versão inicial do Art.Ask	5
2.2	Interpretação da pergunta na versão inicial do Art.Ask	6
2.3	Interpretação da pergunta no Art.Ask com recurso a classificadores	6

Lista de Tabelas

2.1	Tabela de probabilidades	8
2.2	Resultados da avaliação da análise semântica superficial	10
2.3	Resultados da avaliação do Sergeant Blackwell	14
2.4	Resultados da avaliação dos vários sistemas	14
5.1	Avaliação do <i>Voting Model</i> entre 0,1 e 1,0	38
5.2	Avaliação do <i>Voting Model</i> na proximidade de 0,233	38
5.3	Resultados da avaliação por tipo de pergunta para o <i>corpus</i> sobre arte	39
5.4	Comparação entre os dois métodos utilizados para a avaliação do <i>corpus</i> em inglês	39
5.5	Comparação entre as duas medidas definidas para a primeira extensão quando aplicada ao <i>Voting Model</i> e com <i>Support Vector Machines</i> sem extensão	40
5.6	Regras utilizadas na segunda extensão proposta	43
5.7	Comparação o <i>Voting Model</i> com a primeira extensão e ambas as extensões	43
5.8	Resultados da avaliação por tipo de pergunta para o <i>corpus</i> sobre cinema	44
5.9	Avaliação do <i>Voting Model</i> entre 0,1 e 1,0	44
5.10	Avaliação do <i>Voting Model</i> na proximidade de 0,422	44
5.11	Resultados da avaliação por tipo de pergunta para o <i>corpus</i> sobre pintura	45
A.1	Constituição de cada um dos conjuntos utilizados na avaliação (<i>corpus</i> sobre pintura)	56
A.2	Constituição de cada um dos conjuntos utilizados na avaliação (<i>corpus</i> sobre arte)	57
A.3	Constituição de cada um dos conjuntos utilizados na avaliação (<i>corpus</i> sobre arte)	58

Lista de Acrónimos

ATIS Air Travel Information System

CMU Carnegie Mellon University

1 Introdução

1.1 Motivação

Os sistemas de pergunta e resposta, frequentemente abreviados como sistemas QA (do inglês *Question Answering*) são sistemas cujo objectivo é responder a questões colocadas pelo utilizador em língua natural.

A utilização de língua natural como forma de interacção entre o utilizador e o sistema tem várias vantagens (Dourado, 2009):

- Permite que a comunicação seja feita da mesma forma que seria com um agente humano;
- Serve-se do poder expressivo da língua natural, i.e., permite exprimir de uma forma relativamente simples expressões que em linguagens formais como o SQL seriam bastante complexas.

Estes sistemas estão a tornar-se cada vez mais comuns devido ao aumento da quantidade de informação em formatos digitais que se tem vindo a registar, assim como do número de utilizadores capazes de lhe aceder, maioritariamente através da Internet.

Os sistemas de pergunta e resposta podem ser de domínio aberto e serem capazes de lidar com questões sobre qualquer assunto; ou de domínio fechado e estarem restritos a um domínio específico. Um exemplo deste último tipo é o JaTeDigo (Guimarães, 2007) (cujo domínio é o cinema), desenvolvido anteriormente no L2F.

Estes sistemas frequentemente têm como objectivo o mapeamento de frases em *frames*, i.e., pares atributo-valor. Este mapeamento pode ser compreendido como uma tarefa de classificação. Esta perspectiva é abordada com algum detalhe em (Bhagat et al., 2005), demonstrado, entre outros métodos, o *Voting Model*. No entanto, para todos os métodos apresentados existe uma etapa final na qual são usadas heurísticas para escolher a *frame* mais provável.

De forma a classificar as perguntas, os sistemas de pergunta e resposta necessitam de recorrer a uma taxonomia (um conjunto de classes). Nos sistemas de domínio aberto a taxonomia deve abranger

o máximo de conceitos possível de uma forma mais superficial. Por outro lado, quando o domínio é fechado é conveniente que a taxonomia utilizada contenha classes mais específicas para o domínio a ser tratado.

O Art.Ask enquadra-se no âmbito dos sistemas de resposta e resposta de domínio fechado e tem como ponto de partida um trabalho anteriormente realizado no L2F, cujo domínio inicial é a pintura (dados sobre quadros e pintores), mas pretende-se que este seja estendido de forma a cobrir todos os campos da arte.

Este trabalho inclui uma implementação do *Voting Model* de acordo com o descrito em (Bhagat et al., 2005), assim como duas extensões que possibilitam a selecção da *frame* mais provável com recurso a informação sobre o domínio e foi aplicado a dois *corpora*: um sobre cinema e outro sobre arte, tendo este último sido estendido no âmbito deste trabalho.

1.2 Objectivos

Este trabalho centra-se em dois objectivos principais:

- Definição de uma taxonomia a ser utilizada na classificação de perguntas. Esta taxonomia será definida através da extensão da taxonomia de Li e Roth e consistirá numa extensão a esta, sendo adicionadas subclasses às classes já existentes. Estas novas classes serão definidas tendo em conta o domínio do Art.Ask e, como tal, a taxonomia não será a mais indicada para ser utilizada em problemas num domínio diferente;
- Teste de vários métodos para a interpretação da pergunta, com especial incidência em classificadores (nomeadamente *Support Vector Machines* e *Voting Model*). Para além destes métodos, serão também apresentadas e discutidas algumas extensões que podem ser utilizadas para melhorar os resultados obtidos com estes métodos.

1.3 Problemática

1.3.1 A taxonomia do Art.Ask

Uma taxonomia é um conjunto de classes utilizada, neste contexto, para classificar perguntas quanto à sua semântica, i.e., se a pergunta se refere a uma data de nascimento, local de nascimento, local de morte, etc.

Neste trabalho vai ser introduzida no Art.Ask uma nova taxonomia, organizada numa estrutura hierárquica (baseada na taxonomia definida por Li e Roth (Li & Roth, 2006)). Este tipo de estruturas tem a vantagem de permitir que a classificação seja feita de uma forma gradual, i.e., inicialmente é apenas atribuída à pergunta a classe do nível mais elevado da hierarquia, em seguida é-lhe atribuída uma classe do segundo nível (subclasse da primeira) e assim sucessivamente.

1.3.2 A interpretação da pergunta no Art.Ask

A utilização de interfaces em língua natural permite facilitar a utilização do sistema pelo utilizador, pois desta forma a comunicação é feita da mesma forma que seria feita com outro utilizador. Por outro lado, a implementação do sistema torna-se mais complexa, porque passa a ser necessário lidar com problemas que não existem em linguagens formais como o SQL (ambiguidade, várias formas possíveis de formular a mesma pergunta, etc.).

Uma interface em língua natural pode ser construída de forma a aceitar apenas perguntas completas (como “Quem pintou O Grito?”) ou também sequências de palavras isoladas (por exemplo, “autor Grito”). A primeira abordagem é a utilizada no Art.Ask, devido ao facto de facilitar a interacção do utilizador com o sistema (apesar de ser mais susceptível a ambiguidade).

Como a implementação de uma gramática para um idioma natural, ainda que restrita a perguntas, seria demasiado complexa, o Art.Ask determina qual o tipo de pergunta introduzida pelo utilizador através do cálculo da distância de edição entre a pergunta dada e um conjunto predefinido de perguntas.

Neste trabalho, o objectivo da interface em língua natural é mapear uma frase (pergunta) num conjunto de pares atributo-valor (*frames*) e este assunto será abordado como uma tarefa de classificação, indo ser testada a interpretação da pergunta com recurso a classificadores, como por exemplo o *Voting Model*.

Para além disto, serão apresentadas duas extensões que permitem melhorar os resultados obtidos com base em informação sobre o domínio.

1.4 Organização do documento

No capítulo 2 são descritos e analisados alguns sistemas que utilizam classificadores para processar informação, assim como é feita uma descrição mais detalhada de um dos métodos a ser utilizados nesta tese, o *Voting Model*.

No capítulo 3 é analisada a tarefa de classificação de perguntas no Art.Ask, sendo descrita a forma como é feita na versão inicial, são apresentadas as alterações a fazer e discutidos os pontos que continuam em aberto.

No capítulo 4 é discutida a forma como a pergunta do utilizador é interpretada, sendo apresentada a forma como é feita na versão inicial, assim como os classificadores a ser utilizados, tal como as extensões.

No capítulo 5 é apresentada a avaliação dos métodos apresentados no capítulo 4, com e sem o recurso às extensões e utilizando vários *corpora*. Neste capítulo é também feita a análise de erros.

Finalmente, no capítulo 6 são apresentadas as contribuições e conclusões deste trabalho, assim como o trabalho a desenvolver no futuro.

Estado da arte

2.1 Art.Ask – Versão inicial

O Art.Ask tem a arquitectura representada na Figura 2.1.

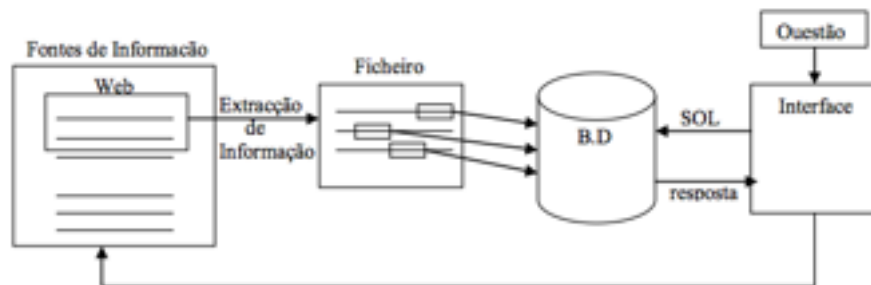


Figura 2.1: Arquitectura da versão inicial do Art.Ask

Este trabalho não prevê alterações à arquitectura global do Art.Ask, existindo apenas alterações no componente “Interface”.

Na versão inicial do Art.Ask a interpretação da pergunta é feita de uma forma bastante simples, sendo calculada a distância de *Levenshtein* entre a pergunta colocada pelo utilizador e as perguntas de um conjunto de perguntas-tipo. A pergunta do utilizador é então considerada equivalente à pergunta do conjunto de perguntas-tipo que estiver à menor distância, sendo-lhe atribuída a mesma classificação. No entanto, este método pode apresentar resultados incorrectos com alguma facilidade. Por exemplo, a pergunta “Onde nasceu Claude Monet?” está à distância 1 de “Onde morreu Claude Monet?” (substituição de uma palavra) e à distância 4 de “Onde é que nasceu o pintor Claude Monet?” (remoção de quatro palavras).

Pretende-se que esta fase de interpretação passe a utilizar classificadores, como por exemplo *Support Vector Machines* ou *Voting Model*. A vantagem desta alteração é que o conhecimento do sistema passa a ser aprendido através de exemplos, deixando de ser necessária a introdução manual de regras ao tentar reutilizar o sistema num novo domínio (David et al., 1996).

As figuras 2.2 e 2.3 ilustram as diferenças entre a interpretação da pergunta sem e com recurso a classificadores: no primeiro caso, a pergunta do utilizador é comparada com todas as perguntas-tipo com o objectivo de encontrar uma equivalente. No segundo caso, as perguntas do conjunto de teste apenas são utilizadas para construir um modelo probabilístico, nunca existindo comparação entre a pergunta do utilizador e as perguntas do conjunto de teste.

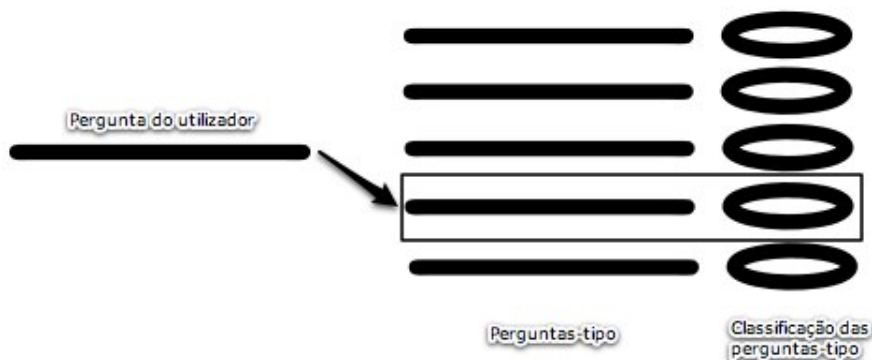


Figura 2.2: Interpretação da pergunta na versão inicial do Art.Ask

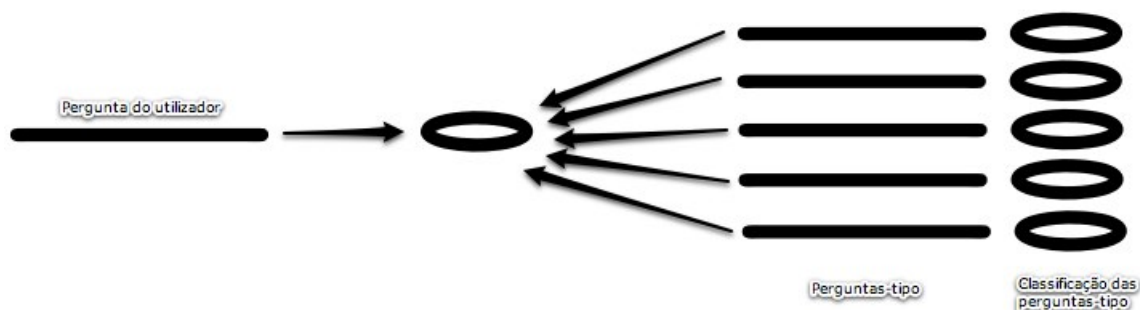


Figura 2.3: Interpretação da pergunta no Art.Ask com recurso a classificadores

Finalmente, com o objectivo de tentar melhorar os resultados obtidos com a aplicação destes métodos, serão apresentadas algumas extensões que podem ser utilizados em conjunto com eles.

2.2 *Interpretação de perguntas baseada em métodos probabilísticos e em classificadores*

Para além do método actualmente utilizado, com base na distância de Levenshtein, a interpretação da pergunta pode ser feita utilizando classificadores. Um exemplo deste tipo de métodos é o *Voting Model*.

O *Voting Model* descrito em (Bhagat et al., 2005) é um método probabilístico baseado em contagens que representa a probabilidade de ser produzido um par atributo-valor, (*frame*) como *output*, dado um certo *input*. São utilizados por este método três modelos: unigramas, bigramas e trigramas.

Seja $C(f_i | w_j)$ o número de vezes que a frase aparece como *output* dado que o unigrama w_j apareceu algures no *input*. De forma análoga, representemos por $C(f_i | w_{j-1}w_j)$ o número de vezes que f_i aparece como *output* dado que no *input* apareceu o bigrama $w_{j-1}w_j$ e $C(f_i | w_{j-2}w_{j-1}w_j)$ o análogo para trigramas.

Para definir a probabilidade de um dado n-grama corresponder a uma *frame*, i.e., para treinar o sistema, utilizam-se as seguintes equações:

$$P(f_i | w_j) = \frac{C(f_i | w_j)}{\sum_{k=1}^n C(f_k | w_j)} \quad (2.1)$$

$$P(f_i | w_{j-1}w_j) = \frac{C(f_i | w_{j-1}w_j)}{\sum_{k=1}^n C(f_k | w_{j-1}w_j)} \quad (2.2)$$

$$P(f_i | w_{j-2}w_{j-1}w_j) = \frac{C(f_i | w_{j-2}w_{j-1}w_j)}{\sum_{k=1}^n C(f_k | w_{j-2}w_{j-1}w_j)} \quad (2.3)$$

Por exemplo, considere-se o seguinte conjunto de frases em que W_i representa uma palavra, e respectivas *frames* para treino do sistema (exemplo adaptado de (Dourado, 2009)):

- $W_1W_2W_3 \rightarrow f_1, f_2$
- $W_2W_3 \rightarrow f_1, f_2$
- $W_4W_3 \rightarrow f_3, f_2$

A partir deste conjunto de treino chegamos às probabilidades representadas na tabela da figura 2.1.

Constituída a tabela resultante do treino, pode-se classificar uma frase de *input* de forma a obter os pares atributo-valor que constituem as *frames* correspondentes, através da seguinte fórmula:

$$W_i(f_i) = \sum_{j=1}^n P(f_i | w_j) + \sum_{j=2}^n P(f_i | w_{j-1}w_j) + \sum_{j=3}^n P(f_i | w_{j-2}w_{j-1}w_j) \quad (2.4)$$

Por exemplo, para o *input* W_3W_2 temos:

Unigramas	Bigramas	Trigramas
$P(f_1 W_1) = 0.5$	$P(f_1 W_1W_2) = 0.5$	$P(f_1 W_1W_2W_3) = 0.5$
$P(f_1 W_2) = 0.5$	$P(f_1 W_2W_3) = 0.5$	$P(f_2 W_1W_2W_3) = 0.5$
$P(f_1 W_3) = 0.34$	$P(f_1 W_4W_3) = 0$	$P(f_3 W_1W_2W_3) = 0$
$P(f_1 W_4) = 0$	$P(f_2 W_1W_2) = 0.5$	
$P(f_2 W_1) = 0.5$	$P(f_2 W_2W_3) = 0.5$	
$P(f_2 W_2) = 0.5$	$P(f_2 W_4W_3) = 0.5$	
$P(f_2 W_3) = 0.5$	$P(f_3 W_1W_2) = 0$	
$P(f_2 W_4) = 0.5$	$P(f_3 W_2W_3) = 0$	
$P(f_3 W_1) = 0$	$P(f_3 W_4W_3) = 0.5$	
$P(f_3 W_2) = 0$		
$P(f_3 W_3) = 0.17$		
$P(f_3 W_4) = 0.5$		

Tabela 2.1: Tabela de probabilidades

- $W_t(f_1) = \frac{5}{6}$
- $W_t(f_2) = 1$
- $W_t(f_3) = \frac{1}{6}$

Após calcular estas probabilidades o objectivo é construir a *frame* dos pares atributo-valor com maior probabilidade. Para isso é utilizado um procedimento com três níveis.

1. Uma heurística baseada no pressuposto que dois pares atributo-valor com o mesmo atributo raramente ocorrem na mesma *frame*;
2. Uma heurística baseada no pressuposto que dois pares atributo-valor com o mesmo valor raramente ocorrem na mesma *frame*;
3. Um *cutoff* que selecciona apenas os candidatos de maior valor.

As heurísticas utilizadas durante esta fase estão dependentes do domínio da aplicação.

2.3 Aplicações de classificadores

2.3.1 Análise semântica superficial em árvores sintácticas

2.3.1.1 Descrição

Esta experiência é descrita em (Pradhan et al., 2004) e o seu objectivo é a identificação e classificação de funções semânticas (i.e., agente, paciente, etc.) em árvores sintácticas. O *corpus* utilizado para esta

experiência foi o PropBank¹, um *corpus* com cerca de 300000 palavras no qual as funções semânticas foram previamente etiquetadas.

Em relação à classificação, o método utilizado foi uma versão do *Support Vector Machines* adaptado para ser utilizado em problemas não-binários. Através de uma abordagem “um contra todos”, na qual k classificadores são utilizados para classificar cada classe contra as restantes (esta abordagem é semelhante à testada no âmbito do Art.Ask). para além disto, e com o objectivo de reduzir o tempo necessário para a experiência, o processo de treino foi dividido em duas fases:

- Como a maioria dos nós numa árvore são `null` (i.e., não contém informação semântica), o conjunto de treino é classificado inicialmente com um classificador “`null` contra não-`null`”;
- Os nós classificados como não-`null` são treinados com os classificadores “um contra todos”.

Ao utilizar esta estratégia a maioria dos nós (aqueles classificados como `null`) apenas serão classificados pelo primeiro classificador, reduzindo drasticamente o tempo necessário para o treino.

Em relação às features utilizadas para o *Support Vector Machines*, esta experiência utilizou as seguintes:

- O predicado correspondente à árvore;
- O caminho (movimento ao longo da árvore) até ao nó seguinte;
- O tipo de sintagma do constituinte (e.g., NP ou PP);
- Se o sintagma está antes ou depois do predicado;
- Se o predicado está na voz activa ou passiva;
- A *headword* do sintagma;
- A regra sintáctica utilizada para expandir o nó (e.g., $VP \rightarrow VBD PP$).

2.3.1.2 Avaliação

Esta experiência avaliou três tarefas diferentes:

- Identificação de argumentos: determinação de quais as sequências de palavras que representam um argumento;

¹<http://verbs.colorado.edu/~mpalmer/projects/ace.html>

- Classificação de argumentos: Atribuição da classe correcta a cada argumento;
- Uma combinação das duas anteriores.

Para cada uma destas tarefas o *corpus* foi avaliado de acordo com dois métodos: um em que apenas são considerados os argumentos principais (os argumentos verbais descritos anteriormente, como o agente ou o paciente) e outro no qual também são considerados outros argumentos como os locativos e os temporais.

Os resultados obtidos estão descritos na tabela 2.3.1.2.

	Tarefa	Precisão	Cobertura	Medida-F
Argumentos principais	Identificação	0,909	0,898	0,904
	Classificação	0,879		
	Ambos	0,833	0,785	0,808
Todos os argumentos	Identificação	0,947	0,901	0,923
	Classificação	0,914		
	Ambos	0,884	0,841	0,862

Tabela 2.2: Resultados da avaliação da análise semântica superficial

Não é apresentado um valor para a medida-F nos problemas de classificação porque a todos os nós é atribuída uma classe (eventualmente `null`).

2.3.2 Tina

2.3.2.1 Descrição

Tina é um sistema de língua natural descrito em (Seneff, 1992) e desenhado para reconhecer texto falado e identificar a sua semântica. Em vez de uma abordagem estatística ao processo de classificação, este sistema utiliza uma abordagem baseada numa lista de regras, que são internamente convertidas numa rede probabilística, i.e., um grafo em que a probabilidade de cada transição é dada pela análise de um conjunto de treino, e.g., se um símbolo é expandido pela regra A quatro vezes no conjunto de treino e pela regra B apenas uma vez, a transição correspondente à regra A tem probabilidade 0,8 e a correspondente à regra B tem probabilidade 0,2.

O Tina analisa as frases através de uma gramática probabilística livre de contexto (na qual cada regra tem uma probabilidade associada) utilizando uma abordagem conhecida como “bigramas em camadas” (Goblirsch, 1996) para calcular a probabilidade de cada frase. Um bigrama em camadas é um bigrama no qual cada um dos dois nós é uma palavra, uma classe ou outro bigrama, obtendo-se assim

uma estrutura recursiva. Desta forma, cada símbolo não-terminal é expandido para dois símbolos (terminais ou não). Esta abordagem tem algumas vantagens sobre os bigramas comuns:

- Permitem a modelação de dependências para além das palavras (ou classes) adjacentes;
- Fornecem um maior controlo sobre o espaço de pesquisa, pois o alisamento de parâmetros não possibilita o aparecimento de sequências de palavras inválidas noutra bigrama;
- Facilitam a integração num módulo de língua natural, pois o reconhecedor pode fornecer uma análise parcial ao atribuir a um nó a mesma etiqueta que foi atribuída ao nó a partir do qual foi expandido.

O Tina foi integrado num sistema de diálogo completo chamado Voyager, projectado para dar indicações sobre viagens, assim como informações sobre restaurantes e outros pontos de interesse. Alguns exemplos de frases no domínio do Voyager são:

- “Do you know the most direct route to Broadway Avenue from here?”
- “What kind of restaurant is located around Mount Auburn in Kendall Square of East Cambridge?”

2.3.2.2 Avaliação

De forma a obter dados para treino e teste, foram seleccionados alguns indivíduos para que tentassem utilizar o sistema. Cerca de 500 frases (no domínio do Voyager) foram recolhidas desta forma.

Deste conjunto de 5000 frases foi criado um conjunto de treino com 3312 frases. A partir destas, foram desenvolvidas algumas regras para as quais o sistema não estava preparado inicialmente.

Esta experiência resultou numa cobertura de 76%, o que significa que 76% das frases recebeu uma interpretação.

2.3.3 Phoenix

2.3.3.1 Descrição

O Phoenix (Ward, 1990) é um sistema desenvolvido pela CMU projectado para compreender texto falado no âmbito do ATIS (Seneff et al., 1991), uma tarefa cujo objectivo é a recolha de informação de uma base de dados sobre viagens aéreas, como horários e aeronaves. Este sistema utiliza um analisador baseado em *frames* para processar o *input*, no qual as *frames* são utilizadas para representar a informação

semântica e são preenchidas através de uma gramática semântica, i.e., uma gramática que converte partes do discurso em conceitos semânticos. esta gramática permite que um sintagma seja interpretado de forma isolada, para que seja possível a análise em situações em que não exista uma frase completa. Nesse caso, apenas é preenchida a parte da *frame* que for possível. Para além disto, é também possível descartar elementos desnecessários, como ruído.

Em seguida é apresentado um exemplo da *frame* correspondente à frase “*Show me...ah...I want to see all flights to Denver after two PM.*”:

```
[list]: list
[flights]: flights
[arrive_loc]: ``DEN``
[depart_loc]: ``PIT``
[depart_time_range]: 1400 2400
```

2.3.3.2 Avaliação

O vocabulário reconhecido pelo Phoenix consiste num conjunto de 484 palavras. O conjunto de teste é composto por 93 elocuições “classe-A”² recolhidas de 5 oradores distintos.

O Phoenix foi avaliado de duas formas: através da análise do discurso e através da análise directa das transcrições.

Um teste inicial resultou em 39% de frases interpretadas correctamente para a análise do discurso e 48% para as transcrições. Sendo um valor baixo, os resultados foram analisados e confirmou-se que a maioria dos erros era de um de três tipos:

- **Datas:** As datas geradas pelo Phoenix eram relativas ao tempo de execução e não ao tempo de recolha do *corpus* (e.g., palavras como “hoje” estavam a ser expandidas para a data de recolha do *corpus* em vez da data do teste);
- **Abreviaturas:** O sistema estava a gerar abreviaturas em alguns casos, em vez do texto completo;
- **Informação sobre viagens de ida e volta:** O sistema calculava a informação sobre viagens de ida e volta de uma forma incorrecta.

²Uma elocução “classe-A” é uma elocução que pode ser interpretada sem contexto adicional.

Após a correcção destes erros, o teste foi repetido e desta vez resultou em 42% para o discurso e 65% para a transcrição.

Uma outra tarefa na avaliação do Phoenix consistiu na avaliação do reconhecimento do *input*: esta tarefa resultou numa taxa de erros em elocuições de 97% (97% das elocuições foram analisadas incorrectamente) e de 44% para palavras. Apesar de se poder considerar que estes valores são elevados, o sistema conseguiu interpretar correctamente 75% das elocuições (10% foram analisados correctamente, mas interpretados incorrectamente, obtendo-se assim 65%), mostrando que até certo ponto o analisador é capaz de “corrigir” erros.

2.3.4 Sergeant Blackwell

2.3.4.1 Descrição

O Sergeant Blackwell (Leuski, Patel, & Traum, 2006) (Leuski, Pair, et al., 2006) foi projectado para ser usado como quiosque informativo numa conferência do exército. Este sistema recolhe *input* do utilizador através de um microfone e utiliza o motor de reconhecimento de voz Sonic, desenvolvido pela Universidade de Colorado (Salor et al., 2007) para o processar. É capaz de responder a um vasto conjunto de perguntas, tal como informação sobre a conferência, a sua identidade, a tecnologia utilizada no seu desenvolvimento e algumas questões diversas, como “Que horas são?”.

O processo de classificação utilizado no Sergeant Blackwell utiliza um classificador multi-classe baseado em *Support Vector Machines* que utiliza o algoritmo $tf-idf^3$ e um *kernel* exponencial (a combinação que permite obter os resultados mais satisfatórios).

O Sergeant Blackwell diverge da maioria dos sistemas semelhantes no sentido em que em vez de procurar uma resposta numa base de dados, apenas responde a um número fixo de perguntas (60).

2.3.4.2 Avaliação

O Sergeant Blackwell foi testado com três algoritmos de classificação distintos: *Support Vector Machines*, *Mono-lingual Information Retrieval* e *Cross-lingual Information Retrieval* (Ponte & Croft, 1998).

As precisões para cada um dos três algoritmos podem ser encontradas na tabela 2.3.4.2.

Podemos verificar que existe uma vantagem nos algoritmos que utilizam recuperação de

³term frequency – inverse document frequency

Algoritmo	Precisão
SVM	0,5313
MIR	0,5780
CIR	0,6199

Tabela 2.3: Resultados da avaliação do Sergeant Blackwell

informação (MIR e CIR) em relação àqueles que utilizam *Support Vector Machines*.

De forma a suportar os resultados obtidos com este teste, ele foi repetido com sete outros sistemas (de outros projectos semelhantes). A tabela 2.3.4.2 mostra a precisão média para cada algoritmo.

Algoritmo	Precisão
SVM	0,4816
MIR	0,4893
CIR	0,5314

Tabela 2.4: Resultados da avaliação dos vários sistemas

Apesar destes resultados serem inferiores aos obtidos para o Sergeant Blackwell, os resultados dos algoritmos que utilizam recuperação de informação continuam a ser melhores que o resultado obtido com *Support Vector Machines*.

Um aspecto relevante desta experiência que não é visível nestas tabelas é que as abordagens baseadas em recuperação de informação funcionam melhor em problemas com mais classes e mais dados de treino.

2.3.5 Comparação

Todos sistemas descritos têm como objectivo o processamento semântico. No entanto, existem alguns pontos nos quais eles divergem:

- O problema de análise semântica superficial tem como objectivo identificar funções semânticas numa árvore sintáctica e utiliza *Support Vector Machines*;
- O Tina tem como objectivo o reconhecimento da semântica em texto (dado na forma de discurso falado) e utiliza uma abordagem baseada em regras;
- O Phoenix tem como finalidade obter informação de uma base de dados através de uma interface em língua natural com o utilizador. A classificação do *input* do utilizador é feito com recurso a uma gramática semântica;

- O Sergeant Blackwell tem como objectivo fornecer uma resposta a um conjunto limitado de perguntas, utilizando, entre outros métodos, *Support Vector Machines* para identificar qual das respostas é a mais adequada.

Podemos dividir os sistemas analisados em dois grupos: os que pretendem avaliar a semântica de uma frase dada como *input* (problema de análise superficial e Tina) e os que têm como objectivo responder a pedidos do utilizador (Phoenix e Sergeant Blackwell). O Art.Ask enquandra-se no segundo grupo.

Dos dois sistemas incluídos no segundo grupo, o Sergeant Blackwell diverge ligeiramente do Art.Ask porque em vez de uma base de dados com respostas apenas responde a um conjunto limitado de perguntas. O Phoenix e o HARC são, portanto, os sistemas que mais se aproximam do Art.Ask.

Classificação de perguntas no Art.Ask

3.1 Introdução

Na versão actual do Art.Ask, as perguntas podem ser classificadas de acordo com um conjunto predefinido de tipos. Estes tipos são completamente independentes entre si. No entanto, algumas perguntas podem ser respondidas com vários níveis de precisão (e.g., a pergunta “Onde se situa a Mona Lisa?” pode ter como resposta o museu onde se encontra o quadro – “Louvre” – ou de uma forma mais geral a cidade – “Paris”). Para simplificar a forma como este tipo de situações é resolvida, os tipos de perguntas passarão a estar integrados numa taxonomia (uma hierarquia de tipos utilizados para classificar conceitos).

Esta nova abordagem tem algumas vantagens sobre a anterior, sendo a mais relevante a possibilidade de agrupar várias classes numa só classe, num nível superior da hierarquia. Com a utilização desta hierarquia, será possível usar um classificador inicial que atribui à pergunta uma classe do nível superior; em seguida utilizar um outro classificador para obter no segundo nível uma subclasse da classe obtida no nível anterior; e assim sucessivamente até se atingir uma classe sem subclasses.

Por exemplo, vamos supor que temos uma classe para representar países (COUNTRY) e uma outra para representar cidades (CITY). Como ambas as classes se referem a localizações, é possível definir uma superclasse que englobe as duas anteriormente descritas (LOCATION).

A base desta taxonomia será a taxonomia de Li e Roth, descrita em (Li & Roth, 2006), definida no âmbito dos sistemas de pergunta e resposta de domínio aberto, tendo sido definida com base num *corpus* formado a partir de perguntas do TREC. A utilização desta taxonomia como base tem a vantagem de permitir reutilizar trabalho já feito adaptando-o a um novo contexto. Uma alternativa poderia ser a utilização da taxonomia da Webclopedia¹, no entanto esta taxonomia tem a desvantagem de ser mais horizontal, i.e., está definida de forma menos hierárquica. Além disto, inclui classes demasiado específicas para o domínio em causa (e.g., C-PROPER-SPORTS-TEAM para equipas desportivas) em comparação com a taxonomia de Li e Roth, que apenas inclui classes genéricas.

¹<http://www.isi.edu/natural-language/projects/webclopedia>

3.2 Classificação inicial do Art.Ask

Na versão do Art.Ask considerada como ponto de partida para este trabalho, a taxonomia utilizada é bastante simples: é constituída apenas por um conjunto de classes correspondentes a um conjunto predefinido de tipos, tendo sido apenas considerada uma parte do domínio do Art.Ask (mais especificamente o subdomínio da pintura).

As classes consideradas nesta taxonomia inicial são:

- **Quem** – Utilizada quando se pretende obter o autor de uma obra (e.g., “Quem é o autor da \$OBRA?”);
- **Onde** – Utilizada quando se pretende obter a localização (museu ou cidade) onde está uma obra (e.g., “Onde se encontra a \$OBRA?”);
- **Tamanho** – Utilizada quando se pretende obter as dimensões de uma obra (e.g., “Quais são as dimensões da \$OBRA?”);
- **Material** – Quando se pretende obter qual o material usado na concepção da obra (e.g., “Qual o material usado na obra \$OBRA?”);
- **Data** – Quando se pretende obter a data de concepção de uma obra (e.g., “Em que ano foi feito o \$OBRA?”);
- **Estilo** – Quando se pretende obter o movimento artístico associado a uma obra (e.g., “Qual o estilo da \$OBRA?”);
- **Data_nascimento** – Quando se pretende obter a data de nascimento de um artista (e.g., “Quando nasceu o pintor \$ARTISTA?”);
- **Data_morte** – Quando se pretende obter a data de morte de um artista (e.g., “Quando morreu \$ARTISTA?”);
- **Onde_nasceu** – Quando se pretende obter a cidade de nascimento de um artista (e.g., “Onde nasceu \$ARTISTA?”);
- **Pais_nascimento** – Quando se pretende obter o país de nascimento de um artista (e.g., “Em que país nasceu \$ARTISTA?”);
- **Onde_morreu** – Quando se pretende obter a cidade de morte de um artista (e.g., “Onde morreu \$ARTISTA?”);

- Pais_morte – Quando se pretende obter o país de morte de um artista (e.g., “Em que país morreu \$ARTISTA?”);
- Principais_trabalhos – Quando se pretende obter as principais obras de um artista (e.g., “Quais os principais trabalhos de \$ARTISTA?”);
- Nacionalidade – Quando se pretende obter a nacionalidade de um artista (e.g., “Qual a nacionalidade de \$ARTISTA?”);
- Estilo_pintor – Quando se pretende obter o movimento artístico associado a um artista (e.g., “Qual o estilo praticado por Nome_Completo – Quando se pretende obter o nome completo de um artista (e.g. “Qual o nome completo de \$ARTISTA?”);
- Ocupacao – Quando se pretende obter a ocupação de um artista (e.g., “Em que área se destacou \$ARTISTA?”).

3.3 Definição da nova taxonomia

A taxonomia original de Li e Roth é composta por dois níveis. O primeiro, mais abrangente, contém seis classes (ABBREVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION e NUMERIC) e o segundo contém cinquenta, das quais cada uma corresponde a uma (e apenas uma) classe no primeiro nível. Em qualquer um dos níveis as classes são disjuntas (i.e., não devem existir sobreposições entre elas).

Por exemplo, a pergunta “*What Canadian city has the largest population?*” é do tipo LOCATION:CITY e a pergunta “*Who was the first woman killed in the Vietnam War?*” corresponde ao tipo HUMAN:INDIVIDUAL.

De forma a manter a compatibilidade com a taxonomia de Li e Roth, a nova taxonomia tem as mesmas classes que esta nos dois primeiros níveis, sendo as alterações realizadas apenas nos níveis seguintes. Este facto é particularmente visível na existência de subclasses de ENTITY:OTHER e LOCATION:OTHER em vez de ENTITY e LOCATION, mantendo assim a equivalência nos dois níveis superiores da hierarquia.

A definição das novas classes a figurar na nova taxonomia foi feita com base em dois *corpora*:

- Um *corpus* composto por perguntas aceites pela versão inicial do Art.Ask, com 624 perguntas em português sobre pinturas e dados biográficos de pintores. As classes definidas a partir deste *corpus* foram mapeadas directamente nos tipos da taxonomia inicial do Art.Ask (e.g., Onde, Data, etc.). Esta decisão permite em simultâneo manter informação sobre a classe a que cada entidade

pertence e distinguir duas perguntas com tipos semelhantes (e.g., datas de nascimento e morte) – referido como *corpus* sobre pintura;

- Um *corpus* composto por um conjunto de perguntas sobre arte e literatura, com 233 perguntas em inglês, retirado do Trivia². As perguntas deste *corpus* eram de um âmbito mais geral (não incluíam perguntas apenas sobre pintura) e tinham uma estrutura mais irregular, não estando limitadas a nenhum conjunto restrito de tipos – referido como *corpus* sobre arte.

3.4 Descrição da taxonomia

Nesta secção são descritas as classes adicionadas à taxonomia a ser utilizada no Art.Ask, indicando o conceito que identificam, acompanhados de exemplos de perguntas a que corresponde essa classificação.

3.4.1 Subclasses de DESCRIPTION

- DESCRIPTION:DESCRIPTION:SETTING

– Esta classe é utilizada para o contexto de obras, incluindo detalhes como a localização ou a época histórica. Foi definida uma classe SETTING como subclasse de DESCRIPTION:DESCRIPTION para este efeito em vez da utilização das classes LOCATION (para o caso de localizações) e NUMERIC:TIME (para o caso de épocas históricas) porque define algo mais abrangente que estas classes (e.g., uma realidade alternativa), apesar de em algumas situações ser suficiente utilizar uma destas classes. No contexto deste trabalho, a classe SETTING será utilizada sempre que a pergunta se refira ao contexto da obra sem referir se se trata de uma localização ou de uma época, sendo nesse caso utilizadas as classes apropriadas.

– Exemplo: “*What is the fairytale world of C S Lewis’s The Lion, The Witch And The Wardrobe?*”

- DESCRIPTION:DESCRIPTION:SUBJECT

– Esta classe é utilizada para assuntos, podendo referir-se a assuntos de obras (como em “*What is the subject of Kyle Onstott’s Mandingo?*”) ou a outros assuntos que possam ser referidos dentro de obras (como na pergunta “*What subject did Mr Chips teach?*”). Optou-se por não diferenciar os assuntos de obras de outros tipos de assuntos devido à semelhança entre os dois.

- DESCRIPTION:DESCRIPTION:TRAIT

²Um jogo de perguntas de cultura geral, *download* disponível em http://www.eggdroprsus.co.uk/download/myscripts/Trivia_2000v2.3.zip

– Esta classe é utilizada para características de indivíduos (físicas ou psicológicas). Devido à existência de um baixo número de perguntas nesta categoria, optou-se por não definir uma classe para cada um dos tipos.

– Exemplo: *“What facial feature does not feature on the Mona Lisa?”*

- DESCRIPTION:DESCRIPTION:TRAIT:GENDER

– Esta classe é utilizada quando se pretende obter o género de um indivíduo (i.e., masculino ou feminino).

– Exemplo: *“Was Richmal Crompton a man or woman?”*

- DESCRIPTION:REASON:CAUSE

– Esta classe é utilizada para causas de acontecimentos. Decidiu-se criar uma subclasse de DESCRIPTION:REASON para distinguir causas (aquilo que provoca um acontecimento, como na pergunta *“What caused D H Lawrence’s early death?”*) de motivações (o que leva alguém a agir de uma certa forma³).

– Exemplo: *“What caused D H Lawrence’s early death?”*

3.4.2 Subclasses de ENTITY

- ENTITY:ANIMAL:MYTH

– Esta classe é utilizada para criaturas mitológicas. Optou-se por definir esta classe como uma especialização de ENTITY:ANIMAL porque também esta se refere a criaturas vivas e não-humanas.

– Exemplo: *“What is the name of the winged horse in Greek mythology?”*

- ENTITY:ANIMAL:NAME

– Esta classe é utilizada para nomes (próprios) de animais. Apesar de alguns animais serem também personagens (classe HUMAN:INDIVIDUAL:CHARACTER) convencionou-se que todos os nomes de animais (personagens ou não) pertencem à classe

– Exemplo: *“What is the name of Little Orphan Annie’s dog?”* ENTITY:ANIMAL:NAME.

- ENTITY:ANIMAL:REAL

– Esta classe é utilizada para espécies animais reais (por oposição à classe ENTITY:ANIMAL:MYTH que representa criaturas fictícias).

³Devido à inexistência desta situação no corpus não foi definida uma classe para esta situação

- Exemplo: *“What Aesop animal assumed the grapes he couldn’t reach were sour anyway?”*
- ENTITY:CREATIVE:BOOK
 - Esta classe é utilizada para títulos de livros.
 - Exemplo: *“What Herman Hess book gave its name to a rock group?”*
- ENTITY:CREATIVE:BOOK:NOVEL
 - Esta classe é utilizada para títulos de romances.
 - Exemplo: *“What James Clavell novel recounts John Blackthorne’s adventures in 16th century Japan?”*
- ENTITY:CREATIVE:COMIC
 - Esta classe é utilizada para títulos de bandas desenhadas. Apesar de em várias situações uma banda desenhada ser um livro, esta classe não foi considerada uma subclasse de ENTITY:CREATIVE:BOOK para que pudesse também incluir bandas desenhadas publicadas de outra forma (tiras em jornais ou *webcomics*).
 - Exemplo: *“What comic features Dennis the Menace?”*
- ENTITY:CREATIVE:PAINTING
 - Esta classe é utilizada para títulos de pinturas e corresponde ao tipo de pergunta `Principais_trabalhos` da versão inicial do Art.Ask.
 - Exemplo: *“Who painted The Scream?”*
- ENTITY:CREATIVE:PLAY
 - Esta classe é utilizada para títulos de peças de teatro.
 - Exemplo: *“What Anthony Shaffer two man play opens in the living room of Andrew Wykes Norman manor house?”*
- ENTITY:CREATIVE:POEM
 - Esta classe é utilizada para títulos de poemas. Tal como ENTITY:CREATIVE:COMIC, esta classe não foi considerada subclasse de ENTITY:CREATIVE:BOOK pois um poema pode não ser um livro (pode ser apenas parte de um livro).
 - Exemplo: *“What poem begins ‘Twas brillig, and the slithy toves..’?”*
- ENTITY:INSTRUMENT:MUSICAL
 - Esta classe é utilizada para instrumentos musicais.
 - Exemplo: *“What musical instrument did Sherlock Holmes play?”*

- ENTITY:OTHER:MOVEMENT⁴

- Esta classe é utilizada para movimentos artísticos e correntes filosóficas. Decidiu-se que perguntas referindo movimento artísticos (como *“What painting movement was conceived by Pablo Picasso and Georges Braque?”*) e correntes filosóficas (como *“What philosophy is associated with Jean Paul Sartre?”*) seriam do mesmo tipo porque são conceitos semelhantes e este facto permite simplificar a taxonomia.

- ENTITY:OTHER:MOVEMENT:PHILOSOPHY

- Este classe é utilizada para correntes filosóficas.
- Exemplo: *“What philosophy is associated with Jean Paul Sartre?”*

- ENTITY:OTHER:MOVEMENT:PAINTING:PAINTERMOVEMENT

- Esta classe é utilizada para correntes artísticas, quando associadas a um pintor e corresponde ao tipo de pergunta `Estilo_pintor` da versão inicial do Art.Ask.
- Exemplo: *“What painting movement was conceived by Pablo Picasso and Georges Braque?”*

- ENTITY:OTHER:MOVEMENT:PAINTING:PAINTINGMOVEMENT

- Esta classe é utilizada para correntes artísticas, quando associadas a um quadro e corresponde ao tipo de pergunta `Estilo` da versão inicial do Art.Ask.
- Exemplo: *“What painting movement is represented in The Scream?”*

- ENTITY:OTHER:ROCK

- Esta classe é utilizada para nomes de estruturas rochosas.
- Exemplo: *“What rock did Graham Greene write about?”*

- ENTITY:OTHER:ROOM

- Esta classe é utilizada para nomes de salas e outras espaços. Apesar de à primeira vista parecer que esta classe faria mais sentido como subclasse de `LOCATION`, ela foi definida como subclasse de `ENTITY:OTHER` por não se referir a um espaço em concreto, mas sim a um tipo de espaço.
- Exemplo: *“Which room did Napoleon keep the Mona Lisa in?”*

- ENTITY:TECHNIQUE:PAINTING

⁴Como já foi referido anteriormente, esta classe e as seguintes são subclasses de `ENTITY:OTHER` com o objectivo de manter a compatibilidade com a taxonomia original de Li e Roth.

- Esta classe é utilizada para técnicas de pintura. Foi definida uma especialização da classe ENTITY : TECHNIQUE para que fosse possível distinguir técnicas de pintura de outras técnicas (de escultura, por exemplo).
- Exemplo: *“What is Ray Bradbury’s illustrated man illustrated with?”*

3.4.3 Subclasses de HUMAN

- HUMAN : DESCRIPTION : OCCUPATION

- Esta classe é utilizada para nomes de ocupações sejam elas profissões ou passatempos e corresponde ao tipo de pergunta Ocupacao da versão inicial do Art.Ask.
- Exemplo: *“What is Sir Robert Helpman’s field of fame?”*

- HUMAN : GROUP : SURNAME

- Esta classe é utilizada para apelidos, sendo utilizada quer para apelidos de indivíduos (como em *“What is Juliet’s last name?”*), quer como referência a famílias (como em *“What is the family name in the H E Bates story A little of what you fancy?”*). Apesar de no primeiro caso não estarmos na presença de um HUMAN : GROUP mas sim de um HUMAN : INDIVIDUAL, optou-se por utilizar a mesma classe, porque o conceito em questão é o mesmo.

- HUMAN : INDIVIDUAL : ARTIST

- Esta classe é utilizada para nomes de artistas, sendo utilizada apenas quando a pergunta não refere em que campo se notabilizou o artista.
- Exemplo: *“What artist had the surname Buonarrotti?”*

- HUMAN : INDIVIDUAL : ARTIST : COMPOSER

- Esta classe é utilizada para nomes de compositores.
- Exemplo: *“What composer had the forenames Piotr Illyich?”*

- HUMAN : INDIVIDUAL : ARTIST : PAINTER

- Esta classe é utilizada para nomes de pintores e corresponde ao tipo de pergunta Quem da versão inicial do Art.Ask.
- Exemplo: *“What dutch master painted 64 self portraits?”*

- HUMAN : INDIVIDUAL : ARTIST : PAINTER : FULLNAME

- Esta classe é utilizada para nomes completos de pintores e corresponde ao tipo de pergunta Nome_completo da versão inicial do Art.Ask.

– *“What was the full name of Van Gogh?”*

- HUMAN:INDIVIDUAL:ARTIST:SCULPTOR

– Esta classe é utilizada para nomes de escultores.

– Exemplo: *“What sculptor is responsible for The Kiss?”*

- HUMAN:INDIVIDUAL:ARTIST:WRITER

– Esta classe é utilizada para nomes de escritores.

– Exemplo: *“What Greek slave wrote fables?”*

- HUMAN:INDIVIDUAL:CHARACTER

– Esta classe é utilizada para nomes de personagens. Como já referido anteriormente, esta classe apenas é utilizada para personagens humanos, sendo que personagens animais pertencem à classe ENTITY:ANIMAL:NAME.

– Exemplo: *“What Peanuts character has a brother named Spike?”*

- HUMAN:INDIVIDUAL:PHILOSOPHER

– Esta classe é utilizada para nomes de filósofos.

– Exemplo: *“Who is known as the Father of Western Philosophy?”*

3.4.4 Subclasses de LOCATION

- LOCATION:CITY:CITYOFBIRTH

– Esta classe é utilizada para cidades de nascimento de indivíduos e corresponde ao tipo de pergunta Onde_nasceu da versão inicial do Art.Ask.

– Exemplo: *“In which city was Leonardo Da Vinci born?”*

- LOCATION:CITY:CITYOFDEATH

– Esta classe é utilizada para cidades de morte de indivíduos e corresponde ao tipo de pergunta Onde_morreu da versão inicial do Art.Ask.

– Exemplo: *“In which city did Leonardo Da Vinci die?”*

- LOCATION:COUNTRY:COUNTRYOFBIRTH

– Esta classe é utilizada para países de nascimento de indivíduos e corresponde ao tipo de pergunta Pais_nascimento da versão inicial do Art.Ask.

– Exemplo: *“In which country was Leonardo Da Vinci born?”*

- LOCATION:COUNTRY:COUNTRYOFDEATH

- Esta classe é utilizada para países de morte de indivíduos e corresponde ao tipo de pergunta Pais_morte da versão inicial do Art.Ask.

- Exemplo: *“In which country did Leonardo Da Vinci die”*

- LOCATION:COUNTRY:NATIONALITY

- Esta classe é utilizada para nacionalidades de indivíduos e corresponde ao tipo de pergunta Nacionalidade da versão inicial do Art.Ask.

- Exemplo: *“What nationality was Pablo Picasso?”*

- LOCATION:OTHER:ISLAND

- Esta classe é utilizada para nomes de ilhas.

- Exemplo: *“What imaginary island did Sir Thomas More create in a 1516 work?”*

- LOCATION:OTHER:MUSEUM

- Esta classe é utilizada para nomes de museus. É discutível se terá sido correcto colocar esta classe como uma subclasse de LOCATION porque um museu pode ser visto como uma instituição. Dado que a taxonomia proposta não inclui uma classe que representa “Instituição”, optou-se por definir esta classe abaixo de LOCATION na hierarquia.

- Esta classe corresponde ao tipo de pergunta “Onde” da versão inicial do Art.Ask.

- Exemplo: *“What museum provides a home for Whistler’s Mother?”*

- LOCATION:OTHER:PLANET

- Esta classe é utilizada para nomes de planetas (reais ou fictícios).

- Exemplo: *“What planet is taken over by aliens called overlords in Arthur C Clarke’s novel Childhood’s End?”*

- LOCATION:OTHER:SCHOOL

- Esta classe é utilizada para nomes de escolas e universidades. A inclusão desta classe como uma subclasse de LOCATION tem uma motivação análoga à da classe MUSEUM (É uma instituição, mas não existe uma classe na hierarquia para este conceito).

- Exemplo: *“What school does Tom attend in Tom Brown’s School Days?”*

3.4.5 Subclasses de NUMERIC

- NUMERIC:DATE:DATEOFBIRTH
 - Esta classe é utilizada para datas de nascimento de indivíduos e corresponde ao tipo de pergunta `Data_nascimento` da versão inicial do Art.Ask.
 - Exemplo: *“When was Leonardo da Vinci born?”*
- NUMERIC:DATE:DATEOFDEATH
 - Esta classe é utilizada para datas de morte de indivíduos e corresponde ao tipo de pergunta `Data_morte` da versão inicial do Art.Ask.
 - Exemplo: *“When did Leonardo da Vinci die?”*
- NUMERIC:DATE:DATEOFPAINTING
 - Esta classe é utilizada para datas de quadros e corresponde ao tipo de pergunta `Data` da versão inicial do Art.Ask.
 - Exemplo: *“When did Edvard Munch paint The Scream?”*
- NUMERIC:DATE:DAY
 - Esta classe é utilizada para dias.
 - Exemplo: *“In which day was Star Wars: A New Hope released?”*
- NUMERIC:DATE:MONTH
 - Esta classe é utilizada para meses.
 - Exemplo: *“Which month is named after Roman Emperor Julius Caesar?”*
- NUMERIC:DATE:SEASON
 - Esta classe é utilizada para estações do ano (Primavera, Verão, Outono ou Inverno).
 - Exemplo: *“What season is the setting for Shakespeare’s A Midsummer Nights Dream?”*
- NUMERIC:DATE:TIME
 - Esta classe é utilizada para horas.
 - Exemplo: *“What was the time for J L Herlihy’s cowboy?”*
- NUMERIC:DATE:YEAR
 - Esta classe é utilizada para anos.
 - Exemplo: *“In which year was War and Peace published?”*

- NUMERIC:SIZE:DIMENSIONS

- Esta classe é utilizada para dimensões de quadros e corresponde ao tipo de pergunta Tamanho da versão inicial do Art.Ask.
- Exemplo: “Which size is The Scream?”

3.5 Questões em aberto

Esta taxonomia não pretende ser definitiva, podendo ser alterada. Esta necessidade pode surgir caso existam alterações significativas ao domínio (e.g., a introdução de perguntas sobre arquitectura implicaria a introdução da classe HUMAN:INDIVIDUAL:ARCHITECT).

Por outro lado, foram definidas algumas classes às quais apenas corresponde uma pergunta no *corpus* (e.g., à classe ENTITY:FOOD:SEASONING apenas corresponde a pergunta “What seasoning is mentioned more than 30 times in the Bible?”). Devido ao facto de terem uma utilização reduzida a sua necessidade pode ser debatida. No contexto deste trabalho foi definida uma subclasse sempre que uma pergunta pudesse ter um tipo mais específico que o identificado por uma classe já existente. Uma outra hipótese é apenas definir uma subclasse quando existe no *corpus* um número mínimo de perguntas desse tipo (em termos absolutos ou relativos).

Um outro aspecto que pode suscitar dúvidas é a existência de algumas classes que poderiam ter sido definidas com um nível diferente de especificidade:

- A classe DESCRIPTION:DESCRIPTION:SUBJECT poderia ter sido dividida em duas classes: uma para identificar assuntos de obras (como em “What is the subject of Kyle Onstott’s Mandingo?”) e uma outra para os assuntos referidos dentro de obras (como em “What subject did Mr Chips teach?”);
- A classe DESCRIPTION:DESCRIPTION:TRAIT é utilizada para referir características de indivíduos, sejam estas físicas (como em “What facial feature does not feature on the Mona Lisa?”) ou psicológicas (Exemplo inexistente no *corpus* utilizado). Dependendo do domínio em causa, poderia ter interesse a definição de uma classe para cada um deste tipo de características;
- A classe ENTITY:OTHER:MOVEMENT é utilizada tanto para movimentos artísticos (como em “What painting movement was conceived by Pablo Picasso and Georges Braque?”) como para correntes filosóficas (como em “What philosophy is associated with Jean Paul Sartre?”). Como já referido, esta decisão foi baseada no facto de estarmos perante conceitos semelhantes. Caso seja necessário (ou considerado mais correcto), estes conceitos poderiam passar a ser identificados por classes distintas.

- A classe HUMAN:DESCRIPTION:OCCUPATION é utilizada para ocupações de indivíduos. Caso se considerasse necessário, poderiam ser definidas duas subclasses desta classe, uma para identificação de profissões (como em “*What is James Herriot’s other occupation?*”) e outra para passatempos (Exemplo inexistente no *corpus* utilizado).

Para terminar, existem algumas classes cuja posição na hierarquia não pode ser definida de uma forma consensual:

- A classe SURNAME foi definida como uma subclasse de HUMAN:GROUP, devido ao facto de um apelido ser utilizado com frequência para identificar uma família. No entanto, um apelido também pode ser utilizado para identificar um indivíduo.
 - Exemplo: “*What was Rob Roy’s surname?*” foi classificada como HUMAN:INDIVIDUAL apesar de referir explicitamente que a resposta deverá ser um apelido.
 - No âmbito deste trabalho, apenas se utiliza a classe HUMAN:GROUP:SURNAME quando um apelido é utilizado como referência a uma família.
- De uma forma análoga, a classe SETTING foi considerada uma subclasse de DESCRIPTION:DESCRIPTION. Mas em alguns casos SETTING pode referir a algo mais simples como a localização (LOCATION) ou a época histórica (NUMBER:DATE).
 - Exemplo: “*What country is the setting for Colleen McCullough’s The Thorn Birds?*” foi classificada como LOCATION:SETTING, apesar de se referir a um contexto (DESCRIPTION:DEFINITION:SETTING).
 - No âmbito deste trabalho considera-se que uma pergunta é do tipo DESCRIPTION:DEFINITION:SETTING apenas se não existir uma referência específica a uma localização ou época.
- A classe CHARACTER foi considerada uma subclasse de HUMAN:INDIVIDUAL. No entanto, um CHARACTER não é necessariamente um humano, podendo em algumas situações ser um membro de ANIMAL.
 - Exemplo: “*What is the dogs name in Peter Pan?*” foi classificada como ENTITY:ANIMAL:REAL, apesar de também ser uma personagem (HUMAN:INDIVIDUAL:CHARACTER).
 - No âmbito deste trabalho, consideram-se membros de HUMAN:INDIVIDUAL:CHARACTER apenas personagens humanos.
- A classe PHILOSOPHER foi definida como uma subclasse de HUMAN:INDIVIDUAL. Esta também poderia ter sido definida como subclasse de HUMAN:INDIVIDUAL:ARTIST para manter a coerência (à semelhança do que aconteceu com a classe ENTITY:OTHER:MOVEMENT).

- Finalmente, as classes MUSEUM e SCHOOL foram consideradas localizações (subclasses de LOCATION) devido à inexistência de uma classe que identificasse o conceito “Instituição”. No caso de ser considerada a introdução de uma nova classe para este conceito, poderá também ser considerada a alteração destas duas classes para que passem a ser subclasses dessa nova classe.

3.6 Criação de corpus

Como já foi referido em 3.3, esta taxonomia foi definida com base em dois *corpora*: um em português sobre pintura e um em inglês sobre arte.

O primeiro *corpus*, tendo sido utilizado na versão anterior do Art.Ask, estava etiquetado com a taxonomia utilizada nessa mesma versão, já apresentada em 3.2. Como ficou estabelecido que a cada classe dessa taxonomia iria corresponder exactamente uma classe na nova taxonomia, a alteração da taxonomia nas perguntas deste *corpus* foi trivial.

Quanto ao segundo, como ainda não tinha sido utilizada no contexto de um sistema de pergunta e resposta, havia a necessidade de o classificar. Numa fase inicial, ele foi classificado com a taxonomia de Li e Roth de uma forma automática utilizando *Support Vector Machines*, sendo o conjunto de treino utilizado as 5500 perguntas recolhidas no âmbito da experiência originalmente realizada por Li e Roth (Li & Roth, 2006). Após esta fase inicial, houve uma segunda fase, de classificação manual, na qual foram alteradas as classificações consideradas incorrectas e adicionadas as classes abaixo do segundo nível (apenas existentes na nova taxonomia).

4

Interpretação de perguntas no Art.Ask

4.1 *Classificadores considerados*

4.1.1 **Support Vector Machines**

Um dos classificadores considerados neste trabalho foi o *Support Vector Machines*. A implementação utilizada foi a mesma em utilização no Just.Ask, sendo que como *feature* apenas foram utilizadas unigramas (devido ao facto ser necessária a utilização da WordNet¹ para *features* mais complexas como a *headword*, e não estar disponível uma versão da WordNet em português).

4.1.2 **Voting Model**

O outro dos classificadores a ser utilizados no âmbito deste trabalho é o *Voting Model*, já descrito em 2.2. Este classificador foi implementado no âmbito deste trabalho de acordo com o descrito em (Bhagat et al., 2005), incluindo os três modelos utilizados (unigramas, bigramas e trigramas). O único ponto divergente entre esta implementação e a descrição feita pelo autor está nas heurísticas utilizadas: no contexto deste trabalho apenas é considerado o nível 1 (dois elementos com o mesmo atributo atributo não ocorrem na mesma *frame*).

4.2 *Extensões aos classificadores*

Nesta secção são apresentadas duas extensões que podem ser utilizadas em conjunto com os classificadores descritos anteriormente com o objectivo de melhorar os resultados obtidos por estes nas situações em que apresentem falhas de desempenho.

4.2.1 **Restrições no valor de atributos**

A primeira das extensões propostas baseia-se na utilização de restrições que influenciam o valor de atributos, eliminando as *frames* que contenham combinações inválidas de atributos. No contexto deste

¹Uma base de dados léxica, disponível em <http://wordnet.princeton.edu>

trabalho, os *targets* serão filtrados de acordo com o tipo de pergunta.

Por exemplo, se considerarmos a pergunta “*What is the capital of \$COUNTRY?*”, que deverá ser classificada com o tipo `LOCATION:CITY`, terá obrigatoriamente um país como *target*.

De um modo mais geral, podem também existir restrições sobre a existência de atributos, i.e., é possível ter perguntas com vários *targets*.

Por exemplo, a pergunta “*Who directed the \$YEAR version of \$MOVIE*” tem dois *targets* (um ano e um título de filme). Por outro lado, é possível ter perguntas que não têm *targets*, como “*What time is it?*”.

Uma alteração significativa ao utilizar esta extensão está no facto de deixarmos de obter a pontuação para cada elemento, passando apenas a obter a pontuação agregada de cada *frame*, e apenas para as *frames* válidas. Isto é conseguido devido a ser possível obter previamente quais as combinações de tipo de pergunta e *targets* válidos. No entanto, de forma a obter a pontuação agregada de uma *frame*, é necessário definir uma função de agregação para calcular a pontuação de uma *frame* a partir das pontuações dos elementos que a compõem. Neste trabalho foram consideradas e comparadas duas funções de agregação:

- A média das pontuações de cada elemento;
- A pontuação do tipo de pergunta.

Enquanto que a utilização da média pode parecer a melhor solução, ela levanta um problema: o peso dado à pontuação do tipo de pergunta pode ser demasiado baixo para um atributo que pode influenciar o valor de outros. Por isso, também foi considerado apenas a pontuação do tipo de pergunta como medida. Uma outra hipótese seria atribuir um peso fixo ao tipo de pergunta (e.g., 0,5 para o tipo de pergunta e 0,5 para a média dos *targets*), mas devido à existência de tipos de perguntas sem *targets* associados esta opção não foi considerada.

4.2.2 Restrições baseadas em keywords

A segunda extensão proposta baseia-se na existência de *keywords* nas perguntas que permitem que o conjunto de valores possíveis para um atributo. No contexto deste trabalho, o atributo cujos valores possíveis serão afectados será o tipo de pergunta. Por exemplo, para uma pergunta que contenha a palavra “onde” apenas os tipos de pergunta que se refiram a locais (subclasses de `LOCATION`) serão válidos. Note-se que uma *keyword* pode ter vários tipos de pergunta associados.

As restrições utilizadas por esta extensão são diferentes das utilizadas pela extensão anterior porque neste caso estamos a restringir o valor de um atributo (o tipo de pergunta) a partir de uma palavra existente na pergunta, em vez de restringir o valor de um atributo com base no valor de outro.

Por exemplo, se tivermos a pergunta “*When was \$PERSON born?*” o tipo de pergunta deverá ser LOCATION devido à presença da palavra “*where*”. Por outro lado, se a pergunta fosse “*Where was \$PERSON born?*” o tipo de pergunta seria TIME, devido à palavra “*when*”.

Ao contrário da extensão anterior, esta extensão exige algum conhecimento adicional sobre o *corpus*, i.e., é necessário conhecer as perguntas de cada tipo e quais as palavras que podem ajudar a identificar esse tipo. Devido a este factor, quando esta extensão for testada, será corrido um teste inicial sem esta extensão.

Como esta extensão pode eliminar à partida alguns tipos de perguntas inválidos, é conveniente que ela seja aplicada antes da extensão anterior, evitando com isto que sejam calculadas pontuações para *frames* que serão mais tarde eliminadas. Por exemplo, uma pergunta contendo a palavra “onde” apenas pode ser dos tipos QT_BORN_WHERE ou QT_DIED_WHERE. Se esta extensão for aplicada antes da anterior apenas serão calculadas as pontuações para duas *frames* (uma para cada tipo de pergunta válido). Por outro lado, se esta extensão for aplicada depois da anterior serão calculadas as *frames* correspondentes a todos os tipos de pergunta, resultando numa perda de eficiência.

4.2.3 Utilização

No ficheiro QT-target.txt são descritos os *targets* associados a cada tipo de pergunta com o seguinte formato:

```
Tipo de pergunta
:Targets (um por linha, com um ``:`` no início)
```

Exemplo:

```
QT.WHO.ACT
:movie
QT.WHO.ACTS.IN.TWO
:movie
:movie
```

...

Neste exemplo, o tipo de pergunta `QT_WHO_ACT` tem um *target* do tipo *movie* e ao tipo de pergunta `QT_WHO_ACTS_IN_TWO` correspondem dois *targets* do tipo *movie*.

No ficheiro `regras.txt` são descritas as regras para filtragem de tipos de pergunta, com a seguinte sintaxe:

```
Keywords Tipos_de_pergunta_válidos ...
```

As *keywords* são descritas na forma de expressões regulares, utilizando a mesma sintaxe que a linguagem Java.

Exemplo:

```
.*contracen.* QT_WHICH_FILMS_MAIN_ACT_TWO QT_WHO_ACTS_WITH_IN  
.*interpret.* QT_WHICH_CHARACTER QT_WHO_ACTS_LIKE_IN  
...
```

Neste caso, as perguntas onde existam palavras que contenham a sequência “contracen” apenas podem corresponder os tipos `QT_WHICH_FILMS_MAIN_ACT_TWO` e `QT_WHO_ACTS_WITH_IN` e as perguntas onde existam palavras que contenham a sequência “interpret” apenas podem corresponder os tipos `QT_WHICH_CHARACTER` e `QT_WHO_ACTS_LIKE_IN`.

5 Avaliação

5.1 *Medidas utilizadas*

Esta avaliação será semelhante à da versão anterior, i.e., a abordagem inicial à interpretação das perguntas será comparada a novas abordagens (envolvendo classificadores) com recurso às medidas de precisão, cobertura e medida-F.

5.1.1 **Precisão**

No contexto dos sistemas de pergunta e resposta, a precisão representa a relação entre o número de perguntas respondidas correctamente e o número total de perguntas. Pode ser obtida através da seguinte expressão:

$$P = \frac{C}{M} \quad (5.1)$$

Em que C representa o número de perguntas respondidas correctamente e M representa o número total de perguntas.

5.1.2 **Cobertura**

Por outro lado, a cobertura mede a fracção de perguntas respondidas em relação ao número de perguntas que a informação disponível. Calcula-se através da seguinte equação:

$$R = \frac{C}{N} \quad (5.2)$$

Em que N representa o número de perguntas que seria possível responder com a base de dados actual.

5.1.3 Medida-F

Ambas as medidas anteriores apresentam desvantagens: a precisão permite obter valores elevados respondendo correctamente a um conjunto limitado de perguntas e ignorando as restantes, enquanto a cobertura toma valores altos se todas as perguntas forem respondidas de forma errada. Estas situações levam a que as tentativas de aumentar uma delas faça com que a outra diminua. O ideal é usar uma terceira medida, que seja função destas duas, como a medida-F.

A medida-F é dada pela seguinte expressão:

$$F_{\beta} = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \quad (5.3)$$

Sendo β uma constante com o objectivo de alterar o peso de cada uma das duas medidas. Como o valor mais habitual para β é 1, a expressão da medida-F pode ser simplificada:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (5.4)$$

Estas três medidas estão aqui apresentadas no contexto dos sistemas de pergunta e resposta, mais informação sobre elas está disponível (e num âmbito mais geral) em (Makhoul et al., 1999) e (Sokolova & Lapalme, 2009).

5.2 Corpora utilizados

Foram utilizados três *corpora* para comparar o desempenho entre os diferentes métodos para interpretação da pergunta.

Os dois primeiros *corpora* já foram apresentados anteriormente na secção 3.3. O terceiro *corpus* é um *corpus* composto por 260 perguntas em português sobre cinema (Dourado, 2009). Uma diferença relevante entre este *corpus* e os outros dois é que as perguntas que o formam não estão classificadas de acordo com a taxonomia desenvolvida no âmbito deste trabalho (apesar de poder ser desenvolvida uma correspondência entre classes desses domínio e a taxonomia já apresentada).

5.3 Avaliação da versão inicial

A interpretação da pergunta na versão inicial do Art.Ask, ao contrário de outros aspectos do sistema, não foi avaliada tendo em conta as medidas apresentadas (Fonseca, 2009). Em vez disso, a avaliação

considerou que era sempre atribuído às perguntas o tipo correspondente à pergunta mais próxima.

Assim, a avaliação foi feita considerando apenas a fracção de perguntas à qual foi atribuído o tipo correcto (precisão).

Para avaliar o sistema, foi utilizado um conjunto de teste com 50 perguntas, tendo destas 42 sido identificadas com o tipo correcto, resultando numa precisão de 84%.

5.4 Método utilizado

Para a avaliação da interpretação da pergunta utilizando o *Voting Model*, cada *corpus* foi dividido em dois conjuntos. A cada pergunta foi atribuído de forma aleatória um dos conjuntos, sendo de 85% a probabilidade de uma pergunta ser atribuída ao conjunto de treino e de 15% a probabilidade de ser atribuída ao conjunto de teste.

As tabelas A.1, A.2 e A.3, presentes no anexo A ilustram o número de perguntas em cada conjunto definido a partir de cada *corpus*, assim como a forma como estão divididas entre os vários tipos. Em relação ao *corpus* sobre cinema, são também apresentados quais os *targets* correspondentes a cada pergunta.

Em relação aos dois primeiros *corpora* considerados, e de forma a eliminar a classificação de perguntas a que o *Voting Model* tenha atribuído pontuações mais baixas (e que seria mais provável estarem classificadas incorrectamente), foi decidido à partida apenas atribuir uma classificação às perguntas que tivessem uma pontuação acima de um limiar. Para além disto, como na avaliação utilizando estes *corpora* não foram definidos *targets*, as extensões propostas não serão utilizadas.

Após o cálculo das medidas apresentadas na secção 5.5.2.1, decidiu-se que a forma mais adequada de definir seria através da maximização da medida-F, devido à existência de alguns problemas na utilização das outras medidas:

- A precisão atinge o valor máximo quando o limiar considerado é ∞ , sendo este o caso em que nenhuma pergunta é classificada;
- A cobertura atinge o valor máximo quando o limiar considerado é 0, sendo este o caso em que todas as perguntas são classificadas (independentemente desta classificação ser correcta);

No terceiro *corpus*, a avaliação é feita de forma ligeiramente diferente: como para este *corpus* foram consideradas as extensões propostas anteriormente, todas as classificações atribuídas pelo *Voting Model*

foram consideradas válidas.

5.5 Resultados

5.5.1 Corpus sobre arte

5.5.1.1 Determinação do limiar e da medida-F

Para determinar o limiar mais apropriado, foram calculadas as medidas de precisão, cobertura e medida-F considerando limiares entre 0,1 e 1 com intervalos de 0,1. Os resultados estão na tabela 5.1.

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Precisão	0,447	0,447	0,432	0,444	0,387	0,476	0,529	0,500	0,429	0,300
Cobertura	1,000	1,000	0,941	0,889	0,632	0,370	0,300	0,267	0,200	0,097
Medida-F	0,618	0,618	0,593	0,593	0,480	0,417	0,383	0,348	0,273	0,146

Tabela 5.1: Avaliação do *Voting Model* entre 0,1 e 1,0

O comportamento da medida-F neste caso é semelhante ao encontrado no caso anterior. Procedemos então da mesma forma para encontrar o limiar mais adequado.

	0,232	0,233	0,234
Precisão	0,447	0,447	0,432
Cobertura	1,000	1,000	0,941
Medida-F	0,618	0,618	0,593

Tabela 5.2: Avaliação do *Voting Model* na proximidade de 0,233

Na tabela 5.2, vemos que o limiar adequado é neste caso 0,233 e a medida-F correspondente a este valor é 0,618.

5.5.1.2 Análise de erros

A tabela 5.3 mostra os resultados da avaliação para o *corpus*, explicitando os resultados obtidos para cada tipo de pergunta.

5.5.1.3 Comparação com Support Vector Machines

Após termos apresentado os resultados obtidos pela avaliação deste *corpus* com o *Voting Model*.

	Total	Certas	Erradas	Sem resposta
ENTITY:ANIMAL	1	0	0	1
ENTITY:COLOR	1	1	0	0
ENTITY:CREATIVE:BOOK:NOVEL	1	1	0	0
ENTITY:CREATIVE:COMIC	1	0	1	0
ENTITY:CREATIVE:PLAY	1	0	1	0
ENTITY:OTHER	4	0	4	0
ENTITY:SUBSTANCE	1	1	0	0
ENTITY:TECHNIQUE:PAINTING	1	0	1	0
HUMAN:GROUP:SURNAME	1	0	1	0
HUMAN:INDIVIDUAL	2	0	2	0
HUMAN:INDIVIDUAL:ARTIST:SCULPTOR	1	0	1	0
HUMAN:INDIVIDUAL:CHARACTER	7	7	0	0
HUMAN:INDIVIDUAL:PHILOSOPHER	1	0	1	0
LOCATION:CITY	1	1	0	0
LOCATION:COUNTRY	2	0	2	0
LOCATION:COUNTRY:NATIONALITY	2	1	1	0
LOCATION:OTHER	3	2	1	0
LOCATION:OTHER:PLANET	1	0	1	0
LOCATION:OTHER:SCHOOL	1	1	0	0
NUMERIC:COUNT	4	3	1	0

Tabela 5.3: Resultados da avaliação por tipo de pergunta para o *corpus* sobre arte

	<i>Support Vector Machines</i>	<i>Voting Model</i>
Precisão	0,711	0,447
Cobertura	1,000	1,000
Medida-F	0,831	0,618

Tabela 5.4: Comparação entre os dois métodos utilizados para a avaliação do *corpus* em inglês

Na tabela 5.4 são apresentados os resultados utilizando os dois métodos utilizados. Como podemos ver, utilizando *Support Vector Machines* os resultados são melhores em duas das três medidas utilizadas (precisão e medida-F). Na terceira medida, a cobertura, os resultados são iguais. No entanto, como o programa utilizando para a avaliação com *Support Vector Machines* atribui sempre uma classificação a todas as perguntas, a cobertura obtida por este método será sempre 1,000, pelo que esta medida não é indicada para ser utilizado em comparações.

5.5.2 Corpus sobre cinema

5.5.2.1 Avaliação inicial

Como para utilizar a segunda extensão é necessária uma lista de *keywords* e respectivos tipos de pergunta correspondentes, começamos por apresentar a avaliação considerando apenas a primeiras das

extensões apresentadas.

Como já foi dito anteriormente, a avaliação com o *corpus* sobre cinema processou-se de forma diferente: como a todas as perguntas foi atribuída uma classificação, a avaliação apenas teve em conta os valores obtidos para a precisão.

A tabela 5.5.2.1 mostra os valores de precisão obtidos para o *Voting Model* considerando ambas as medidas de agregação descritas em 5.5.2.1.

	VM (Média)	VM (Tipo de pergunta)	SVM
Precisão	0,610	0,659	0,951

Tabela 5.5: Comparação entre as duas medidas definidas para a primeira extensão quando aplicada ao *Voting Model* e com *Support Vector Machines* sem extensão

Como podemos ver, os resultados obtidos com o *Voting Model* são inferiores aos obtidos quando é utilizado *Support vector machines*, independentemente da medida considerada para agregação. Por outro lado, podemos também ver que ao utilizar a pontuação do tipo de pergunta como pontuação para a *frame* os resultados obtidos são melhores que ao usar a média de todos os pesos.

5.5.2.2 Determinação das keywords

Para utilizar a segunda extensão é necessário definir um conjunto de *keywords*. Para isto, foi criado um novo conjunto de teste a partir do conjunto de treino existente, novamente com 15% das perguntas, mantendo-se os restantes 85% como conjunto de treino.

Para cada pergunta do novo conjunto de teste que fosse classificada de forma incorrecta, foram adicionadas regras correspondentes a palavras que pudessem identificar o tipo de pergunta em causa, independentemente de ser suficiente para que a pergunta fosse classificada correctamente.

A tabela 5.6 mostra quais as regras adicionadas, assim como os tipos de pergunta que lhes correspondem.

É importante reparar que estas regras são dependentes do domínio em questão e devem ser redefinidas quando esta extensão for utilizada num domínio diferente.

Após aplicação da segunda extensão, os resultados passam a ser os descritos na tabela 5.5.2.2

Apesar de a segunda extensão ter contribuído para que os resultados melhorassem, eles continuam abaixo daqueles obtidos quando é utilizado *Support Vector Machines*.

5.5.2.3 Análise de erros

A tabela 5.8 mostra os resultados da avaliação para este *corpus*, explicitando os resultados obtidos para cada tipo de pergunta.

Podemos verificar que existem alguns tipos de pergunta (QT_BORN_WHEN, QT_DIED_WHERE, QT_WHO_ACT, QT_WHO_ACTS_WITH_IN, QT_WHO_DIRECTED e QT_WHO_WON) que são classificados correctamente em todos os testes. Destes tipos, todos têm pelo menos uma *keyword* associada, o que indica que de certa forma a sua identificação é simples.

Por outro lado, existem três tipos de pergunta (QT_WHO_ACTS_LIKE_IN, QT_WHO_IS e QT_WHO_WON) que em nenhum dos casos são classificados correctamente. Todos estes tipos de pergunta tem *keywords* associadas, QT_WHO_ACTS_LIKE_IN tem duas *keywords* e os restantes dois tipos têm uma *keyword* cada um (a mesma *keyword* está associada ao dois tipos, e também é uma das *keywords* associada ao tipo QT_WHO_ACTS_LIKE_IN).

Em relação ao *Support Vector Machines*, existem apenas duas perguntas que não são classificadas correctamente: uma do tipo QT_BORN_WHERE e uma do tipo QT_DIED_WHERE. Uma destas perguntas poderia ser classificada correctamente caso as extensões fossem também aplicadas ao *Support Vector Machines*, aumentando a precisão obtida por esse método para 0,976.

5.5.3 Corpus sobre pintura

5.5.3.1 Determinação do limiar e da medida-F

De uma forma análoga ao que foi feito para o *corpus* sobre arte, a tabela 5.9 mostra os resultados da avaliação para o *corpus* em inglês.

Como se pode ver, em todo o intervalo considerado, a precisão é crescente e a cobertura decrescente. Para os valores entre 0,1 e 0,4 a medida-F tem o mesmo valor (0,864). Para os valores considerados entre 0,4 e 1,0, a medida-F é decrescente.

Como existe um intervalo de valores no qual a medida-F tem um valor máximo (com início em 0 e fim entre 0,4 e 0,5), decidiu-se utilizar como critério de desempate o valor com maior precisão.

Ao testar os valores entre 0,400 e 0,499, foi encontrado 0,422 como o limiar que satisfaz as condições apresentadas.

Temos então que o limiar mais adequado a este *corpus* é 0,422 e a medida-F que lhe corresponde é 0,864.

Como já referido anteriormente, a versão inicial do Art.Ask reportava uma precisão de 84%, o que é um valor superior aos 77% obtidos quando é utilizado o *Voting Model*. No entanto, é possível obter um resultado superior a este ao utilizar *Support Vector Machines*, método este que obtém uma precisão de 96%. Note-se, no entanto, que como o *corpus* utilizado para a avaliação da versão original não foi o mesmo utilizado no âmbito deste trabalho (o primeiro era composto por um número mais reduzido de perguntas), pelo que esta comparação é aproximada.

5.5.3.2 Análise de erros

A tabela 5.11 mostra os resultados da avaliação para o *corpus*, explicitando os resultados obtidos para cada tipo de pergunta.

Nesta tabela podemos verificar que os tipos de pergunta que aparecem mais vezes no conjunto de treino são os que obtêm melhores resultados (e.g., para o tipo ENTITY:OTHER:MOVEMENT:PAINTING:PAINTERMOVEMENT, o mais frequente, todas as perguntas são classificadas de forma correcta). Por outro lado, os tipos com menor representação (como LOCATION:COUNTRY:COUNTRYOFDEATH) têm maior número de perguntas com classificação errada (ou sem classificação).

Keyword	Tipos de pergunta válidos
Palavras começadas com “contracen” (e.g., “contracena”)	QT_WHICH_FILMS_MAIN_ACT_TWO ou QT_WHO_ACTS_WITH_IN
Palavras começadas com “interpret” (e.g., “interpreta”)	QT_WHICH_CHARACTER ou QT_WHO_ACTS_LIKE_IN
Palavras começadas com “morr” ou “mort” (e.g., “morte” ou “morreu”)	QT_DIED_WHEN ou QT_DIED_WHERE
Palavras começadas com “nasc” (e.g., “nascimento” ou “nasceu”)	QT_BORN_WHEN ou QT_BORN_WHERE
Palavras começadas com “nomea” (e.g., “nomeação” ou “nomeado”)	QT_HOW_MANY_NOMINATIONS, QT_HOW_MANY_NOMINATIONS_FILM or QT_WHO_NOMINATED
Palavras começadas com “particip” (e.g., “participou” or “participação”)	QT_LAST_FILM, QT_WHICH_FILMS, QT_WHO_ACTS, QT_WHO_ACTS_IN_TWO ou QT_WHO_MAIN_ACT
A palavra “principal”	QT_MAIN_CHARACTER, QT_WHICH_FILMS_MAIN_ACT ou QT_WHO_MAIN_ACT
Palavras começadas com “protagoni” (e.g., “protagonista” ou “protagonizou”)	QT_WHICH_FILMS_MAIN_ACT, QT_WHO_ACTS_IN_TWO ou QT_WHO_MAIN_ACT
Palavras começadas com “roda” (e.g., “rodado” ou “rodagem”)	QT_WHICH_YEAR_FILM
A palavra “quando”	QT_BORN_WHEN, QT_DIED_WHEN ou QT_WHICH_YEAR_FILM
A palavra “onde”	QT_BORN_WHERE ou QT_DIED_WHERE
A palavra “quanto” e variações	QT_BORN_WHEN, QT_HOW_MANY_NOMINATIONS_FILM, QT_HOW_MANY_OSCARS ou QT_HOW_MANY_OSCARS_FILM
A palavra “quem”	QT_MAIN_CHARACTER, QT_WHO_ACT, QT_WHO_ACTS_IN_TWO, QT_WHO_ACTS_LIKE_IN, QT_WHO_ACTS_WITH_IN, QT_WHO_DIRECTED, QT_WHO_NOMINATED, QT_WHO_IS, QT_WHO_MAIN_ACT, QT_WHO_PRODUCED, QT_WHO_WON ou QT_WHO_WROTE

Tabela 5.6: Regras utilizadas na segunda extensão proposta

	VM (Média)	VM (Tipo de pergunta)	VM e regras (Média)	VM e regras (Tipo de pergunta)
Precision	0,610	0,659	0,732	0,707

Tabela 5.7: Comparação o *Voting Model* com a primeira extensão e ambas as extensões

	VM (Média)	VM (Tipo de pergunta)	VM e regras (Média)	VM e regras (Tipo de pergunta)	SVM	Total
QT_BORN_WHEN	3	3	3	3	3	3
QT_BORN_WHERE	1	2	3	4	4	5
QT_DIED_WHEN	0	1	1	1	1	1
QT_DIED_WHERE	1	1	1	1	0	1
QT_HOW_MANY_NOMINATIONS	1	1	2	1	2	2
QT_HOW_MANY_OSCARS	1	1	1	1	2	2
QT_HOW_MANY_OSCARS_FILM	1	0	1	0	1	1
QT_WHICH_FILMS	1	2	1	2	2	2
QT_WHICH_YEAR_FILM	2	2	3	2	3	3
QT_WHO_ACT	6	6	6	6	6	6
QT_WHO_ACTS_LIKE_IN	0	0	0	0	1	1
QT_WHO_ACTS_WITH_IN	2	2	2	2	2	2
QT_WHO_DIRECTED	3	3	3	3	3	3
QT_WHO_IS	0	0	0	0	2	2
QT_WHO_MAIN_ACT	3	3	3	3	3	3
QT_WHO_WON	0	0	0	0	4	4

Tabela 5.8: Resultados da avaliação por tipo de pergunta para o *corpus* sobre cinema

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Precisão	0,760	0,760	0,760	0,770	0,778	0,800	0,836	0,828	0,875	0,920
Cobertura	1,000	1,000	1,000	0,983	0,949	0,918	0,875	0,828	0,721	0,648
Medida-F	0,864	0,864	0,864	0,864	0,855	0,855	0,855	0,828	0,790	0,760

Tabela 5.9: Avaliação do *Voting Model* entre 0,1 e 1,0

	0,421	0,422	0,423
Precisão	0,770	0,770	0,767
Cobertura	0,983	0,983	0,966
Medida-F	0,864	0,864	0,855

Tabela 5.10: Avaliação do *Voting Model* na proximidade de 0,422

	Total	Certas	Erradas	Sem resposta
HUMAN:INDIVIDUAL:ARTIST:PAINTER	3	2	1	0
ENTITY:OTHER:MOVEMENT:PAINTING:PAINTERMOVEMENT	22	22	0	0
NUMERIC:DATE:DATEOFPAINTING	12	12	0	0
ENTITY:CREATIVE:PAINTING	5	3	2	0
LOCATION:COUNTRY:NATIONALITY	9	7	2	0
NUMERIC:DATE:DATEOFBIRTH	1	0	0	1
LOCATION:CITY:CITYOFDEATH	7	4	1	2
LOCATION:OTHER:MUSEUM	6	5	0	1
NUMERIC:SIZE:DIMENSIONS	3	3	0	0
LOCATION:CITY:CITYOFBIRTH	4	0	4	0
LOCATION:COUNTRY:COUNTRYOFDEATH	1	0	0	1
LOCATION:COUNTRY:COUNTRYOFBIRTH	1	0	1	0
ENTITY:OTHER:MOVEMENT:PAINTING:PAINTINGMOVEMENT	1	0	1	0

Tabela 5.11: Resultados da avaliação por tipo de pergunta para o *corpus* sobre pintura

Conclusões e trabalho futuro

6.1 Contribuições

Este trabalho teve como ponto de partida o Art.Ask, um sistema de pergunta e resposta desenvolvido anteriormente no L2F. O seu objectivo foi estender o sistema existente, focando-se em dois aspectos:

- Definição de uma nova taxonomia a ser utilizada na classificação de perguntas: a versão inicial do Art.Ask classificava as perguntas com base numa taxonomia composta por dezasseis tipos de pergunta, todos no domínio da pintura. A nova taxonomia está definida com base na taxonomia de Li e Roth, tendo sido construída através da adição de algumas classes pertencentes ao domínio do Art.Ask;
- Comparação de vários métodos para a interpretação de perguntas: na versão inicial do Art.Ask as perguntas eram classificadas de uma forma algo simples, sendo obtida através da distância de *Levenshtein* entre a pergunta do utilizador e um conjunto de perguntas tipo, previamente classificadas, sendo o tipo de pergunta definido como o que correspondesse à menor distância. Neste trabalho foi testada a interpretação de perguntas com recurso a classificadores (*Support Vector Machines* e *Voting Model*).

6.2 Conclusões

Neste trabalho pudemos verificar que a interpretação de perguntas é feita com recurso a *Support Vector Machines* os resultados são melhores que os obtidos quando se utiliza *Voting Model*. Ao utilizar as extensões propostas para melhorar os resultados obtidos pelo *Voting Model*, apesar de haver melhorias nos resultados, eles continuam a ser inferiores aos obtidos com a utilização de *Support Vector Machines*.

Outro aspecto relevante é que os resultados obtidos utilizando o *Voting Model* são inferiores ao resultado da avaliação utilizando a distância de *Levenshtein*. No entanto, como em nenhum dos casos a avaliação foi feita com recurso ao mesmo *corpus* utilizado com a versão inicial, é possível que o resultado obtido inicialmente esteja inflacionado devido ao tamanho reduzido do *corpus* (apenas 50 perguntas).

Em relação à primeira das extensões propostas, ela faz com que combinações inválidas de tipos de pergunta e *targets* sejam consideradas inválidas, podendo assim eliminar classificações inválidas.

Quanto à segunda das extensões, ela pode fazer com que algumas perguntas que seriam classificadas incorrectamente passem a ser classificadas de forma correcta (assumindo que a lista de tipos de pergunta associados a cada *keyword* está correcta), sem no entanto alterar as classificações das perguntas que já foram classificadas correctamente.

6.3 Trabalho futuro

Ainda existem alguns aspectos desta aplicação que podem ser melhorados:

6.3.1 Extracção de informação

O componente de extracção de informação do Art.Ask manteve-se inalterado no âmbito deste trabalho. No entanto, existem alguns aspectos nos quais este pode ser melhorado:

- A versão actual do Art.Ask não está preparado para responder a perguntas em que existem erros ortográficos. Por exemplo, a pergunta “Quando nasceu Eduard Munch?” não pode ser respondida mesmo que seja interpretada correctamente, devido à incapacidade de reconhecer Eduard Munch como o nome do pintor Edvard Munch. A introdução de um corrector ortográfico poderia resolver este problema;
- Para além deste problema, existe o problema das páginas de desambiguação da Wikipedia: expressões como “A Primavera” podem referir-se a vários conceitos (a estação do ano ou o quadro de Botticelli). O Art.Ask pode ser melhorado de forma a identificar o *link* correcto a seguir quando encontra uma página de desambiguação.

6.3.2 Interactividade

Outro aspecto que pode vir a ser implementado numa versão futura do Art.Ask é a capacidade do sistema interagir com o utilizador, com o objectivo de eliminar casos de ambiguidade que possam surgir. Isto pode ser útil quando o sistema não é capaz de interpretar a pergunta do utilizador, seja devido a ser difícil encontrar o tipo de pergunta correcto ou identificar as entidades mencionadas apresentado nesse caso um pedido de clarificação.

Um exemplo de um diálogo de clarificação seria:

Utilizador: Quem realizou o filme Ocean's Eleven?

Art.Ask: Refere-se ao filme de 1960 ou ao de 2001?

Utilizador: 2001.

Art.Ask: Steven Soderbergh.

6.3.3 Classificação por níveis

Apesar de ter sido definida uma nova taxonomia para a classificação das perguntas, a classificação continua a ser feita sem ter em conta os vários níveis da hierarquia, i.e, quando uma pergunta é classificada é-lhe atribuída directamente uma classe. De forma a utilizar devidamente uma taxonomia com vários níveis o sistema deve ser adaptado para que inicialmente seja atribuída uma classe do nível superior, em seguida uma classe do segundo nível e assim sucessivamente. Por exemplo, a pergunta "Qual é a capital de \$PAIS?" deve ser classificada inicialmente com o tipo `LOCATION` e numa segunda fase com o tipo `LOCATION:CITY`.

Bibliografia

- Bhagat, R., Leuski, A., & Hovy, E. (2005, October). Statistical shallow semantic parsing despite little training data. 186–187.
- David, S. M., Stallard, D., Bobrow, R., & Schwartz, R. (1996). A fully statistical approach to natural language interfaces. 55–61.
- Dourado, M. (2009). *Em que filmes participou a atriz scarlett johansson?* Unpublished master's thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Fonseca, P. (2009). *Porque sorri a mona lisa?* Unpublished master's thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Goblirsch, D. M. (1996). Viterbi beam search with layered bigrams. 2131–2134.
- Guimarães, A. (2007). *Já te digo – uma interface em língua natural para uma base de dados de cinema.* Unpublished master's thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- Leuski, A., Pair, J., Traum, D., McNerney, P. J., Georgiou, P., & Patel, R. (2006). How to talk to a hologram. 360–362.
- Leuski, A., Patel, R., & Traum, D. (2006). Building effective question answering characters. 18–27.
- Li, X., & Roth, D. (2006). Learning question classifiers: the role of semantic information. *Nat. Lang. Eng.*, 12(3), 229–249.
- Makhoul, J., Kubala, F., Schwartz, R., & Weischedel, R. (1999). Performance measures for information extraction. 249–252.
- Ponte, J. M., & Croft, W. B. (1998). A language modeling approach to information retrieval. 275–281.
- Pradhan, S., Ward, W., Hacioglu, K., Martin, J. H., & Jurafsky, D. (2004). Shallow semantic parsing using support vector machines. 233–240.
- Salor, O., Pellom, B. L., Ciloglu, T., & Demirekler, M. (2007). Turkish speech corpora and recognition tools developed by porting sonic: Towards multilingual speech recognition. *Comput. Speech Lang.*, 21(4), 580–593.

- Seneff, S. (1992). Tina: a natural language system for spoken language applications. *Comput. Linguist.*, 18(1), 61–86.
- Seneff, S., Glass, J., Goddeau, D., Goodine, D., Hirschman, L., Leung, H., et al. (1991). Development and preliminary evaluation of the mit atis system. 88–93.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.*, 45(4), 427–437.
- Ward, W. (1990). The cmu air travel information service: understanding spontaneous speech. 127–129.

I Apêndices

A
tabelas

	Treino	Teste	Total
HUMAN:INDIVIDUAL:ARTIST:PAINTER	35	3	38
ENTITY:TECHNIQUE:MOVEMENT:PAINTING:PAINTERMOVEMENT	183	22	205
NUMERIC:DATE:DATEOFPAINTING	39	12	51
ENTITY:CREATIVE:PAINTING	55	5	60
LOCATION:COUNTRY:NATIONALITY	45	9	54
NUMERIC:DATE:DATEOFBIRTH	22	1	23
LOCATION:CITY:CITYOFDEATH	29	7	36
LOCATION:OTHER:MUSEUM	36	6	42
NUMERIC:SIZE:DIMENSIONS	56	3	59
NUMERIC:DATE:DATEOFDEATH	3	0	3
LOCATION:CITY:CITYOFBIRTH	19	4	23
LOCATION:COUNTRY:COUNTRYOFDEATH	5	1	6
LOCATION:COUNTRY:COUNTRYOFBIRTH	15	1	16
ENTITY:OTHER:MOVEMENT:PAINTING:PAINTINGMOVEMENT	3	1	4
ENTITY:SUBSTANCE	3	0	3
HUMAN:INDIVIDUAL:ARTIST:PAINTER:FULLNAME	1	0	1

Tabela A.1: Constituição de cada um dos conjuntos utilizados na avaliação (*corpus* sobre pintura)

	Treino	Teste	Total
DESCRIPTION:DESCRIPTION:SETTING	2	0	2
DESCRIPTION:DESCRIPTION:SUBJECT	3	0	3
DESCRIPTION:DESCRIPTION:TRAIT	1	0	1
DESCRIPTION:DESCRIPTION:TRAIT:GENDER	1	0	1
DESCRIPTION:REASON:CAUSE	1	0	1
ENTITY:ANIMAL	1	1	2
ENTITY:ANIMAL:MYTH	1	0	1
ENTITY:ANIMAL:NAME	5	0	5
ENTITY:ANIMAL:REAL	5	0	5
ENTITY:COLOR	3	1	4
ENTITY:CREATIVE:BOOK	3	0	3
ENTITY:CREATIVE:BOOK:NOVEL	9	1	10
ENTITY:CREATIVE:COMIC	2	1	3
ENTITY:CREATIVE:PLAY	5	1	6
ENTITY:CREATIVE:POEM	2	0	2
ENTITY:FOOD	1	0	1
ENTITY:FOOD:SEASONING	1	0	1
ENTITY:INSTRUMENT:MUSICAL	1	0	1
ENTITY:OTHER	9	4	13
ENTITY:OTHER:MOVEMENT	1	1	2
ENTITY:OTHER:MOVEMENT:PAINTING	1	0	1
ENTITY:OTHER:ROCK	1	0	1
ENTITY:OTHER:ROOM	1	0	1
ENTITY:PLANT	1	0	1
ENTITY:SPORT	1	0	1
ENTITY:SUBSTANCE	1	1	2
ENTITY:TECHNIQUE:PAINTING	1	1	2
ENTITY:WORD	3	0	3
HUMAN:DESCRIPTION:OCCUPATION	2	0	2
HUMAN:GROUP	1	0	1
HUMAN:GROUP:SURNAME	4	1	5
HUMAN:INDIVIDUAL	6	2	8
HUMAN:INDIVIDUAL:ARTIST	4	0	4
HUMAN:INDIVIDUAL:ARTIST:COMPOSER	1	0	1
HUMAN:INDIVIDUAL:ARTIST:PAINTER	8	0	8
HUMAN:INDIVIDUAL:ARTIST:SCULPTOR	1	0	1
HUMAN:INDIVIDUAL:ARTIST:WRITER	1	0	1
HUMAN:INDIVIDUAL:CHARACTER	49	7	56
HUMAN:INDIVIDUAL:PHILOSOPHER	0	1	1
LOCATION:CITY	7	1	8
LOCATION:COUNTRY	10	2	12
LOCATION:COUNTRY:NATIONALITY	9	2	11
LOCATION:OTHER	5	3	8
LOCATION:OTHER:ISLAND	2	0	2
LOCATION:OTHER:MUSEUM	1	0	1
LOCATION:OTHER:PLANET	1	1	2
LOCATION:OTHER:SCHOOL	3	1	4
LOCATION:OTHER:STATE	1	0	1
NUMERIC:CODE	1	0	1
NUMERIC:COUNT	9	4	13
NUMERIC:DATE:SEASON	1	0	1
NUMERIC:DATE:TIME	1	0	1

Tabela A.2: Constituição de cada um dos conjuntos utilizados na avaliação (*corpus* sobre arte)

	Targets	Treino	Teste	Total
QT_BORN_WHEN	actor	12	3	15
QT_BORN_WHERE	actor	12	5	77
QT_DIED_WHEN	actor	4	1	5
QT_DIED_WHERE	actor	7	1	8
QT_HOW_MANY_NOMINATIONS	actor	3	2	5
QT_HOW_MANY_NOMINATIONS_FILM	movie	3	0	3
QT_HOW_MANY_OSCARS	actor	4	2	6
QT_HOW_MANY_OSCARS_FILM	movie	3	1	4
QT_LAST_FILM	actor	3	0	3
QT_MAIN_CHARACTER	movie	1	0	1
QT_WHICH_CHARACTER	actor	1	0	1
QT_WHICH_FILMS	actor	5	2	7
QT_WHICH_FILMS_DIRECTED	actor	2	0	2
QT_WHICH_FILMS_MAIN_ACT	actor	3	0	3
QT_WHICH_FILMS_MAIN_ACT_TWO	actor,actor	2	0	2
QT_WHICH_WON		5	0	5
QT_WHICH_YEAR_FILM	movie	5	3	8
QT_WHO_ACT	movie	26	6	32
QT_WHO_ACTS_IN_TWO	movie	7	0	7
QT_WHO_ACTS_LIKE_IN	character,movie	13	1	14
QT_WHO_ACTS_WITH_IN	actor,movie	12	2	14
QT_WHO_DIRECTED	movie	32	3	35
QT_WHO_IS	actor	6	2	8
QT_WHO_MAIN_ACT	movie	29	3	32
QT_WHO_NOMINATED		7	0	7
QT_WHO_PRODUCED	movie	1	0	1
QT_WHO_WON		15	4	19
QT_WHO_WROTE	movie	3	0	3

Tabela A.3: Constituição de cada um dos conjuntos utilizados na avaliação (*corpus* sobre arte)