

C96-1058

5 sentences summary:

1: yet, they can be parsed in **$O(n^3)$** time (eisner, 1996).

2: if projectivity (no crossing branches) is desired, eisner's (1996) **dynamic programming algorithm** (similar to cyk) for dependency parsing can be used instead.

3: first, in supervised models, a head out-ward process is modeled (eisner, 1996; collins, 1999).

4: see for example (lombardi, 1996; eisner, 1996), who also discuss early-style parsers for **projective dependency grammars**.

5: in the context of dps, this **edge based factorization method** was proposed by (eisner, 1996).

Fact id pyramid tier allocation: {'11': 2, '10': 3, '12': 2, '1': 1, '3': 4, '2': 2, '5': 2, '4': 7, '7': 6, '6': 2, '9': 4, '8': 12}

Optimal summary total facts weights: 43

Optimal summary covered facts: {'10': ['sum of the score of all edges'], '12': ['data-driven'], '1': ['pos tags'], '3': ['bottom-up', 'span', 'notion of a span'], '5': ['data-driven dependency'], '4': ['generative model'], '7': ['projective trees'], '6': ['penn treebank'], '9': ['edge based factorization method', 'edge based factorization'], '8': [' $O(n^3)$ ', ' $O(n^3)$ time']}

Summary total facts weights: 24

Summary covered facts: {'11': ['dynamic programming algorithm'], '8': [' $O(n^3)$ '], '9': ['edge based factorization method', 'edge based factorization'], '7': ['projective dependency grammars']}

Summary pyramid score: 0.558139534884

P97-1003

5 sentences summary:

1: in particular, the model in collins (1997) failed to generate **punctuation**, a deficiency of the model.

2: in the field of **statistical parsing**, various **probabilistic evaluation models** have been proposed where different models use different feature types [black, 1992] [briscoe, 1993] [brown, 1991] [charniak, 1997] [collins, 1996] [collins, 1997] [magerman, 1991] [magerman, 1992] [magerman, 1995] [eisner, 1996].

3: (collins 1997, 1999; charniak 2000), and the current paper has shown the importance of including two and more **nonhead words**.

4: table 1 shows the lp and lr scores obtained with our base line subtree set, and compares these scores with those of previous stochastic parsers tested on the wsj (respectively charniak 1997, collins 1999, ratnaparkhi 1999, and charniak 2000).

5: there has been a great deal of progress in **statistical parsing** in the past decade (collins, 1996; collins, 1997; chaniak, 2000).

Fact id pyramid tier allocation: {'1': 11, '3': 6, '2': 3, '5': 4, '4': 9}

Optimal summary total facts weights: 33

Optimal summary covered facts: {'1': ['probabilistic context-free grammars'], '3': ['punctuation'], '2': ['constituents'], '5': ['language'], '4': ['head-child']}

Summary total facts weights: 26

Summary covered facts: {'1': ['statistical parsing', 'probabilistic evaluation models', 'statistical parsing'], '3': ['punctuation'], '4': ['nonhead words']}

Summary pyramid score: 0.787878787879

P99-1065

5 sentences summary:

1: **czech** results for the **czech** data, we used the predefined training, development and testing split of the prague dependency treebank (hajič et al, 2001), and the automatically generated pos tags supplied with the data, which we reduce to the **pos tag** set from collins et al (1999).

2: the appendices of collins (1999) give a precise description of the parsing algorithms, an analysis of their computational complexity, and also a description of the pruning methods that are employed.

3: the usage of special knowledge bases to determine projections of categories (xia and palmer, 2001) would have presupposed language-dependent knowledge, so we investigated two other options: **flat rules** (collins et al, 1999) and binary rules.

4: the trees are then transformed into penn treebank style **constituencies** using the technique described in (collins et al, 1999).

5: see appendix a of collins (1999) for a description of how **the head rules** treat phrases involving coordination.

Fact id pyramid tier allocation: {'1': 6, '3': 6, '2': 13, '5': 2, '4': 2, '7': 5, '6': 2}

Optimal summary total facts weights: 34

Optimal summary covered facts: {'1': ['lexicalized'], '3': ['tag classification'], '2': ['czech treebank', 'czech', 'czech', 'czech parser'], '5': ['punctuation signs'], '4': ['flat rules'], '7': ['constituency-based parsers', 'constituency']}

Summary total facts weights: 26

Summary covered facts: {'3': ['pos tag'], '2': ['czech'], '4': ['flat rules', 'the head rules'], '7': ['constituencies']}

Summary pyramid score: 0.764705882353

P05-1013

5 sentences summary:

1: graph transformations for recovering **non-projective** structures (nivre and nilsson, 2005).

2: recent work by nivre and nilsson introduces a technique where the **projectivization transformation** is encoded in the non-terminals of constituents during parsing (nivre and nilsson, 2005).

3: although the parser only derives **projective graphs**, the fact that graphs are labeled allows **non-projective** dependencies to be captured using the pseudoprojective approach of nivre and nilsson (2005) .

4: we projectivize **training data** by a minimal transformation, lifting **non-projective** arcs one step at a time, and extending the arc label of lifted arcs using the encoding scheme called head by nivre and nilsson (2005), which means that a lifted arc is assigned the label $r\hat{\langle}86\rangle\langle 91\rangle h$, where r is the original label and h is the label of the original head in the nonprojective dependency graph.

5: nivre and nilsson (2005) presented a parsing model that allows for the introduction of **non-projective** edges into dependency trees through learned edge transformations within their memory-based parser.

Fact id pyramid tier allocation: {'1': 19, '3': 1, '2': 6, '5': 4, '4': 8}

Optimal summary total facts weights: 38

Optimal summary covered facts: {'1': ['projectivization transformation', 'projectivizing'], '3': ['nonterminal categories in constituency'], '2': ['data-driven', 'training data'], '5': ['maltparser'], '4': ['czech']}

Summary total facts weights: 25

Summary covered facts: {'1': ['non-projective', 'projectivization transformation', 'projective graphs', 'non-projective', 'non-projective', 'non-projective'], '2': ['training data']}

Summary pyramid score: 0.657894736842

P05-1012

5 sentences summary:

1: 3.1 decoding mcdonald et al (2005b) use the chu-liuedmonds (cle) algorithm to solve the maximum **spanning tree** problem.

2: to learn these structures we used online **large-margin** learning (mcdonald et al, 2005) that empirically provides state-of-the-art performance for czech.

3: while we have presented significant improvements using additional constraints, one may wonder even when caching feature extraction during training mcdonald et al (2005a) still takes approximately 10 minutes to train.

4: mcdonald et al (2005a) introduce a dependency parsing framework which treats the task as searching for the **projective tree** that **maximises the sum of local dependency scores**.

5: we take as our starting point a re-implementation of mcdonald's state-of-the-art dependency parser (mcdonald et al, 2005a).

Fact id pyramid tier allocation: {'1': 10, '3': 7, '2': 2, '5': 7, '4': 13, '7': 3, '6': 2}

Optimal summary total facts weights: 44

Optimal summary covered facts: {'1': ['spanning tree'], '3': ['large margin'], '2': ['sum of local score'], '5': ['mira'], '4': ['non-projective', 'projective'], '7': ['intervening material'], '6': ['o(n²) time']}

Summary total facts weights: 32

Summary covered facts: {'1': ['spanning tree'], '3': ['large-margin'], '2': ['maximises the sum of local dependency scores'], '4': ['projective tree', 'projective']}

Summary pyramid score: 0.727272727273

N03-1017

5 sentences summary:

1: using giza++ model 4 alignments and **pharaoh** (koehn et al, 2003), we achieved a bleu score of 0.3035.

2: word alignment is an important component of a complete **statistical** machine translation pipeline (koehn et al, 2003).

3: the baseline system we used for comparison was **pharaoh** (koehn et al, 2003; koehn, 2004a), as publicly distributed.

4: modifiers within german clauses were translated using the phrase-based model of koehn et al (2003).

5: **phrase-pairs** are then extracted from the word alignments (koehn et al, 2003).

Fact id pyramid tier allocation: {'1': 21, '3': 26, '2': 12, '5': 10, '4': 10, '7': 21, '6': 11, '8': 3}

Optimal summary total facts weights: 114

Optimal summary covered facts: {'1': ['phrase pair'], '3': ['statistical'], '2': ['linear interpolation of these scores'], '5': ['refined alignment'], '4': ['distortion model'], '7': ['pharaoh'], '6': ['heuristic'], '8': ['limiting phrase length to three']}

Summary total facts weights: 68

Summary covered facts: {'1': ['phrase-pair'], '3': ['statistical'], '7': ['pharaoh', 'pharaoh']}

Summary pyramid score: 0.59649122807

W03-0301

5 sentences summary:

1: the experiments reported here were carried out using data from the **workshop** on building and using parallel texts held at hlt-naacl 2003 (mihalcea and pedersen, 2003).

2: we also compared the **performance** on the 447 test sentences to 1/ the intersection of the alignments produced by the top ibm4 alignments in either directions, and 2/ the best systems from (mihalcea and pedersen, 2003).

3: it was the basis for a system that performed very well in a comparison of several alignment systems (dejean et al, 2003; mihalcea and pedersen, 2003).

4: we trained and evaluated our various modifications to model 1 on data from the bilingual word alignment **workshop** held at hlt-naacl 2003 (mihalcea and pedersen, 2003).

5: for parameter tuning, we used the 17 sentence trial set from the romanian-english corpus in the **2003 naacl task** (mihalcea and pedersen, 2003).

Fact id pyramid tier allocation: {'1': 2, '3': 2, '2': 6}

Optimal summary total facts weights: 10

Optimal summary covered facts: {'1': ['french/english'], '3': ['performance', 'performance measures'], '2': ['workshop', 'workshop', '2003 naacl shared task', 'shared task']}

Summary total facts weights: 8

Summary covered facts: {'3': ['performance'], '2': ['workshop', 'workshop', '2003 naacl task']}

Summary pyramid score: 0.8

5 sentences summary:

1: unless otherwise noted, the following discussion is generally applicable to **alignment template** systems (och and ney 2004) as well.

2: a lexicalized phrase **reordering model** like that in use in isi's system (och et al, 2004) might be able to learn a **better reordering**, but simpler distortion models will probably not.

3: the current state of the art is represented by the so-called **phrase-based** translation approach (och and ney, 2004; koehn et al, 2003).

4: phrase pairs are extracted by following the method described in (och and ney, 2004) where all contiguous phrase pairs having **consistent** alignments are extraction candidates.

5: nonetheless, attempts to incorporate richer linguistic features have generally met with little success (och et al, 2004a).

Fact id pyramid tier allocation: {'1': 4, '3': 9, '2': 15, '5': 5, '4': 4}

Optimal summary total facts weights: 37

Optimal summary covered facts: {'1': ['notion of consistency'], '3': ['alignment template'], '2': ['phrase-based'], '5': ['reordering model', 'better reordering'], '4': ['bilingual phrases', 'bilingual phrase']}

Summary total facts weights: 33

Summary covered facts: {'1': ['consistent'], '3': ['alignment template'], '2': ['phrase-based'], '5': ['reordering model', 'better reordering']}

Summary pyramid score: 0.891891891892

N04-1033

5 sentences summary:

1: (zens and ney, 2004) obtain $p(s|t)$ from **smoothed relative-frequency** estimates in a wordaligned corpus.

2: we define the source and target projection of a hypothesis h by the proj operator which collects in order the words of a hypothesis along one language: $\text{proj}_f(h) = \text{braceleftbig fp} : \text{p}\hat{\text{a}}\langle 88 \rangle \langle 88 \rangle \text{uniontextui}=1 \{ \text{j} \text{in} \} \hat{\text{n}}\langle 88 \rangle \langle 88 \rangle [1, \text{ni}] \text{bracerightbig}$ $\text{proj}_e(h) = \text{braceleftbig ep} : \text{p}\hat{\text{a}}\langle 88 \rangle \langle 88 \rangle \text{uniontextui}=1 \{ \text{lim} \} \hat{\text{m}}\langle 88 \rangle \langle 88 \rangle [1, \text{mi}] \text{bracerightbig}$ if we denote by h_f the set of hypotheses that have f as a source projection (that is, $h_f = \{ h : \text{proj}_f(h) \hat{\text{a}}\langle 89 \rangle_i f \}$), then our translation engine seeks $\hat{e} = \text{proj}_e(\hat{h})$ where: $\hat{h} = \text{argmax}_h \hat{h} \langle 88 \rangle \langle 88 \rangle h_f s(h)$ the function we seek to maximize $s(h)$ is a loglinear combination of 9 components, and might be better understood as the numerator of a maximum entropy model popular in several **statistical mt** systems(ochandney, 2002; bertoldietal., 2004; zens and ney, 2004; simard et al, 2005; quirk et al, 2005).

3: for the confidence measures which will be introduced in section 5, we use a state-of-the-art phrasebased approach as described in (zens and ney, 2004).

4: our approach to **phrase-table smoothing** contrasts to previous work (zens and ney, 2004) in which smoothed phrase probabilities are constructed from word-pair probabilities and combined in a log-linear model with an unsmoothed phrase-table.

5: finally, we use pos features to parameterize a distortion model in a **limited distortion decoder** (zens and ney, 2004; tillmann and zhang, 2005).

Fact id pyramid tier allocation: {'1': 4, '3': 1, '2': 3, '5': 1, '4': 5}

Optimal summary total facts weights: 14

Optimal summary covered facts: {'1': ['statistical machine translation model'], '3': ['supervised manner'], '2': ['two levels of distortion'], '5': ['bilingual phrases'], '4': ['phrase-table smoothing']}

Summary total facts weights: 12

Summary covered facts: {'1': ['statistical mt'], '2': ['a limited distortion decoder'], '4': ['smoothed relative-frequency', 'phrase-table smoothing']}

Summary pyramid score: 0.857142857143

P05-1033

5 sentences summary:

1: so i co-developed several fast and exact algorithms for kbest parsing in the general framework of directed monotonic hypergraphs (Huang and Chiang, 2005).

2: so instead of determinization, here we present a simple-yet-effective extension to the algorithm 3 of Huang and TARGETANCHOR that guarantees to output unique translated strings: • keep a hash-table of unique strings at each vertex in the hypergraph • when asking for the next-best derivation of a vertex, keep asking until we get a new string, and then add it into the hash-table this method should work in general for any equivalence relation (say, same derived tree) that can be defined on derivations.

3: as illustrated in figure 2, vpp-vp has contiguous spans on both source and target sides, so that we can generate a binary-branching **scfg**: $(2) s \rightarrow np(1) v pp-vp(2), np(1) v pp-vp(2) vpp-vp \rightarrow vp(1) pp(2), pp(2) vp(1)$ in this case m-gram integrated decoding can be done in $O(|w|^{3+4(m-1)})$ time which is much lower order polynomial and no longer depends on rule size (Wu, 1996), allowing the search to be much faster and more accurate facing **pruning**, as is evidenced in the **hiero** system of TARGETANCHOR where he restricts the **hierarchical** phrases to be a binary **scfg**.

4: these algorithms have been re-implemented by other researchers in the field, including Eugene Charniak for his n-best parser, Ryan McDonald for his dependency parser (McDonald et al, 2005), Microsoft Research NLP group (Simon Kevin Duh, p.c.) for a similar model, Jonathan Graehl for the ISI **syntax-based** mt decoder, David A. Smith for the Dyna language (Eisner et al, 2005), 223 and Jonathan May for ISI's tree automata package Tiburon.

5: variations of scfgs go back to Aho and Ullman (1972) **syntax-directed** translation schemata, but also include the inversion transduction grammars in Wu (1997), which restrict grammar rules to be binary, the **synchronous grammar** in TARGETANCHOR, which use only **a single nonterminal symbol**, and the multitext grammars in Melamed (2003), which allow independent rewriting, as well as other tree-based models such as Yamada and Knight (2001) and Galley et al (2004).

Fact id pyramid tier allocation: {'1': 20, '3': 12, '2': 20, '5': 2, '4': 2, '6': 3}

Optimal summary total facts weights: 59

Optimal summary covered facts: {'1': ['synchronous context-free grammars', 'synchronous context-free grammar', 'synchronous cfg', 'scfg', 'synchronous grammar', 'scfg'], '3': ['syntax-directed'], '2': ['hiero', 'hierarchical'], '5': ['a single nonterminal symbol'], '4': ['chinese-english'], '6': ['pruning']}

Summary total facts weights: 57

Summary covered facts: {'1': ['scfg', 'synchronous grammar', 'scfg'], '3': ['syntax-based', 'syntax-directed'], '2': ['hiero', 'hierarchical'], '5': ['a single nonterminal symbol'], '6': ['pruning']}

Summary pyramid score: 0.966101694915

A00-1043

5 sentences summary:

1: many algorithms exploit parallel corpora (jing 2000; knight and marcu 2002; riezler et al 2003; nguyen et al 2004a; turner and charniak 2005; mcdonald 2006) to learn the correspondences between long and short sentences in a supervised manner, typically using a rich feature space **induced from parse trees**.

2: for example, jing (2000) trained her system on a set of 500 sentences from the benton foundation (<http://www.benton.org>) and their reduced forms written by humans.

3: depending on the chosen task, such systems either generate single-sentence “headlines” for multi-sentence text (witbrock and mittal, 1999), or they provide a sentence condensation module designed for combination with sentence extraction systems (knight and marcu, 2000; jing, 2000).

4: in contrast to jing (2000), the bulk of the research on sentence compression relies exclusively on corpus data for modelling the compression process without recourse to extensive knowledge sources (e.g., wordnet).

5: jing (2000) proposes a novel algorithm for sentence reduction that takes into account different sources of information to decide whether or not to remove a particular component from a sentence to be included in a summary.

Fact id pyramid tier allocation: {'1': 4, '3': 1, '2': 2, '4': 1}

Optimal summary total facts weights: 8

Optimal summary covered facts: {'1': ['grammar checking', 'induced from parse trees'], '3': ['based on context-free deletion decisions'], '2': ['evaluation of sentence reduction', 'automatic evaluation method'], '4': ['using multiple source of knowledge']}

Summary total facts weights: 4

Summary covered facts: {'1': ['induced from parse trees']}

Summary pyramid score: 0.5

A00-2024

5 sentences summary:

1: table 3: example compressions compression avg len rating baseline 9.70 1.93 bt-2-step 22.06 3.21 spade 19.09 3.10 humans 20.07 3.83 table 4: mean ratings for automatic compressions naturally, we added a simple baseline **compression algorithm** proposed by jing and mckeown (2000) which removed all prepositional phrases, clauses, toinfinitives, and gerunds.

2: we analyzed a set of articles and identified six major operations that can be used for **editing the extracted sentences**, including removing extraneous phrases from an extracted sentence, combining a reduced sentence with other sentences, syntactic transformation, substituting phrases in an extracted sentence with their paraphrases, substituting phrases with more general or specific descriptions, and reordering the extracted sentences (jing and mckeown, 1999; jing and mckeown, 2000).

3: while earlier approaches for text compression were based on symbolic reduction rules (grefenstette 1998; mani, gates, and bloedorn 1999), more recent approaches use an aligned corpus of documents and their **human written summaries** to determine which constituents can be reduced (knight and marcu 2002; jing and mckeown 2000; reizler et al 2003).

4: as previously observed in the literature (mani, gates, and bloedorn 1999; jing and mckeown 2000), such components include a clause in the clause conjunction, relative clauses, and some elements within a clause (such as adverbs and prepositions).

5: because of this, it is generally accepted that some kind of **postprocessing should be performed** to improve the final result, by shortening, fusing, or otherwise revising the material (grefenstette 1998; mani, gates, and bloedorn 1999; jing and mckeown 2000; barzilay et al 2000; knight and marcu 2000).

Fact id pyramid tier allocation: {'1': 5, '3': 2, '2': 4, '5': 2, '4': 2}

Optimal summary total facts weights: 15

Optimal summary covered facts: {'1': ['editing extracted text spans'], '3': ['compression algorithm'], '2': ['cutting and pasting'], '5': ['human-written'], '4': ['rule-based algorithm']}

Summary total facts weights: 9

Summary covered facts: {'1': ['editing the extracted sentences', 'postprocessing should be performed'], '3': ['compression algorithm'], '5': ['human written summaries']}

Summary pyramid score: 0.6

C00-1072

5 sentences summary:

1: two methods are used: **topic signature** (lin and hovy, 2000): a **topic signature** is a family of related terms {topic, signature}, where topic is the target concept and signature is a vector related to it.

2: 2 general summarization models many summarization systems (e.g., (teufel and moens, 1997; mckeown et al, 1999; lin and hovy, 2000)) include two levels of analysis: the sentence level, where every textual unit is scored according to c1 c2 c3 c4 c5 t1 t2 t3 t4 table 1: matrix for summarization model the concepts or features it covers, and the text level, where, before being added to the final output, textual units are compared to each other on the basis of those features.

3: to deal with a lot of chinese documents which have free style of writing and flexible themes, a **sentence-extraction summarization** method created by detecting thematic areas is tried following such work as (nomoto and matsumoto, 2001; salton et al, 1996; salton et al, 1997; carbonell and goldstein, 1998; lin and hovy, 2000).

4: to date, researchers have harvested, with varying success, several resources, including concept lists (lin and pantel 2002), **topic signatures** (lin and hovy 2000), facts (etzioni et al 2005), and word similarity lists (hindle 1990).

5: in order to generate quabs automatically, documents identified from ferret's automatic q/a system are first submitted to a **topic representation** module, which computes both **topic signatures** (lin and hovy, 2000) and enhanced **topic signatures** (harabagiu, 2004) in order to identify a set of topic-relevant passages.

Fact id pyramid tier allocation: {'1': 1, '3': 1, '2': 13}

Optimal summary total facts weights: 15

Optimal summary covered facts: {'1': ['sentence-extraction summarization'], '3': ['co-occurrence of particular concepts'], '2': ['topic signature', 'topic signature', 'n-gram key concepts', 'topic signature']}

Summary total facts weights: 14

Summary covered facts: {'1': ['sentence-extraction summarization'], '2': ['topic signature', 'topic signature', 'topic signature', 'topic representation']}

Summary pyramid score: 0.933333333333

W00-0403

5 sentences summary:

1: but the interpretation of the results is not simple; studies (jing et al 1998; donaway, drummey, and mather 2000; radev, jing, and budzikowska 2000) 404 computational linguistics volume 28, number 4 show how the same summaries receive different scores under different measures or when compared to different (but presumably equivalent) ideal summaries created by humans.

2: other research rewards passages that include topic words, that is, words that have been determined to correlate well with the topic of interest to the user (for topic-oriented summaries) or with the general theme of the source text (buckley and cardie 1997; strzalkowski et al 1999; radev, jing, and budzikowska 2000).

3: we observed the best results with maximal marginal relevance (mmr) (carbonell and goldstein, 1998) reranker and the default reranker of the system based on cross-sentence informational sub-sumption (csis) (radev, 2000).

4: first, our method focuses on subject shift of the documents from the target event rather than the **sets of documents from different events** (radev et al, 2000).

5: **mead** (radev et al, 2000): **mead** is a **centroid-based extractive summarizer** that scores sentences based on sentence-level and inter-sentence features which indicate the quality of the sentence as a summary sentence.

Fact id pyramid tier allocation: {'1': 7, '3': 4, '2': 7, '5': 4, '4': 1, '6': 3}

Optimal summary total facts weights: 26

Optimal summary covered facts: {'1': ['relative utility'], '3': ['cluster'], '2': ['top scoring sentences', 'centroid-based summarization', 'centroid'], '5': ['multi-document summarization'], '4': ['statistical methods'], '6': ['mead']}

Summary total facts weights: 14

Summary covered facts: {'2': ['centroid-based extractive summarizer', 'centroid'], '5': ['sets of documents from different events'], '6': ['mead']}

Summary pyramid score: 0.538461538462

W03-0510

5 sentences summary:

1: **unigram co-occurrence** metric in a recent study (lin and hovy, 2003a), we showed that the recall-based unigram cooccurrence automatic scoring metric correlates highly with human evaluation and has high recall and precision in predicting the statistical significance of results comparing with its human counterpart.

2: the issue of subjectivity gains prominence as the compression ratio increases, i.e., the shorter the summary, the larger the number of correct summaries (lin and hovy, 2003b).

3: these findings are additionally supported by the fact that automatic **n-gram-based evaluation measures** now being used to assess predominately extractive multi-document summarization systems correlate strongly with human judgments when restricted to the usage of unigrams and bigrams, but correlate weakly when longer n-grams are factored into the equation (lin & hovy, 2003).

4: it is also notable the study reported in (lin and hovy, 2003b) **discussing the usefulness and limitations of automatic sentence extraction** for text summarization, 23 single document meta<9d> summarization algorithm summarization algo.

5: let us imagine, for instance, that the best metric turns out to be a **rouge** (lin and hovy, 2003a) variant that only considers unigrams to compute similarity.

Fact id pyramid tier allocation: {'1': 5, '3': 5, '2': 3}

Optimal summary total facts weights: 13

Optimal summary covered facts: {'1': ['rouge', 'rouge'], '3': ['unigram co-occurrence', 'ngram statistics'], '2': ['discussing the usefulness and limitations of automatic sentence extraction', 'discussing the usefulness and limitations of automatic sentence extraction']}

Summary total facts weights: 13

Summary covered facts: {'1': ['rouge'], '3': ['unigram co-occurrence', 'n-gram-based evaluation measure'], '2': ['discussing the usefulness and limitations of automatic sentence extraction']}

Summary pyramid score: 1.0

A00-1023

5 sentences summary:

1: examples of using nlp and ie in question answering include **shallow parsing** [kupiec 1993] [srihari & li 2000], deep parsing [li et al 2002] [litkowski 1999] [voorhees 1999], and ie [abney et al 2000] [srihari & li 2000].

2: it is worth noticing that in our experiment, the structural support used for answer-**point identification** only checks the binary links involving the asking point and the candidate answer points, instead of **full template** matching as proposed in (srihari and li, 2000).

3: in response, factoid question answering systems have evolved into two types: • use-knowledge: extract query words from the input question, perform ir against the source corpus, possibly segment resulting documents, identify a set of segments containing likely answers, apply a set of heuristics that each consults a different source of knowledge to score each candidate, rank them, and select the best (harabagiu et al, 2001; hovy et al, 2001; srihari and li, 2000; abney et al, 2000).

4: assuming that it is very likely that the answer is a named entity, (srihari and li, 2000) describes a ne-supported q&a system that functions quite well when the expected answer type is one of the categories covered by the ne recognizer.

5: qa is different than search engines in two aspects: (i) instead of a string of keyword search terms, the query is a natural language question, necessitating question parsing, (ii) instead of a list of documents or urls, a list of candidate answers at phrase level or sentence level are expected to be returned in response to a query, hence the need for text processing beyond keyword indexing, typically supported by natural language processing (nlp) and information extraction (ie) (chinchor and marsh 1998, hovy, hermjakob and lin 2001, li and srihari 2000).

Fact id pyramid tier allocation: {'1': 3, '3': 1, '2': 4}

Optimal summary total facts weights: 8

Optimal summary covered facts: {'1': ['shallow parsing', 'shallow approach'], '3': ['full template'], '2': ['ne-supported q&a', 'typical named entities', 'point identification']}

Summary total facts weights: 8

Summary covered facts: {'1': ['shallow parsing'], '3': ['full template'], '2': ['point identification']}

Summary pyramid score: 1.0

W00-0603

5 sentences summary:

1: keywords in questions it has been observed in the work of (riloff and thelen, 2000) that certain words in a when or where question tend to indicate that the **dateline** is an $\sim n \sim$ wer sentence to the question.

2: refer to the readme le of minipar downloaded from <http://www.cs.ualberta.ca/index/minipar.htm> 5 experimental results we selected the features used in **quarc** (riloff and thelen, 2000) to establish the reference performance level.

3: 4 evaluation to evaluate our learning approach, we trained aquarea\$ on the same development set of stories and tested it on the same test set of stories as those used in all past work on the **reading comprehension** task (hirschman et al, 1999; charniak et al, 2000; riloffand thelen, 2000; wang et al, 2000).

4: based on these technologies, riloff and thelen (2000) improved the **humsent accuracy to 40%** by applying a set of heuristic rules that assign **handcrafted weights to matching words** and ne.

5: it is interesting to note that the words automatically determined by our procedure are also part of those **words found manually** in the prior work of (1:riloff and thelen, 2000).

Fact id pyramid tier allocation: {'1': 3, '3': 3, '2': 2, '5': 3, '4': 2, '7': 1, '6': 2}

Optimal summary total facts weights: 16

Optimal summary covered facts: {'1': ['ading comprehension'], '3': ['manually generated rules', 'handcrafted weights to matching words'], '2': ['quarc'], '5': ['accuracy to 40%'], '4': ['humsent'], '7': ['"happen", "take place" "this", "story"'], '6': ['dateline']}

Summary total facts weights: 15

Summary covered facts: {'1': ['ading comprehension'], '3': ['handcrafted weights to matching words', 'words found manually'], '2': ['quarc'], '5': ['accuracy to 40%'], '4': ['humsent'], '6': ['dateline']}

Summary pyramid score: 0.9375

P02-1006

5 sentences summary:

1: (ravichandran and hovy 2002) also use **bootstrapping**, and **learn** simple **surface patterns** for extracting **binary relations** from the **web**.

2: ravichandran and hovy (2002) present an **alternative ontology** for type preference and describe a method for using this **alternative ontology** to extract particular answers using **surface text patterns**.

3: this method was first described in ravichandran and hovy (2002).

4: for instance, ravichandran and hovy (2002) report the following patterns for the relationships inventor, discoverer and location: relation prec.

5: in order to **train a rote extractor** from the **web**, this procedure is usually followed (ravichandran and hovy, 2002).

Fact id pyramid tier allocation: {'1': 21, '3': 12, '2': 15, '5': 1, '4': 3, '7': 3, '6': 1}

Optimal summary total facts weights: 56

Optimal summary covered facts: {'1': ['learn', 'bootstrap', 'learn'], '3': ['surface pattern', 'surface pattern', 'surface text pattern'], '2': ['web'], '5': ['rules based'], '4': ['binary relations', 'binary semantic relations'], '7': ['wildcard'], '6': ['alternative ontology']}

Summary total facts weights: 52

Summary covered facts: {'1': ['learn', 'bootstrap', 'train a rote extractor'], '3': ['surface pattern', 'surface text pattern'], '2': ['web', 'web'], '4': ['binary relations'], '6': ['alternative ontology']}

Summary pyramid score: 0.928571428571

D03-1017

5 sentences summary:

1: so far research in automatic opinion recognition has primarily addressed learning **subjective** language (wiebe et al, 2004; riloff et al, 2003; riloff and wiebe, 2003), **identifying opinionated documents** (yu and hatzivassiloglou, 2003) and sentences (yu and hatzivassiloglou, 2003; riloff et al, 2003; riloff and wiebe, 2003), and **discriminating between positive and negative language** (yu and hatzivassiloglou, 2003; turney and littman, 2003; pang et al, 2002; dave et al, 2003; nasukawa and yi, 2003; morinaga et al, 2002).

2: recent computational work either focuses on sentence ‘subjectivity’ (wiebe et al 2002; riloff et al 2003), concentrates just on explicit statements of evaluation, such as of films (turney 2002; pang et al 2002), or focuses on just one aspect of opinion, e.g., (hatzivassiloglou and mckeown 1997) on adjectives.

3: there is a large body of work on classifying the **polarity** of a document (e.g., pang et al (2002), turney (2002)), **a sentence** (e.g., liu et al (2003), yu and hatzivassiloglou (2003), kim and hovy (2004), gamon et al (2005)), a phrase (e.g., wilson et al (2005)), and a specific object (such as a product) mentioned in a document (e.g., morinaga et al (2002), yi et al (2003), popescu and etzioni (2005)).

4: dave et al (2003), riloff and wiebe (2003), bethard et al (2004), pang and lee (2004), wilson et al (2004), yu and hatzivassiloglou (2003), wiebe and riloff (2005)).

5: this amounts to performing binary text categorization under categories **objective and subjective** (pang and lee, 2004; yu and hatzivassiloglou, 2003); 2. determining document orientation (or **polarity**), as in deciding if a given **subjective** text expresses a positive or a negative opinion on its subject matter (pang and lee, 2004; turney, 2002); 3. determining the strength of document orientation, as in deciding e.g.

Fact id pyramid tier allocation: {'1': 1, '3': 1, '2': 14, '5': 10, '4': 7, '7': 3, '6': 14}

Optimal summary total facts weights: 50

Optimal summary covered facts: {'1': ['compared an individual-preference classifier'], '3': ['lexical cues'], '2': ['polarity', 'polarity', 'the polarity of opinion'], '5': ['opinion sentences'], '4': ['document-level', 'document-level subjectivity classification'], '7': ['semantically oriented words'], '6': ['subjective', 'subjectivity classification']}

Summary total facts weights: 45

Summary covered facts: {'2': ['discriminating between positive and negative language', 'polarity', 'polarity'], '5': ['a sentence'], '4': ['identifying opinionated documents'], '6': ['subjective', 'objective and subjective', 'subjective']}

Summary pyramid score: 0.9

P03-1001

5 sentences summary:

1: data fleischman et al (2003) describe a dataset of **concept-instance pairs** extracted automatically from a very large corpus of newspaper articles.

2: another direction we are pursuing is the use of machine learning techniques to **learn predictors of good nuggets**, much like the work of fleischman et al (2003).

3: 3.2 questions and corpus to get a clear picture of the impact of using different information extraction methods for the offline construction of knowledge bases, similarly to (fleischman et al, 2003), we focused only on questions about persons, taken from the trec8 through trec 2003 question sets.

4: in particular, we use the **name/instance lists** described by (fleischman et al, 2003) and available on fleischman's web page to generate features between names and nominals (this list contains a110a111a85 pairs mined from a112a73a96 gbs of news data).

5: at the same time, research efforts in data acquisition promise to deliver increasingly larger question-answer datasets (girju et al, 2003; fleischman et al, 2003).

Fact id pyramid tier allocation: {'1': 2, '3': 3, '2': 1, '5': 4, '4': 5, '6': 3}

Optimal summary total facts weights: 18

Optimal summary covered facts: {'1': ['filtering methods'], '3': ['instance/concept relations'], '2': ['lexico-semantic information'], '5': ['part of speech patterns'], '4': ['learning components'], '6': ['hyponym relations']}

Summary total facts weights: 8

Summary covered facts: {'3': ['concept-instance pairs', 'name/instance lists'], '4': ['learn predictors of good nuggets']}

Summary pyramid score: 0.444444444444

D04-9907

5 sentences summary:

1: the **tease algorithm** (szpektor et al, 2004) is an **unsupervised method** for **acquiring entailment relations from the web** for a given input template.

2: we then incorporate **paraphrase** similarity within the lexical similarity model by allowing, for some unaligned node $h \hat{=} ph$, where $t \hat{=} pt$: $\text{sim}(h,t) = \max(\text{mn}(h,t), \text{score}(ph,pt))$ 38 our approach to **paraphrase** detection is most similar to the tease algorithm (szpektor et al, 2004), and bears similarity to both dirt (lin and pantel, 2001) and knowitall (etzioni et al, 2004).

3: szpektor et al (2004) **measured yield**, the number of correct rules learned for an input re|see the 3rd iwp workshop for a sample of recent works on paraphrasing (<http://nlp.nagaokaut.ac.jp/iwp2005/>).

4: these retrieved text fragments are then considered good candidate for paraphrasing x bought y . anchor-based learning methods have been used to investigate many semantic relations ranging from very general ones as the isa relation in (morin, 1999) to very specific ones as in (ravichandran and hovy, 2002) where **paraphrases of question-answer pairs are searched in the web** or as in (szpektor et al, 2004) where a method to scan the web for searching textual entailment prototype relations is presented.

5: such transformations are typically denoted as **paraphrases** in the literature, where a wealth of methods for their automatic acquisition were proposed (lin and pantel, 2001; shinyama et al, 2002; barzilay and lee, 2003; szpektor et al, 2004).

Fact id pyramid tier allocation: {'1': 1, '3': 4, '2': 5, '5': 2, '4': 3, '7': 1, '6': 5, '8': 1}

Optimal summary total facts weights: 22

Optimal summary covered facts: {'1': ['learning entailment'], '3': ['paraphrase'], '2': ['acquiring entailment relations'], '5': ['unsupervised method'], '4': ['tease algorithm'], '7': ['anchors'], '6': ['entailment relations from the web', 'entailment relations from the web', 'in web corpus data'], '8': ['measured yield']}

Summary total facts weights: 20

Summary covered facts: {'3': ['paraphrase', 'paraphrase', 'paraphrase'], '2': ['acquiring entailment relations'], '5': ['unsupervised method'], '4': ['tease algorithm'], '6': ['entailment relations from the web', 'question-answer pairs are searched in the web'], '8': ['measured yield']}

Summary pyramid score: 0.909090909091

H05-1047

5 sentences summary:

1: these axioms express knowledge that could not be derived from wordnet regarding employment⁹, family relations, awards, etc. 5 semantic calculus the **semantic calculus axioms** combine two semantic relations identified within a text fragment and increase the semantic connectivity of the text (tatu and moldovan, 2005).

2: many previous approaches have used a logical form representation of the text and hypothesis sentences, focusing on deriving a proof by which one can infer the hypothesis logical form from the **text logical form** (bayer et al, 2005; bos and markert, 2005; raina et al, 2005; tatu and moldovan, 2005).

3: our overall test set **accuracy of 62.50% represents a 2.1% absolute improvement** over the task-independent system described in (tatu and moldovan, 2005), and a 20.2% relative improvement in accuracy over their system with respect to an uninformed baseline accuracy of 50%.

4: attempts have been made to remedy this deficit through various techniques, including modelbuilding (bos and markert, 2005) and the addition of **semantic axioms** (tatu and moldovan, 2005).

5: in id 152, we would like the hypothesis to align with the first part of the text, to 1this is the same problem labeled and addressed as context in tatu and moldovan (2005).

Fact id pyramid tier allocation: {'1': 2, '3': 1, '2': 1, '4': 2}

Optimal summary total facts weights: 6

Optimal summary covered facts: {'1': ['semantic calculus axioms'], '3': ['text logical form'], '2': ['world knowledge'], '4': ['60.4% accuracy', 'accuracy of 62.50% represents a 2.1% absolute improvement']}

Summary total facts weights: 5

Summary covered facts: {'1': ['semantic calculus axioms', 'semantic axioms'], '3': ['text logical form'], '4': ['accuracy of 62.50% represents a 2.1% absolute improvement']}

Summary pyramid score: 0.833333333333

H05-1079

5 sentences summary:

1: finally, a few efforts (akhmatova, 2005; fowler et al, 2005; bos and markert, 2005) have tried to 42 translate sentences into formulas of first-order logic, in order to **test logical entailment** with a theorem prover.

2: attempts have been made to remedy this deficit through various techniques, including **modelbuilding** (bos and markert, 2005) and the addition of semantic axioms (tatu and moldovan, 2005).

3: for example, two high-accuracy systems are those described in (tatu and moldovan, 2005), achieving **60.4% accuracy** with no task-specific information, and (bos and markert, 2005), which achieves 61.2% task-dependent accuracy, i.e.

4: many previous approaches have used a **logical form representation** of the text and hypothesis sentences, focusing on deriving a proof by which one can infer the hypothesis logical form from the text logical form (bayer et al, 2005; bos and markert, 2005; raina et al, 2005; tatu and moldovan, 2005).

5: the rte problem as presented in the pascal rte dataset is particularly attractive in that it is a reasonably simple task for human annotators with **high inter-annotator agreement** (95.1% in one independent labeling (bos and markert, 2005)), but an extremely challenging task for automated systems.

Fact id pyramid tier allocation: {'1': 1, '3': 4, '2': 2, '5': 1, '4': 1}

Optimal summary total facts weights: 9

Optimal summary covered facts: {'1': ['lexical based word overlap measures'], '3': ['logical form representation'], '2': ['modelbuilding'], '5': ['60.4% accuracy'], '4': ['high inter-annotator agreement']}

Summary total facts weights: 8

Summary covered facts: {'3': ['test logical entailment', 'logical form representation'], '2': ['modelbuilding'], '5': ['60.4% accuracy'], '4': ['high inter-annotator agreement']}

Summary pyramid score: 0.888888888889

W05-1203

5 sentences summary:

1: table 1: experimental results **lexical similarity** $\text{siml}(t,h)$ as defined in (corley and mihalcea, 2005).

2: in line with many other researches (e.g., (corley and mihalcea, 2005)), we determine these anchors using **different similarity** or relatedness detectors: the exact matching between tokens or lemmas, a similarity between tokens based on their edit distance, the derivationally related form relation and the verb entailment relation in wordnet, and, finally, a wordnet-based similarity (jiang and conrath, 1997).

3: first, as observed in (corley and mihalcea, 2005) the **lexical-based distance kernel** kl shows an **accuracy significantly higher** than the random baseline, i.e.

4: in line with (corley and mihalcea, 2005), we define it as: $s_l(t,h) = \sum_{w \in w} \text{sim}_w(w,t) \times \text{idf}(w)$ where $\text{idf}(w)$ is the inverse document frequency of the word w .

5: although these implications are uncontroversial, their automatic recognition is complex if we rely on models based on **lexical distance** (or similarity) between hypothesis and text, e.g., (corley and mihalcea, 2005).

Fact id pyramid tier allocation: {'1': 5, '3': 3, '2': 6}

Optimal summary total facts weights: 14

Optimal summary covered facts: {'1': ['lexical based word overlap', 'lexical distance'], '3': ['dramatic improvement'], '2': ['different similarity score functions', 'different similarity', 'of the j-c similarity', 'different relation between words']}

Summary total facts weights: 14

Summary covered facts: {'1': ['lexical similarity', 'lexical-based distance kernel', 'lexical distance'], '3': ['accuracy significantly higher'], '2': ['different similarity']}

Summary pyramid score: 1.0

P05-1014

5 sentences summary:

1: a similar idea by geffet and dagan (geffet and dagan, 2005) was proposed for capturing lexical entailment.

2: previous attempts have used, for instance, the similarities between case frames (lin and pan57 tel, 2001), anchor words (barzilay and lee, 2003; shinyama et al, 2002; szepektor et al, 2004), and a **web-based method** (szepektor et al, 2004;geffet and dagan, 2005).

3: for example, two verbs will be considered similar if they have large common sets of modifying subjects, objects, adverbs etc. **distributional similarity** does not capture directly meaning equivalence and entailment but rather a looser notion of meaning similarity (geffet and dagan, 2005).

4: recent attention to knowledge-rich problems such as question answering (pasca and harabagiu 2001) and textual entailment (geffet and dagan 2005) has encouraged natural language processing researchers to develop algorithms for automatically harvesting shallow semantic resources.

5: the method for **noun entailment** acquisition by (geffet and dagan, 2005) is based on the idea of **distributional inclusion**, according to which one noun is entailed by the other if the set of **occurrence contexts** of the former subsumes that of the latter.

Fact id pyramid tier allocation: {'1': 4, '3': 2, '2': 2, '4': 1}

Optimal summary total facts weights: 9

Optimal summary covered facts: {'1': ['distributional similarity', 'distributional similarity', 'distributional inclusion', 'distributional similarity'], '3': ['occurrence contexts'], '2': ['web-based method', 'over the web'], '4': ['noun entailment']}

Summary total facts weights: 9

Summary covered facts: {'1': ['distributional similarity', 'distributional inclusion'], '3': ['occurrence contexts'], '2': ['web-based method'], '4': ['noun entailment']}

Summary pyramid score: 1.0