



**UNIVERSITY
OF LONDON** | INTERNATIONAL
PROGRAMMES

Introduction to natural language processing

R. Kibble

C03354

2013

Undergraduate study in
Computing and related programmes

This is an extract from a subject guide for an undergraduate course offered as part of the University of London International Programmes in Computing. Materials for these programmes are developed by academics at Goldsmiths.

For more information, see: www.londoninternational.ac.uk

Goldsmiths
UNIVERSITY OF LONDON

Contents

Preface	1
About this half unit	1
Assessment	1
The subject guide and other learning resources	2
Suggested study time	2
Acknowledgement	3
1 Introduction: how to use this subject guide	5
1.1 Introduction	5
1.2 Aims of the course	5
1.3 Learning outcomes	6
1.4 Reading list and other learning resources	6
1.5 Software requirements	8
1.6 How to use the guide/structure of the course	8
1.6.1 Chapter 2: Introducing NLP: patterns and structures in language	8
1.6.2 Chapter 3: Getting to grips with natural language data	8
1.6.3 Chapter 4: Computational tools for text analysis	9
1.6.4 Chapter 5: Statistically-based techniques for text analysis	9
1.6.5 Chapter 6: Analysing sentences: syntax and parsing	9
1.6.6 Appendices	9
1.7 What the course does not cover	9
2 Introducing NLP: patterns and structure in language	11
Essential reading	11
Recommended reading	11
Additional reading	11
2.1 Learning outcomes	11
2.2 Introduction	12
2.3 Basic concepts	12
2.3.1 Tokenised text and pattern matching	12
Activity: Recognising names	13
2.3.2 Parts of speech	13
Activity: identify parts of speech	14
2.3.3 Constituent structure	14
Activity: Writing production rules	15
2.4 A closer look at syntax	15
2.4.1 Operation of a finite-state machine	16
Activity: Finite-state machines	17
2.4.2 Representing finite-state machines	17
2.4.3 Declarative alternatives to finite-state machines	18
Activity: Coding regular expressions	19
Activity: tree diagrams for a regular language	21
2.4.4 Limitations of finite-state methods – introducing context-free grammars	21
Activity: Regular grammars	21
Activity: Context-free grammar	23
2.4.5 Looking ahead: some further uses of regular expressions	23



**UNIVERSITY
OF LONDON** | INTERNATIONAL
PROGRAMMES

Introduction to natural language processing

R. Kibble

C03354

2013

Undergraduate study in
Computing and related programmes

This is an extract from a subject guide for an undergraduate course offered as part of the University of London International Programmes in Computing. Materials for these programmes are developed by academics at Goldsmiths.

For more information, see: www.londoninternational.ac.uk

Goldsmiths
UNIVERSITY OF LONDON

Contents

Preface	1
About this half unit	1
Assessment	1
The subject guide and other learning resources	2
Suggested study time	2
Acknowledgement	3
1 Introduction: how to use this subject guide	5
1.1 Introduction	5
1.2 Aims of the course	5
1.3 Learning outcomes	6
1.4 Reading list and other learning resources	6
1.5 Software requirements	8
1.6 How to use the guide/structure of the course	8
1.6.1 Chapter 2: Introducing NLP: patterns and structures in language	8
1.6.2 Chapter 3: Getting to grips with natural language data	8
1.6.3 Chapter 4: Computational tools for text analysis	9
1.6.4 Chapter 5: Statistically-based techniques for text analysis	9
1.6.5 Chapter 6: Analysing sentences: syntax and parsing	9
1.6.6 Appendices	9
1.7 What the course does not cover	9
2 Introducing NLP: patterns and structure in language	11
Essential reading	11
Recommended reading	11
Additional reading	11
2.1 Learning outcomes	11
2.2 Introduction	12
2.3 Basic concepts	12
2.3.1 Tokenised text and pattern matching	12
Activity: Recognising names	13
2.3.2 Parts of speech	13
Activity: identify parts of speech	14
2.3.3 Constituent structure	14
Activity: Writing production rules	15
2.4 A closer look at syntax	15
2.4.1 Operation of a finite-state machine	16
Activity: Finite-state machines	17
2.4.2 Representing finite-state machines	17
2.4.3 Declarative alternatives to finite-state machines	18
Activity: Coding regular expressions	19
Activity: tree diagrams for a regular language	21
2.4.4 Limitations of finite-state methods – introducing context-free grammars	21
Activity: Regular grammars	21
Activity: Context-free grammar	23
2.4.5 Looking ahead: some further uses of regular expressions	23