

Metodologia de Pesquisa em Computação 2015

Aula 2

Profa. Dra. Valéria D. Feltrim
PCC – DIN – UEM

Projeto de Dissertação

- Capa - **Título**
- Folha de rosto
- **Resumo**
- **Abstract**
- Listas
- Sumário
- **Introdução**
- **Desenvolvimento**
 - Revisão Bibliográfica
 - Proposta de Dissertação
- **Conclusão**
- Referências Bibliográficas
- (Apêndices)
- (Anexos)

Título

- O título é o primeiro meio para chamar a atenção de um potencial leitor
- Deve destacar a principal contribuição do trabalho de forma sintética
- Um bom título é **claro e conciso**
 - Evite títulos muito genéricos
 - Evite títulos muito longos
 - *Implementação de um Separador Silábico para o Português Brasileiro*
 - *Implementação de um Separador Silábico Baseado em Regras Linguísticas para o Português Brasileiro*

Resumo

- **Importante** → muitos leitores se limitam a ler o título e o resumo de um trabalho
- Deve ser completo, interessante e informativo
 - Dispensar a consulta ao restante do texto para se ter a ideia do que trata o trabalho
 - Estimular o interesse pela leitura do texto completo
 - Destacar a principal contribuição
 - É o lugar para se “vender o peixe”
- **Não é** como um *trailer* de filme
 - O resumo científico resume a história E CONTA O FINAL!

Resumo

- **Importante** → muitos leitores se limitam a ler o título e o resumo de um trabalho
- Deve ser completo, interessante e informativo
 - Dispensar a consulta ao restante do texto para se ter a ideia do que trata o trabalho
 - Estimular o interesse
 - Destacar a principal contribuição
 - É o lugar para se “vender o peixe”
- **Não é** como um *trailer* de filme
 - O resumo científico resume a história E CONTA O FINAL!

Descrever quais os objetivos, resultados e conclusões do trabalho, em vez de apenas “indicar” o que foi feito

Resumo

- Exemplo de resumo **indicativo**

Um estudo empírico visando avaliar a eficácia em revelar erros, a dificuldade de satisfação e o custo de aplicação do critério Análise de Mutantes é apresentado neste trabalho. Eficácia e custo também foram avaliados para os critérios Potenciais-Usos, comparando-se assim os resultados obtidos para esses critérios. A especificação e implementação de uma estratégia de minimização de conjuntos de casos de teste adequados ao critério Análise de Mutantes também é apresentada.

Quais os resultados?

Quais as conclusões?

 **evitar**

Resumo

- Exemplo de resumo **informativo**

Técnicas de re-teste seletivo tem como objetivo auxiliar a seleção de casos de teste de regressão com base nas modificações realizadas. Este trabalho apresenta uma nova técnica: Weighted Similarity Approach for Regression Testing (WSA-RT). Esta técnica utiliza uma abordagem de teste baseado em modelos e valores, para selecionar os casos de teste mais importantes dentre os que interagem com as modificações realizadas no software. Os estudos de caso realizados com a técnica indicam que WSA-RT é capaz de reduzir significativamente (de 70% a 80%) o tamanho do conjunto de casos de teste, podendo assim contribuir para a redução dos custos do teste de regressão sem afetar a capacidade de detecção de defeitos.

Quais os resultados?

Quais as conclusões?



preferir

Resumo

- A estrutura dos resumos de quase todas as áreas de estudo é muito similar
 - A informação incluída e a ordem em que aparecem são convencionais
 - Variações na ênfase dada a cada componente
- Vários autores apresentam modelos de resumo
 - Embora cada autor tenha a sua forma de expressar o modelo, existe um consenso sobre os elementos típicos e sua ordem

Resumo

- Estrutura típica de resumo
 - Muito parecida com a estrutura da introdução
 - Uma boa estratégia é escrever o resumo depois da introdução

Contexto

- Declarar proeminência do tópico
- Familiarizar termos e conceitos
- Introduzir a pesquisa a partir da grande área

Problema

- Citar problemas/dificuldades
- Citar necessidades/requisitos
- Citar a ausência ou pouca pesquisa anterior

Objetivo

- Indicar o propósito principal
- Detalhar/Especificar o propósito
- Introduzir mais propósitos

Metodologia

- Listar critérios ou condições
- Citar/Descrever materiais e métodos
- Justificar a escolha pelos materiais e métodos

Principais resultados

- Descrever o artefato
- Apresentar resultados
- Comentar/Discutir resultados

Conclusão

- Apresentar conclusões
- Apresentar contribuições/valor do trabalho
- Apresentar recomendação

Exercício

- As sentenças do resumo seguinte estão embaralhadas
- Vamos colocá-las na ordem correta de acordo com o modelo de resumo visto
 - Vamos identificar os componentes e ordená-los
- Título do resumo
 - ***Extração de Relações Semânticas de Textos em Português Explorando a DBpédia e a Wikipédia***

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
2. **Vários trabalhos anteriores** abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, **é um passo importante** para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

C

C

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
8. **No entanto**, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

C

C

P

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. **Este artigo discute** uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

C

O

C

P

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.

M

2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.

C

3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.

4. A relação é atribuída por votação ponderada, com base nestes exemplos.

5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).

O

6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.

C

7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.

8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.

P

9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.

M

2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.

C

3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.

4. A relação é atribuída por votação ponderada, com base nestes exemplos.

5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).

O

6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.

C

7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.

8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.

P

9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.

M

1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia. M
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*. C
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*). O
6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web. C
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*. M
8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas. P
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização. M

- | | |
|---|---|
| 1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia. | M |
| 2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em <i>string kernels</i> . | C |
| 3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1. | |
| 4. A relação é atribuída por votação ponderada, com base nestes exemplos. | M |
| 5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., <i>min-hashing</i>) e em dispersão sensível à localização (i.e., <i>Locality-Sensitive Hashing</i>). | O |
| 6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web. | C |
| 7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por <i>min-hashing</i> . | M |
| 8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas. | P |
| 9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização. | M |

- | | |
|---|---|
| 1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia. | M |
| 2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em <i>string kernels</i> . | C |
| 3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1. | R |
| 4. A relação é atribuída por votação ponderada, com base nestes exemplos. | M |
| 5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., <i>min-hashing</i>) e em dispersão sensível à localização (i.e., <i>Locality-Sensitive Hashing</i>). | O |
| 6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web. | C |
| 7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por <i>min-hashing</i> . | M |
| 8. No entanto, estas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas. | P |
| 9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização. | M |

CONTEXTO

PROBLEMA

OBJETIVO

METODOLOGIA

RESULTADO

6. A identificação de relações semânticas, expressas entre entidades mencionadas em textos, é um passo importante para a extração automática de conhecimento a partir de grandes coleções de documentos, tais como a Web.
2. Vários trabalhos anteriores abordaram esta tarefa para o caso da língua inglesa, usando técnicas de aprendizagem automática supervisionada para classificação de relações, sendo que o atual estado da arte recorre a métodos baseados em *string kernels*.
8. No entanto, essas abordagens requerem dados de treino anotados manualmente para cada tipo de relação, além de que os mesmos têm problemas de escalabilidade para as dezenas ou centenas de diferentes tipos de relações que podem ser expressas.
5. Este artigo discute uma abordagem com supervisão distante para a extração de relações de textos escritos em português, a qual usa uma técnica eficiente para a medição de similaridade entre exemplares de relações, baseada em valores mínimos de dispersão (i.e., *min-hashing*) e em dispersão sensível à localização (i.e., *Locality-Sensitive Hashing*).
1. No método proposto, os exemplos de treino são recolhidos automaticamente da Wikipédia, correspondendo a frases que expressam relações entre pares de entidades extraídas da DBPédia.
9. Estes exemplos são representados como conjuntos de tetragramas de caracteres e de outros elementos representativos, sendo os conjuntos indexados numa estrutura de dados que implementa a ideia da dispersão sensível à localização.
7. Procuram-se os exemplos de treino mais similares para verificar qual a relação semântica que se encontra expressa entre um determinado par de entidades numa frase, com base numa aproximação ao coeficiente de Jaccard obtida por *min-hashing*.
4. A relação é atribuída por votação ponderada, com base nestes exemplos.
3. Testes com um conjunto de dados da Wikipédia comprovam a adequabilidade do método proposto, tendo sido extraídos 10 tipos diferentes de relações, 8 deles assimétricos, com uma pontuação média de 55,6% em termos da medida F1.

Introdução

- Se o resumo “vende o peixe”, a introdução “fecha o negócio”
- Detalhamento das informações que já apareceram no resumo + revisão da literatura
- Deve guiar o leitor de forma coerente e agradável, partindo-se de um contexto geral até o assunto específico tratado pelo trabalho
 - Evite “saltar” de um assunto para outro sem que haja uma ligação clara entre os dois
- É uma das partes mais difíceis de se escrever

Introdução

- Estrutura típica de introdução
- Similar a do resumo, com adição de novos componentes
- Embora resumo e introdução tenham estruturas similares, o nível de detalhamento é diferente
 - No resumo cada componente é realizado por poucas (as vezes uma única) sentenças
 - Na introdução cada componente normalmente é realizado por um ou mais parágrafos

Contexto

Revisão da literatura

Problema de pesquisa

Objetivos

Metodologia

Principais resultados

Limitações

Contribuições

Estrutura do trabalho