주간 활동 보고서-0516

산업보안학과 32200492 김명현

*주제: **향상된 번역 프로그램***

1. Test Plan

사전지식이 필요한 문장이나 긴 글과 일반적으로 진입장벽이 낮은 글의 번역결과에 대해 비교하고 분석해보면서 교정의 성능이 어느 정도인지 확인하여 "향상된 번역 프로그램이 유의미한지"에 대해 test해볼 예정이다. (영문을 한글로 번역해볼 예정)

예문 1: Chow et. al의 "White-Box Cryptography and an AES Implementation"라는 논문의 Abstract (공백 포함: 1010자, 공백 제외: 872자)

<Conventional software implementations of cryptographic algorithms are totally insecure where a hostile user may control the execution environment, or where co-located with malicious software. Yet current trends point to increasing usage in environments so threatened. We discuss encrypted-composed-function methods intended to provide a practical degree of protection against white-box (total access) attacks in untrusted execution environments. As an example, we show how aes can be implemented as a series of lookups in key-dependent tables. The intent is to hide the key by a combination of encoding its tables with random bijections representing compositions rather than individual steps, and extending the cryptographic boundary by pushing it out further into the containing application. We partially justify our aes implementation, and motivate its design, by showing how removal of parts of the recommended implementation allows specified attacks, including one utilizing a pattern in the aes SubBytes table.>

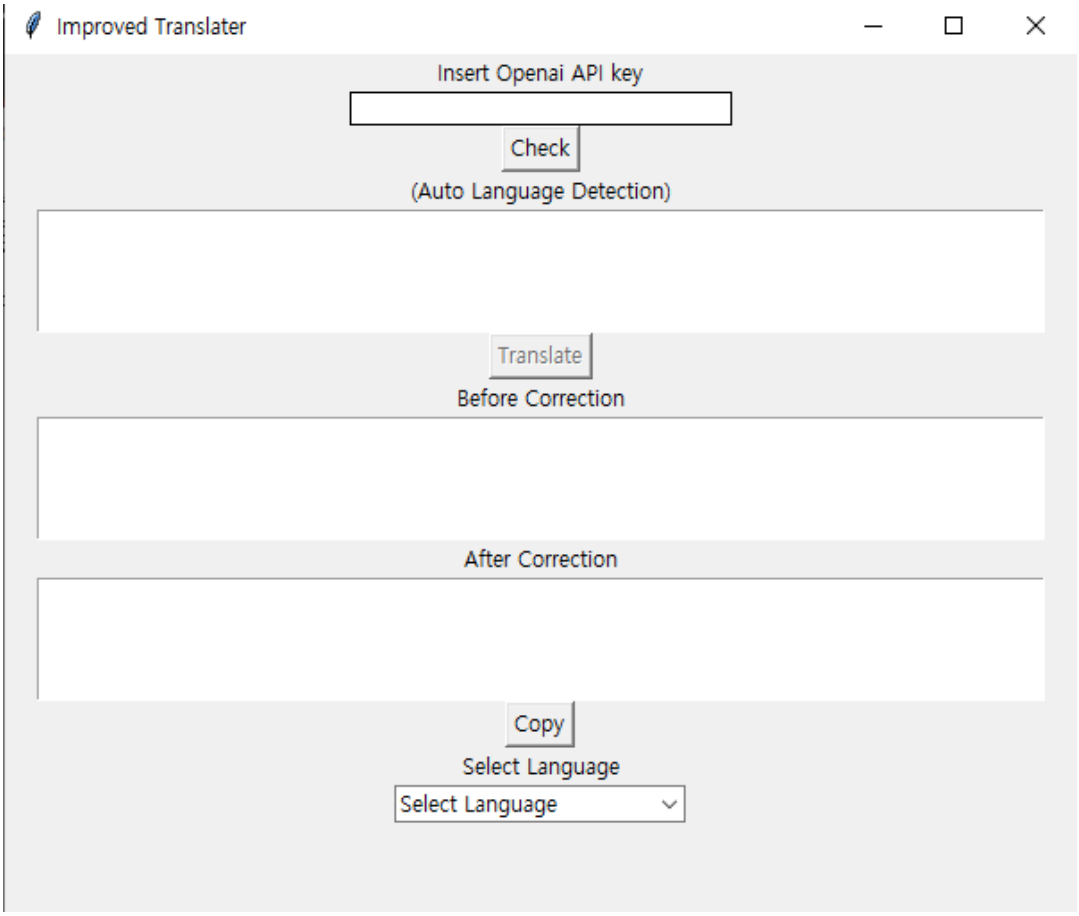
예문 2: 소설 '위대한 개츠비'의 첫 문단

<In my younger and more vulnerable years my father gave me some advice that I've been turning over in my mind ever since. "Whenever you feel like criticizing anyone," he told me, "just remember that all the people in this world haven't had the advantages that you've had.">

또한 지금까지 작동하는지에 대한 test를 했을 때 text-davinci-003 모델을 사용했었는

데, gpt-3.5-turbo 모델을 사용하면 1K 토큰당 0.002달러로 10배 저렴하기 때문에 chat 형식으로 태스크 변환을 하고 gpt-3.5-turbo 모델을 사용하는 test 또한 진행해 볼 예정이다.

마지막으로 test를 진행하는 동안 겪은 오류 과정이나 추가했으면 좋을 것 같다고 생각한 것들에 대해 기록할 것이며 내가 프로그램의 사용자가 되어 기존에 만든 데모 버전을 사용하면서 보완할 부분에 대해 찾아보는 test까지 진행해 볼 예정이다.



2. Test

먼저 예문 2의 번역결과에 대한 초벌 번역본과 교정본에 대한 비교를 해보았습니다.

■ *초벌 번역본:*

젊고 더 취약한 몇 년 동안 아버지는 나에게 내가 그 이후로 내 마음 속에서 뒤집어 놓은 조언을 주었다.'T는 당신이 가진 장점이있었습니다. "

■ *교정본:*

젊고 취약한 몇 년 동안, 아버지는 나에게 내 마음 속에서 뒤집어 놓은 조언을 주었습니다. "당신이 가진 장점을 기억하라"는 것이었습니다.

초벌 번역본은 사람들이 일반적으로 사용하지 않는 부자연스러운 문장 구성을 가지고 있고 등장인물의 대사를 제대로 번역하지 못했습니다. 교정본은 그나마 문장 구성이 자연스럽다. 또한 등장인물의 대사는 중간이 사라지고 다음 대사만 잘 번역이 되었다. 기존의 구글 번역기를 사용했을 때와는 초벌 번역본의 퀄리티가 떨어졌기 때문에 성능 향상 여부가 있는지 googletrans API에 대해서 좀 더 알아볼 생각이다.

다음은 사전지식이 필요한 긴 글에 대한 번역을 진행해보았다.

■ *초벌 번역본:*

암호화 알고리즘의 기존 소프트웨어 구현은 적대적인 사용자가 실행 환경을 제어 할 수 있거나 악의적 인 소프트웨어와 공동으로 배치 된 곳에서 완전히 불안하지 않습니다. 그러나 현재 추세는 환경에서 사용량이 증가하는 것을 지적합니다.우리는 신뢰할 수없는 실행 환경에서 화이트 박스 (Total Access) 공격에 대한 실질적인 보호 수준을 제공하기위한 암호화 된 작용 방법에 대해 논의합니다.예를 들어, AES가 키 의존 테이블에서 일련의 조회로 구현 될 수있는 방법을 보여줍니다.의도는 테이블을 개별 단계보다는 조성물을 나타내는 임의의 바이올리로 인코딩하고 포함 된 애플리케이션으로 더 밀어서 암호화 경계를 확장하여 키를 숨기는 것입니다.우리는 AES 구현을 부분적으로 정당화하고 권장 구현의 일부를 제거하면 AES 서브 베이츠 테이블의 패턴을 사용하는 것을 포함하여 지정된 공격을 허용하는 방법을 보여줌으로써 설계에 동기를 부여합니다.

■ *교정본:*

기존 암호화 알고리즘 소프트웨어 구현은 적대적인 사용자가 실행 환경을 제어하거나 악의적인 소프트웨어가 공동으로 배치된 곳에서 완벽하게 보호되지 않을 수 있습니다. 하지만 현재 추세는 사용량이 증가하는 환경에서 화이트박스 공격에 대한 실질적인 보호 수준을 제공하기 위해 암호화된 조치방법에 대해 논의하고 있습니다. 예를 들어, AES는 키 의존 테이블에서 일련의 조회로 구현될 수 있는 방법을 보여주고, 이를 통해 테이블을 개별 단계보다는 조각들로 인코딩하여 애플리케이션에 포함시켜 암호화 경계를 확장하고 키를 숨기는 것이 목적입니다. 우리는 AES 구현을 일부 정당화하고 권장 구현의 일

부를 제거하면 AES 서브 베이즈 테이블의 패턴을 사용하여 지정된 공격을 허용하는 방법을 보여주고, 이를 통해 설계에 동기를 부여합니다.

초벌 번역본 또한 문장 구성이 사람들이 기존에 쓰던 것과 다르게 부자연스럽다. 그런데 단어나 동사 또한 사전지식 없이 그대로 번역만을 한 느낌이 강하다. 그렇기 때문에 인공지능의 검색과 학습 능력을 활용하여 번역기에 사전 지식을 넣고자 했고 chatgpt API를 이용하여 교정을 해보았다. 교정본을 보면 확실히 문장이 읽는데 불편함이 없고 중간중간 어색했던 단어들을 눈치껏 잘 교정한 모습을 볼 수 있었다.

3. UI

UI를 다른 패키지를 사용하려 했는데 그냥 기존에 만들어 놓은 프로토타입 버전으로 사용하기로 했다. 여기에 기존에 틀만 만들어 놓은 API입력창에 기능을 추가하였고 예외처리나 기능 보완을 할 예정이다. 또한 채색까지 해보겠다. 그런데 한 가지 걸리는 게, 번역버튼을 누르면 응답없음으로 인터페이스가 반응을 안하고 렉걸린거 처럼 되다가 번역이 완료되면 번역글이 떠서 이를 매끄럽게 해결할 수 있는지에 대해 찾아보고 있다.