Rangkuman

Towards Robust Indonesian Speech Recognition with

Spontaneous-Speech Adapted Acoustic Models

Paper ini menjelaskan mengenai penerapan spontaneous speech recognition menggunakan speech corpus Bahasa Indonesia. Beberapa kesulitan yang muncul ketika menggunakan spontaneous speech recognition dikarenakan spontaneous speech mengandung penundaan kata, penghubung kata, kesalahan penyebutan kata, dan urutan kata yang berubah-rubah.

Spontaneous speech recognition menggunakan speech corpus dengan banyak pengejaan masing-masing kata dengan variasi logat yang berbeda-beda. Adaptasi dari akustik model ini bertujuan untuk meningkatkan akurasi spontaneous speech recognition. Teknik yang digunakan untuk adaptasi akustik model ini berdasarkan Gaussian Mixture and Hidden Markov Models (GMM-HMM) yaitu

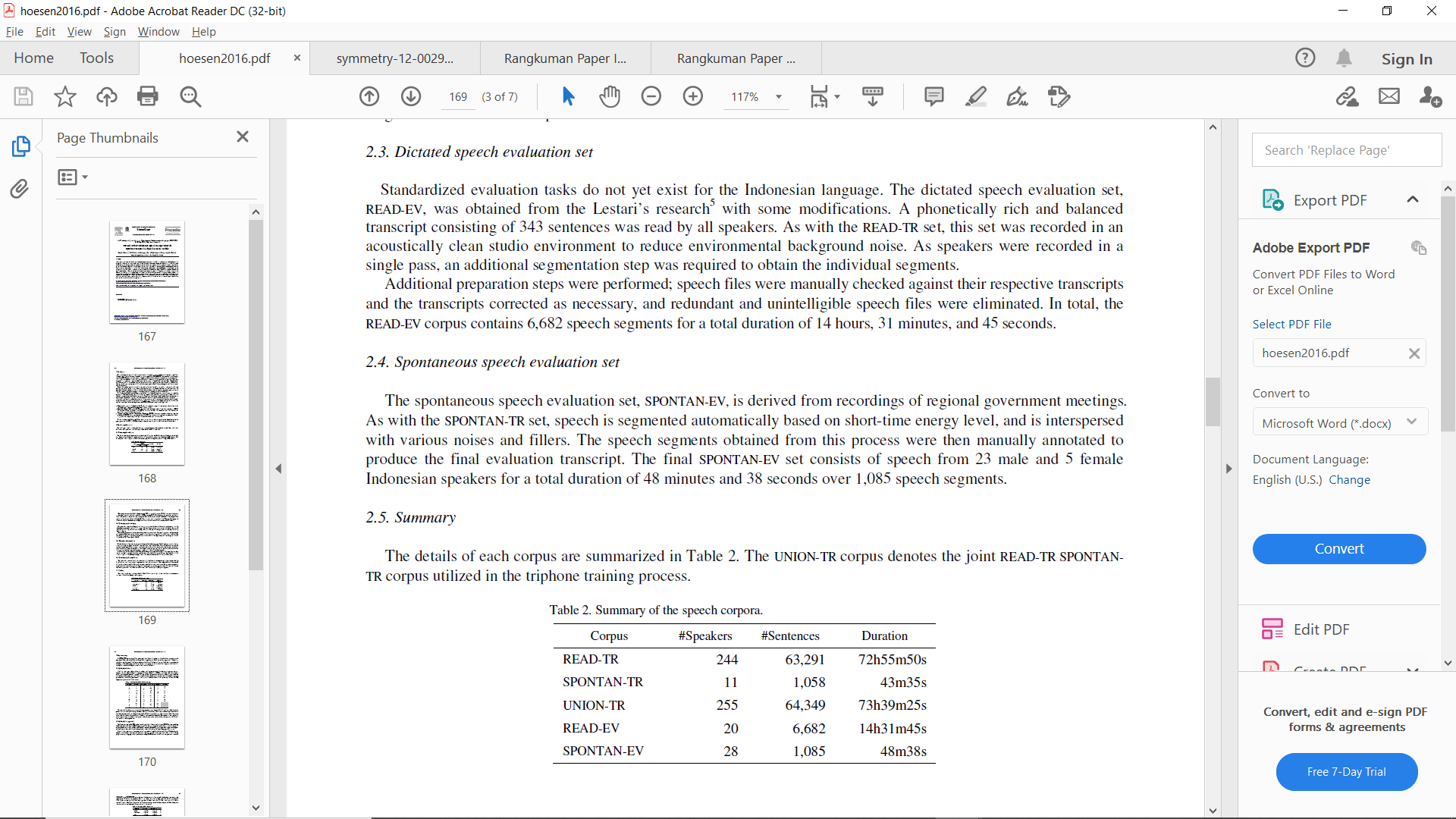
1. Maximum A-posteriori Probability (MAP)
2. Maximum Mutual Information (MMI)
3. Feature-space Maximum Likelihood Linear Regression (fMLLR)

MAP dan MMI sangat berperan dalam mengurangi word error rate untuk spontaneous speech recognition.

Speech Corpus yang digunakan pada implementasi machine learning terdiri dari :

1. Dictated speech training set
2. Spontaneous speech training set
3. Dictated speech evaluation set
4. Spontaneous speech evaluation set

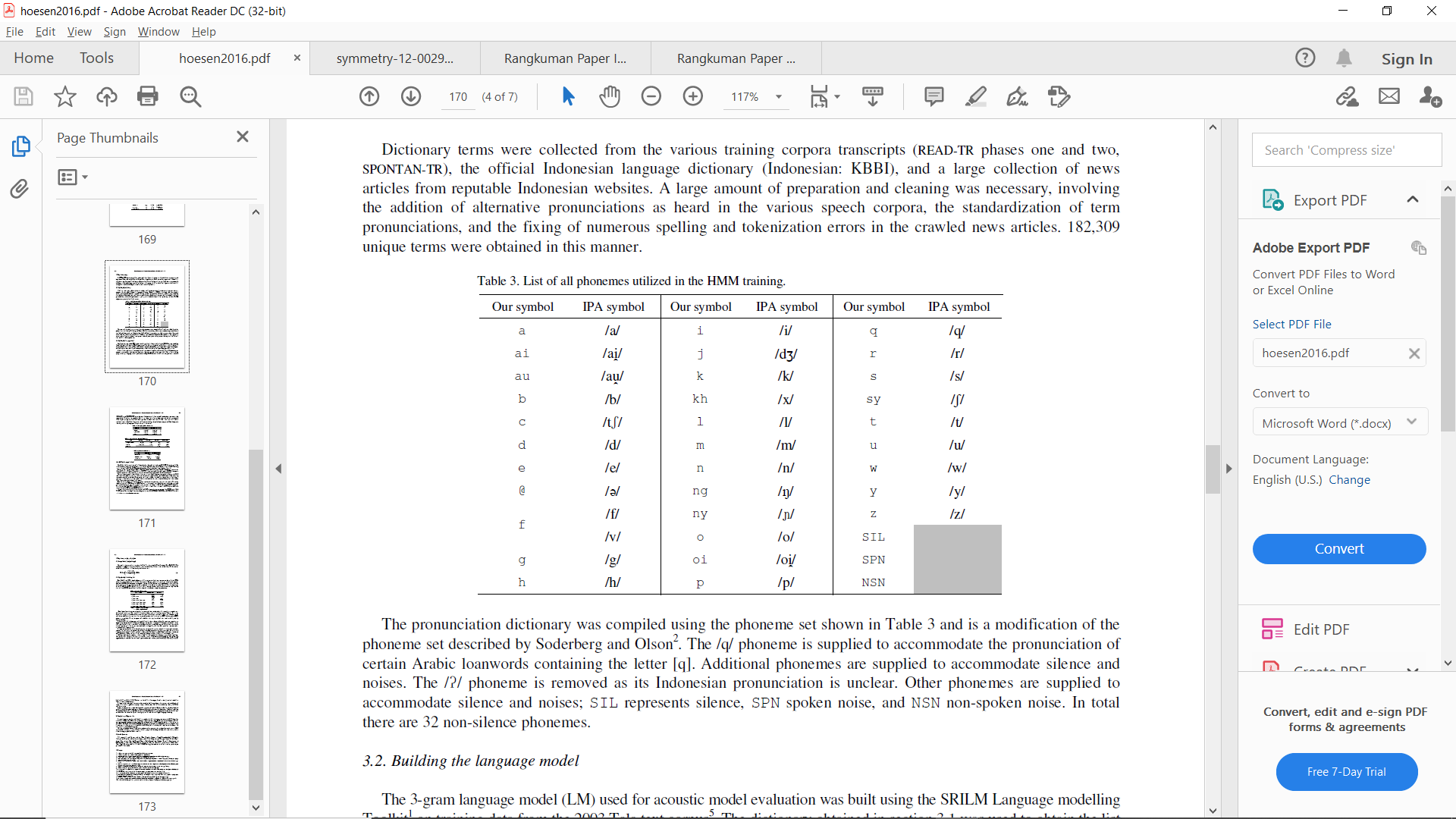
Dengan rangkuman data speech corpus sebagai berikut



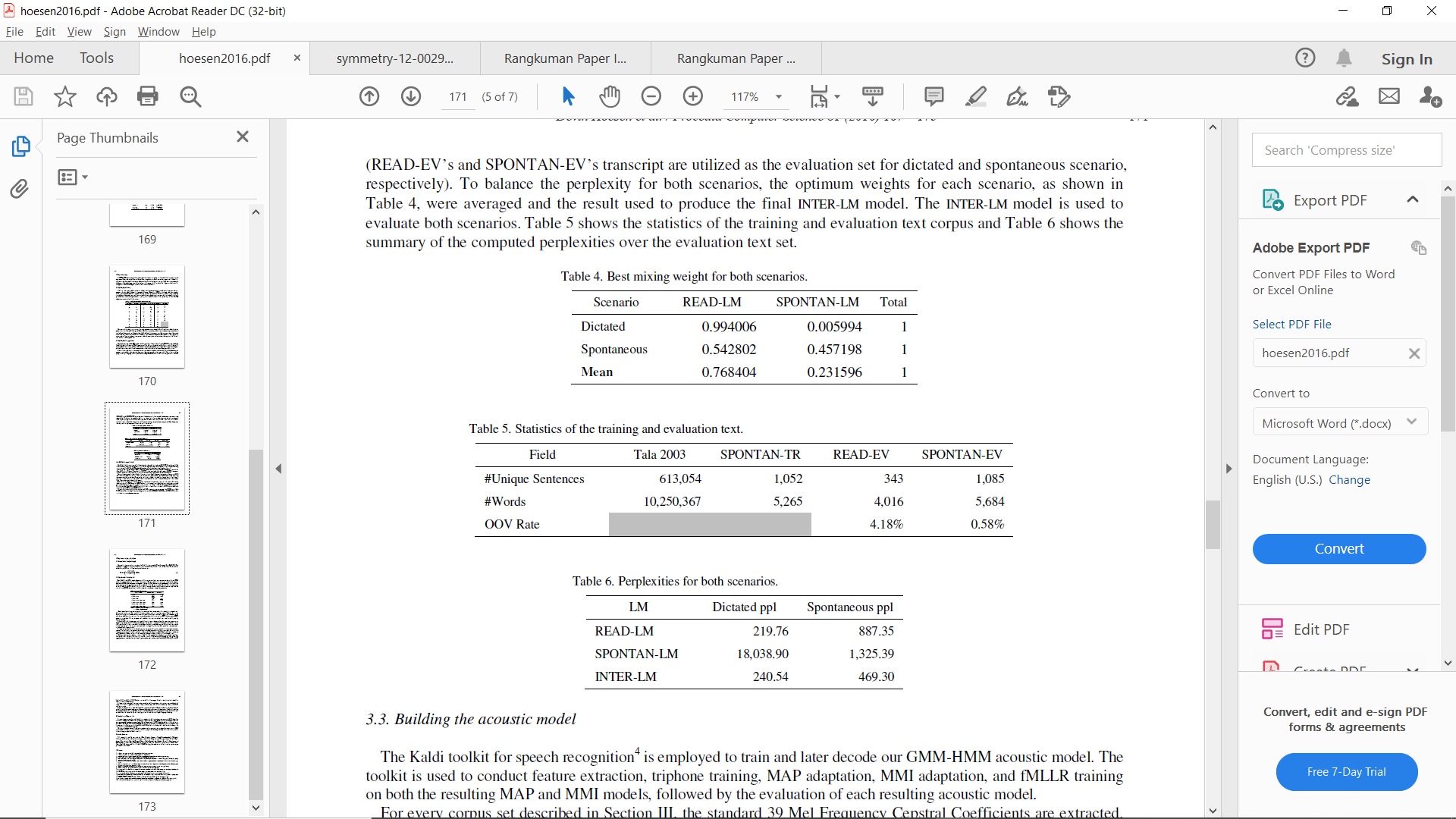
Langkah-langkah yang dilakukan sebelum melakukan implementasi spontaneous speech recognition

1. Building Dictionary

Pembuatan kamus bahasa dengan penyesuaian masalah pada spontaneous speech recognition seperti alternatif pengucapan, standarisasi pengucapan, memperbaiki kesalahan pengucapan,dll.



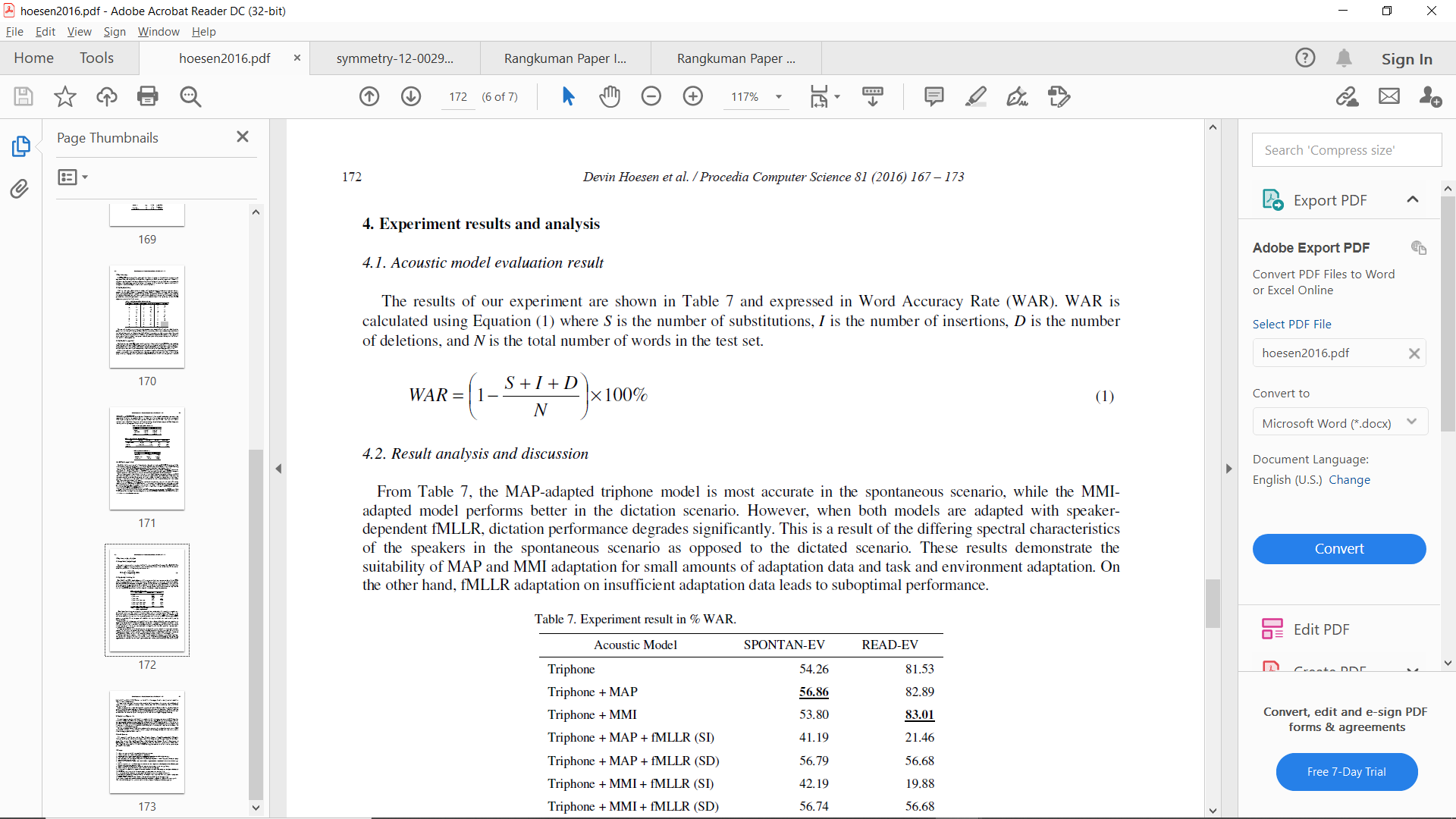
1. Building Language Model

Pembuatan Language Model menggunakan evaluasi akustik model dari pelatihan machine learning. Pelatihan machine learning menggunakan penggabungan data dari 2 speech corpus (dictated dan spontaneous ) dengan rasio tertentu.

1. Building Acoustic Model

Pembuatan akustik model berdasarkan Gaussian Mixture and Hidden Markov Models (GMM-HMM) dilakukan untuk feature extraction dengan MAP, MMI, fMMLR training.

Hasil implementasi spontaneous speech recognition ditampilkan dalam parameter Word Accuracy Rate (WAR).



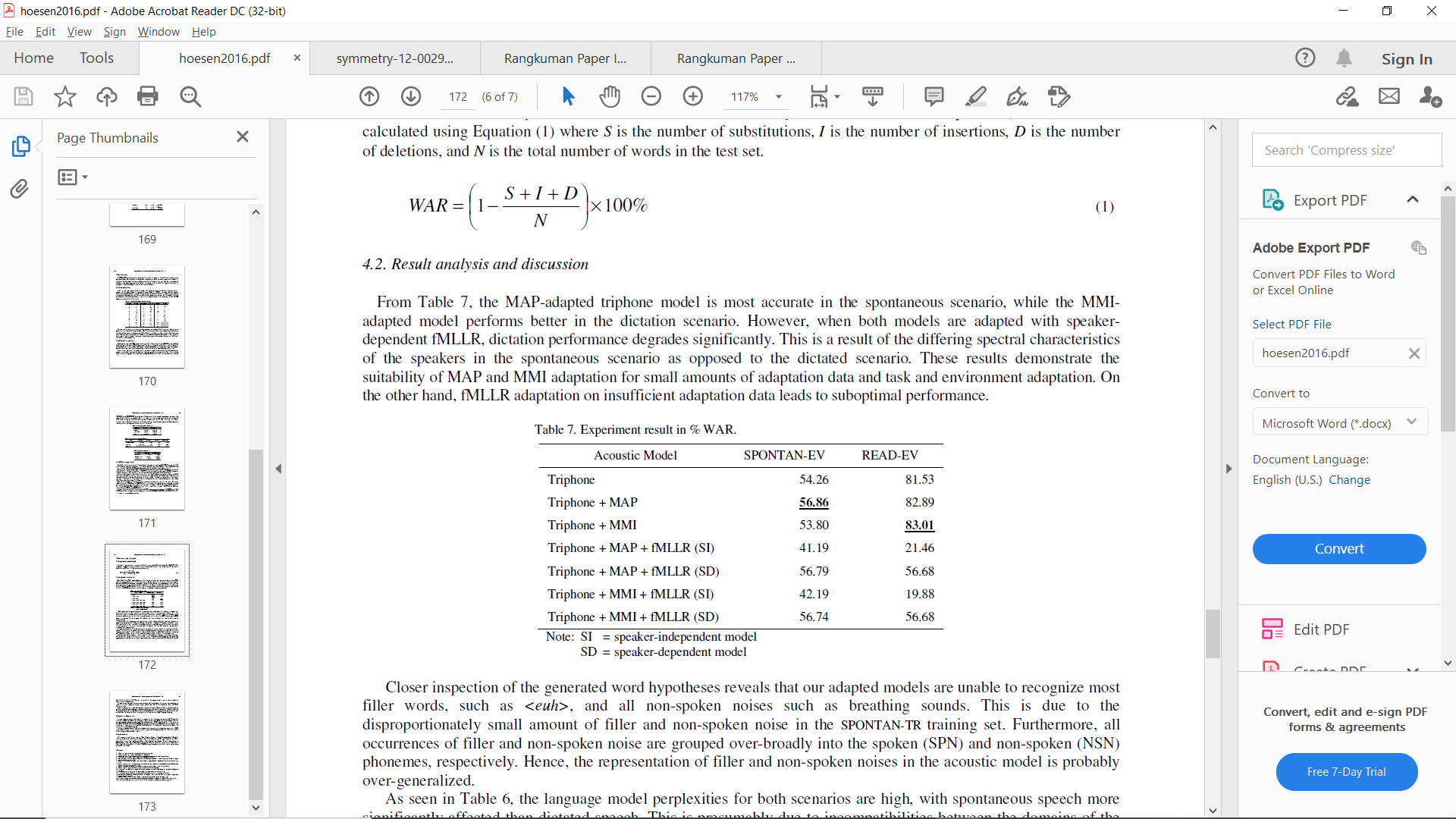
S = Substitution

I = Insertion

D = Deletion

N = Jumlah Kata pada test set

Data hasil experiment dapat dilihat pada table dibawah



Model triphone yang diadaptasi MAP paling akurat dalam skenario spontaneous, sedangkan model yang diadaptasi MMI berkinerja lebih baik dalam skenario dictated. Model akustik dibuat untuk mengenali ucapan dictated dan spontaneous. Adaptasi MAP menghasilkan peningkatan absolut 2,60% dan 1,36% dalam Word Accuracy Rate (WAR).