

Acknowledgement

Brief Contents	v
Contents	vii
1 Introduction	13
1.1 Motivation.....	13
1.2 Outline of the Book.....	13
1.3 The Prague Dependency Treebank 2.0.....	15
1.3.1 The Morphological Layer.....	15
1.3.2 The Analytical Layer.....	15
1.3.3 The Tectogrammatical Layer.....	16
2 The Problem Analysis	17
2.1 Related Work.....	17
2.1.1 More or Less Theoretical Papers.....	17
2.1.2 Existing Search Tools.....	18
Manatee/Bonito.....	18
TGrep.....	18
TGrep2.....	19
TigerSearch.....	19
Oraculum.....	20
TrEd.....	20
VIQTORYA.....	21
Fsq.....	21
2.2 Netgraph 1.0 – The Starting Point.....	22
2.3 Linguistic Phenomena in PDT 2.0.....	23
2.3.1 The Tectogrammatical Layer.....	24
Basic Principles.....	24
Valency.....	24
Coordination and Apposition.....	25
Idioms (Phrasemes) etc.....	25
Complex Predicates.....	26
Predicative Complement (Dual Dependency).....	26
Coreferences.....	27
Topic-Focus Articulation.....	28
2.3.2 Accessing Lower Layers.....	30
2.3.3 The Analytical Layer (and Lower Layers).....	31
Morphological Tags.....	31

Agreement.....	32
Word Order.....	32
2.4 Linguistic Requirements.....	32
2.4.1 Complex Evaluation of a Node.....	32
2.4.2 Dependencies Between Nodes (Vertical Relations).....	33
2.4.3 Horizontal Relations.....	33
2.4.4 Other Features.....	33
3 The Query Language	35
3.1 The Basics.....	35
3.2 Alternative Values and Nodes.....	36
3.2.1 Alternative Values.....	36
3.2.2 Alternative Nodes.....	36
3.3 Wild Cards.....	36
3.4 Regular Expressions.....	37
3.5 Dependencies Between Nodes.....	38
3.6 Arithmetic Expressions.....	39
3.7 Other Relations.....	39
3.8 Meta-Attributes.....	40
3.8.1 <code>_transitive</code>	40
3.8.2 <code>_optional</code>	42
3.8.3 <code>_#sons</code>	45
3.8.4 <code>_#hsons</code>	45
3.8.5 <code>_depth</code>	46
3.8.6 <code>_#descendants</code>	46
3.8.7 <code>_#lbrothers</code>	47
3.8.8 <code>_#rbrothers</code>	47
3.8.9 <code>_#occurrences</code>	47
3.8.10 <code>_name</code>	50
3.8.11 <code>_sentence</code>	50
3.9 References.....	50
3.10 Multi-Tree Queries.....	52
3.11 Hidden Nodes.....	54
4 The Data	57
4.1 The Format.....	57
4.2 Corpus-Specific Features in the Header.....	58
4.3 How Data Can Help.....	59
4.4 References.....	59
4.5 Attribute <code>m/lemma</code>	60
4.6 Hidden Nodes.....	61

5 Using the Query Language	63
5.1 General Requirements	63
5.1.1 Complex Evaluation of a Node	63
5.1.2 Dependencies Between Nodes (Vertical Relations)	63
5.1.3 Horizontal Relations	64
5.1.4 Other Features	65
5.2 Using the Query Language for Searching in PDT 2.0	65
5.2.1 The Tectogrammatical Layer	65
Basic Principles	65
Valency	65
Coordination and Apposition	66
Idioms (Phrasemes) etc	67
Complex Predicates	67
Predicative Complement (Dual Dependency)	68
Coreferences	69
Topic-Focus Articulation	70
5.2.2 Accessing Lower Layers	77
5.2.3 The Analytical Layer	79
Morphological Tags	79
Agreement	79
Word Order	80
6 Notes on the Query Language	81
6.1 Netgraph Query Language vs. FS Query Language	81
6.2 Trees Only	81
6.3 Redundancy	81
6.3.1 Two Types of Redundancy	82
6.4 Result Trees and Result Occurrences	83
6.5 Comparison to Other Treebank Query Systems	84
6.5.1 A Biased Table	84
6.5.2 Comparison to TGrep	86
6.5.3 Comparison to TGrep2	88
6.5.4 Comparison to TigerSearch	90
Node Description	90
Node Relations	91
Negation	92
Graph Description	93
Variables	93
Graph Predicates	93
6.5.5 Why Is It So Complex in Netgraph?	94
6.6 Universality	95
6.7 Feedback From Users	95

7 The Tool	99
7.1 Properties of the Tool.....	99
7.2 Changes since Version 1.0.....	101
7.2.1 Main Extensions to the Query Language.....	101
7.2.2 Main Extensions to the Tool.....	102
8 Real World	103
8.1 The Queries.....	104
8.1.1 One-Node Queries.....	104
8.1.2 Structured Queries without Meta-Attributes.....	105
8.1.3 Queries with Meta-Attributes.....	107
_transitive.....	107
_optional.....	107
_sons.....	108
_depth.....	108
_descendants.....	109
_lbrothers, _rbrothers.....	109
_occurrences.....	110
_name.....	110
_hsons.....	110
_sentence.....	110
8.1.4 Queries with References.....	111
8.1.5 Queries with Hidden Nodes.....	111
9 Conclusion	115
9.1 What Has Been Done.....	115
9.2 Future Work.....	115
9.2.1 The Query Language.....	115
Simplification.....	116
Further Extensions.....	117
Corpus-Wide Comparing and Statistics.....	118
9.2.2 Speed.....	119
9.2.3 Further Improvements.....	119
10 Appendixes	121
10.1 Appendix A: Publications about Netgraph.....	122
10.2 Appendix B: FS File Format Description.....	124
10.2.1 Notes on Metasyntax.....	124
10.2.2 The FS File Structure.....	124
10.2.3 Identifiers, Attribute Names and Values.....	124
10.2.4 Node Attributes Definition.....	125
Properties.....	125
10.2.5 A Tree.....	126

10.2.6 A Node.....	127
10.3 Appendix C: FS Query Format Description.....	128
10.3.1 The FS Query Structure.....	128
10.3.2 A Node.....	128
10.3.3 Attribute Values.....	129
10.4 Appendix D: List of Attributes in PDT 2.0.....	130
10.4.1 The Word Layer.....	130
w/token (w/token at hidden nodes).....	130
w/no_space_after (w/no_space_after at hidden nodes).....	130
w/id.....	130
10.4.2 The Morphological Layer.....	130
m/form (m/form at hidden nodes).....	130
m/form_change.....	130
m/id.....	130
m/lemma (m/lemma at hidden nodes).....	131
m/src.rf.....	131
m/tag (m/tag at hidden nodes).....	131
10.4.3 The Analytical Layer.....	132
afun (a/afun at hidden nodes).....	132
eparents (a/eparents at hidden nodes).....	133
eparents_diff (a/eparents_diff at hidden nodes).....	133
id (a/id at hidden nodes).....	133
is_member (a/is_member at hidden nodes).....	134
is_parenthesis_root.....	134
ord (a/ord at hidden nodes).....	134
s.rf.....	134
- (a/parent at hidden nodes).....	134
- (a/ref_type at hidden nodes).....	134
10.4.4 The Tectogrammatical Layer.....	134
atree.rf.....	134
compl.rf.....	135
coref_gram.rf.....	135
coref_special.....	135
coref_text.rf.....	135
deepord.....	135
eparents.....	135
eparents_diff.....	135
functor.....	136
Grammatemes (attributes gram/*).....	137
id.....	138
is_dsp_root.....	138
is_generated.....	138
is_member.....	138
is_name_of_person.....	138

is_parenthesis.....	138
is_state.....	138
nodetype.....	139
quot/set_id.....	139
quot/type.....	139
sentence.....	139
sentmod.....	139
subfunctor.....	139
t_lemma.....	139
tfa.....	140
val_frame.rf.....	140
hide.....	140
10.5 Appendix E: Other Usages of Netgraph.....	141
10.5.1 Morphological “Trees” of the Czech Academic Corpus 1.0.....	141
10.5.2 Latin IT Treebank.....	142
10.5.3 Arabic Trees.....	142
10.5.4 Chinese Treebank.....	143
10.5.5 Vallex.....	144
10.6 Appendix F: Installation and Usage of Netgraph – A Quick How-To.....	145
10.6.1 Installation.....	145
Java 2 Installation.....	145
Netgraph Client/Server Installation.....	145
10.6.2 Connection to the Public Netgraph Server for PDT 2.0.....	146
10.6.3 Connection to the Local Netgraph Server for PDT 2.0 Sample Data.....	146
10.6.4 Selection of Files for Searching.....	147
10.6.5 Creation of a Simple Query.....	148
Summary	151
Bibliography	153
Index	157