



학회가 열린 Fess Parker 호텔

김종권

서울대학교 프로그래밍 연구실

1. 소개

6월 13-17 동안 미국 산타바바라에서 열린 PLDI 2016에 다녀왔다. PLDI (Programming Language Design and Implementation) 는 프로그래밍 언어 디자인과 구현에 대한 학회이고 POPL과 함께 프로그래밍 언어 분야의 대표적인 학회이다. 이번 PLDI는 LCTES, ARRAY, FMS, SOAP 등의 여러 위성 학회, 워크샵들과 함께 일주일 동안 진행되었다. 나는 PLDI와 SOAP를 주로 들었고, 다른 몇몇 워크샵과 튜토리얼을 들었다. 또한 나는 PLDI Student Research Competition에 참가하여 포스터 발표를 진행하였다.

2. 여정

한국에서 미국, 미국에서 한국으로 올 때 모두 United Airlines 항공편을 이용하였다. 항공권 예매를 조금 늦게 한 탓인지 United Airlines 항공편만 구매할 수 있었다. 미국으로 갈 때에는 나리타 공항, 샌프란시스코 공항을 경유하여 산타바바라 공항에 도착하였다. 한국에 돌아올 때는 산타바바라 공항에서 샌프란시스코 공항을 경유하여 인천 공항으로 도착하였다. 10시간 이 넘는 비행은 너무나 지치고 힘들었다. 영화와 책 논문 등을 준비해 갔지만 좌석이 좁고 불편하여 거의 보지 못하였다.

샌프란시스코 공항에서 입국 심사를 하였는데 연구실 선배들의 Trip report 를 참고하여 학회 등록증 숙소 영수증 등 만반의 준비를 하였다. 그러나 막상 입국 심사에서 아무런 질문도 받지 않고 무사히(?) 통과 하였다. 같이 학회를 참석한 기홍이형과 성근이형은 간단한 질문을 받았다고 한다.



나리타 공항에서 샌프란시스코로 갈 때 탔던 UA 비행기

3. 날씨

산타바바라의 날씨는 우리나라의 초봄, 늦가을 건조한 날씨였다. 낮에는 햇빛은 따뜻하고 바람은 선선하여 밖에서 일광욕을 하고 싶어지는 날씨였다. 밤에는 조금 쌀쌀하여 반팔을 입으면 추운 날씨였다. 또한 학회장이 해안가 바로 앞에 위치하여 바람이 많이 불었다. 혹시 산타바바라 여행을 가게 된다면 가볍게 걸칠 수 있는 걸옷을 꼭 준비해가는 것이 좋을 것 같다.



구름 한 점 없는 날씨가 인상적이다.

4. 음식

산타바바라는 와인과 맥주가 유명하다고 한다. 와이너리 투어가 있을 정도로 와인이 유명하다고 하는데 아쉽게도 와인을 먹을 기회가 없었다. 지역 특산 맥주는 학회 리셉션과 만찬을 통해 맛볼 수 있었다.

학회장이 시내와 조금 멀리 떨어져 있어 주변에 마땅한 음식점이 없었다. 그러나 학회에서 아침에는 간단한 빵을 주었고, 점심으로 뷔페를 제공해주어 밖에 나가서 사먹을 일이 많지 않았다.



저녁으로 먹었던 음식들



오학주 교수님, 성근이형, 기홍이형과 마신 지역 특산 맥주

5. 주변 환경

학회장 바로 앞에 해변가가 있어 아침에 일찍 일어나 산책하는 것이 즐거웠다. 아침 날씨는 조금 쌀쌀하긴 했지만 시원한 바닷바람을 맞으며 산책하는 것이 즐거웠다. 숙소 바로 앞에 수영장이 있었지만 수영하기엔 날씨가 추웠다.

학회장 호텔에서 4일동안 지냈는데, 바로 앞에 기차길이 있어 시끄러웠다. 호텔 바로 앞에 기차길 건널목이 있어 기차가 지나갈 때 마다 경적을 울려 매우 시끄러웠다. 다행히 저녁 늦게는 기차가 다니지 않아 잠을 자는데 방해가 되지 않았다. 학회장 주변에 편의 시설이 많지 않아 사람이 거의 없어 상당히 여유로웠다.



산책하던 길과 모래사장

6. 포스터 발표

이번 PLDI에서 포스터 발표를 한 내용은 Format string bug 경보 분류 시스템인 Poclain에 대한 내용이었다. Poclain의 주요 내용인 사용자 상호작용 부분과 몇 가지 예제와 실험 결과를 주된 내용으로 발표하였다. 같은 내용을 발표하였음에도 모든 사람들이 다른 질문을 하여 상당히 재미있었다. 또한 대부분의 사람들이 많은 버그를 찾은 것에 대해 굉장히 흥미로워하였다.

포스터 발표는 PLDI 학회가 열린 수요일 오후 6시부터 약 3시간 정도 진행되었다. 국제 학회에서 영어로 발표하는 것은 처음이라 많이 긴장하였다. 그러나 막상 발표를 진행 하면서 사람들이 내 연구에 관심을 가져주고 많은 질문들을 해주어 긴장도 풀리고 즐겁게 발표할 수 있었다. 한가지 아쉬운 점은 영어실력이 부족해 대답을 하는데 시간이 좀 오래 걸리고 매끄럽게 대답하지 못하였다. 다행히 많은 사람들이 날카로운 질문들을 해주어 내 발표를 어느 정도 이해하고 있다는 사실을 알고 더 열심히 설명하였다.

많은 사람들과 소통하기 위해 영어 공부를 열심히 해야겠다는 생각이 들었다. 만약 영어를 더 잘 했다면 더 많은 사람들과 연구 이야기를 할 수 있었을 텐데 라는 아쉬움이 많이 남았다.

여 메모리 사용을 줄이고 분석 속도를 향상 시켰다. 다음으로 허위 경보를 줄이기 위해서 분석을 하다가 필요한 부분만 문맥 민감 분석을 하고 SMT Solver를 이용해 경로 민감 분석을 필요한 만큼 진행한다. 선별적으로 문맥 민감 분석을 하는 것과 Sparse하게 분석하는 방법은 우리 연구실에서 PLDI에 논문으로 냈었기 때문에 발표를 들으면서 이해를 잘 할 수 있었다. 이론적으로 새로운 내용은 없었지만 우리 연구실에서 연구한 내용들이 실제 산업 현장에서 핵심적으로 사용되고 있어 뿌듯하면서, 좋은 연구가 세상을 바꾼다는 것을 느낄 수 있었다.

[Tutorial] PROSE: Programming using Examples

이 발표는 Sumit Gulwani가 연구한 프로그램 합성에 대한 이야기이었다. 프로그램 합성이 무엇인지 잘 모르고 있었는데 예제 중심의 발표로 아주 쉽게 프로그램 합성에 대해 사전 지식 없이 어느 정도 쉽게 이해할 수 있었다.

Sumit Gulwani가 진행한 프로그램 합성 연구들의 공통적인 연구 동기는 프로그래밍을 하지 못하는 사용자들을 위한 것이다. 이 연구들의 결과물은 사용자가 만들고 싶은 프로그램에 대해서 입력과 출력을 예제로 주면 입력과 출력을 만족시키면서 사용자가 원하는 프로그램을 만들어내는 것이다.

프로그램 합성에서 가장 중요한 파트 중 하나는 도메인에 특화된 언어 (domain specific language)를 정의하는 것이다. 일반적으로 자동으로 프로그램을 만들기 어려운 이유는 문제공간이 매우 넓기 때문이다. 그러나 도메인에 특화된 언어를 정의함으로써 이 문제공간을 엄청나게 줄일 수 있다. 도메인에 특화된 언어를 정의할 때 중요한 부분은 입력과 출력에 대한 역함수가 존재하는지, 존재 한다면 합리적인 성능으로 구현 가능한지를 잘 선택해야 한다. Sumit Gulwani가 진행한 연구들은 모두 이런 도메인에 특화된 언어를 아주 정교하게 만들었다.

프로그램 합성의 다른 중요한 파트로는 합성된 프로그램이 여러 개 있을

때 적당한 프로그램을 고르는 기술이다. Sumit Gulwani가 푼 문제들은 사용자 편의를 위해 입력과 출력이 보통 한 개만 주어지는데, 보통 입력과 출력을 만족하는 프로그램이 엄청나게 많이 존재한다. 따라서 이 프로그램 중 사용자가 원하는 프로그램을 잘 골라내는 것이 프로그램 합성에서 굉장히 중요하다. 이런 문제를 해결하기 위해 굉장히 많은 예제들을 관찰하였고 데이터의 특징을 기계학습 하였다고 한다.

많은 사용자들이 실제로 프로그램 합성 연구의 결과물을 이용하고 있다는 점이 가장 인상 깊었다. Microsoft Excel에 들어있는 Flash Fill등 다양한 제품에서 프로그램 합성 기술을 이용하고 있다. 발표를 듣고 숙소로 돌아와 Excel에서 Flash Fill을 시험해봤는데 마치 내 생각을 읽고 있는 것처럼 잘 동작하였다.

Types from data: Making structured data first-class citizens in F#

본 연구는 외부 서비스와 통신할 때 쓰이는 XML, JSON 등과 같은 파일 양식에 대해 정적으로 타입 안전성 검사를 해주는 연구이다. 이번 PLDI 2016 발표 중에 가장 인상 깊게 들었던 발표이고, distinguished paper에 뽑힌 연구이기도 하다. 연구의 동기는 프로그램을 개발할 때 외부 서비스와 연동을 하기 위해서 JSON을 많이 사용하는데, 이 때 JSON에 대한 스키마가 제대로 주어지지 않는 경우가 많다. 또한 외부 서비스에서 가져온 JSON을 안전하게 처리하기 위해 상당히 귀찮은 작업을 거쳐야 한다. 이 부분에 대해서 상당히 공감을 했다. 실제 Facebook API나 Kakao API 날씨 API 등을 사용하기 위해 JSON으로 된 데이터를 받아 처리 하는데, 이때 주어진 스키마가 거의 없고 타입 때문에 직접 오류를 경험해봤기 때문에 발표 초반부터 상당히 기대를 하였다.

본 연구의 핵심은 컴파일 타임에 외부 서비스의 데이터를 예제로 받아 정적으로 타입 체크를 수행 하고 해당 데이터를 다룰 수 있는 모듈을 자동으로 생성해주는 것이다. 물론 예제를 기반으로 하는 타입 검사이기 때문에 완전(complete)할 수 는 없지만 현실에서 매우 유용할 것이라고 생각한

다. 실제로 본 연구의 구현이 F# 언어의 Standard Library로 포함되었다고 한다.

실제 발표에서는 타입 검사의 디테일 보다는 자신이 한 연구에 대해서 많은 예제들을 준비하여 발표하였다. 실제 서비스에서 자신의 연구 결과물이 어떻게 동작하는지를 중점적으로 보여주었다. 또한 어떤 부분이 부족한지를 예제로 보여주었다. 어떤 동기로 시작했는지 알기 쉬웠고, 연구의 목적이 명확하여 이해하기 쉬운 발표였다. 타입 검사의 자세한 내용은 논문을 통해 읽어볼 수 있기 때문에 이러한 발표 방식도 괜찮을 것 같다는 생각을 했다.

Statistical Similarity of Binaries

본 연구는 소스코드 접근이 어려운 환경에서 바이너리에 존재하는 취약점을 바이너리 코드 유사도를 통해 찾는 연구이다. 본 연구의 동기도 상당히 공감이 되었는데 일반적으로 프로그램이 많은 라이브러리에 의존하고 있고 라이브러리 소스코드를 직접 모두 관리하는 것은 현실적으로 어렵기 때문에 바이너리를 통해서 취약점을 찾는 아이디어는 상당히 괜찮은 것 같다고 생각한다.

본 연구의 핵심 아이디어는 코드를 쪼개서 얼마나 많은 부분이 비슷한가를 측정하는 것이다. 이 아이디어는 이미지 처리에서 사용하는 방식에서 아이디어를 얻었다고 한다. 이미지 유사도를 측정하기 위해 먼저 이미지를 특정 기준으로 조각낸다. 그리고 조각난 이미지들이 비교하고자 하는 원본 이미지에 얼마나 끼워 맞춰질 수 있는지를 파악한다. 많이 끼워 맞춰질 수록 유사도가 높다. 바이너리도 마찬가지로 특정 기준으로 조각내고, 원본 바이너리에 얼마나 끼워 맞춰지는지를 확인한다. 이러한 방식을 바이너리에 대해 적용하였더니 상당히 유용하다는 것이 이 연구의 핵심이다. 또한 중요한 부분은 유사도 함수를 구성할 때 우연히 한 두 명령어가 비슷할 때 보다 여러 명령어들이 비슷할 때 점수를 훨씬 많이 주는 방식을 사용하였다. 이러한 방식은 전체적인 코드의 흐름이 비슷한 코드를 찾는데 중

요한 역할을 했다고 한다.

본 연구에 실험은 각기 다른 컴파일러로 컴파일된 바이너리로 실험하였는데, 결과가 상당히 좋았다. 또한 재미있는 실험 결과 중 하나는 최적화 없이 컴파일한 바이너리와 최적화를 키고 컴파일한 바이너리는 유사도가 좋지 않았다. 그러나 각기 다른 컴파일러로 최적화를 키고 비교한 결과 O1, O2, O3에서 별 차이 없이 어느 정도 잘 찾았다고 한다.

Just-in-Time Static Type Checking for Dynamic Languages

본 연구는 동적 타입을 갖는 언어에서 런타임에서 필요할 때 타입 검사를 하는 방법에 대한 연구이다. 실제 구현에 사용한 언어는 루비인데 루비에서 정적으로 타입 검사를 하는 것이 어려운 이유는 바로 다단계 프로그래밍 때문이라고 한다. 본 연구는 함수 호출이 일어날 때 마다 해당 함수의 본체를 타입 검사한다. 이를 통해 다단계 프로그램으로 인해 발생하는 함수 호출을 안전하게 검사할 수 있다고 한다.

본 연구에서 핵심 아이디어는 타입 검사의 캐싱이다. 매 함수 호출마다 타입 검사를 하는 것은 매우 비싸다. 또한 흐름 민감 타입 검사를 사용하기 때문에 캐싱 없이 타입 검사를 하는 것은 실용적이지 못하다. 이는 실험 결과에서도 잘 들어나는데, 캐싱을 하기 전과 캐싱을 하고 난 뒤 성능이 약 3배 정도 차이가 났다. 본 연구의 핵심은 캐싱의 유효성 검사를 통해 필요할 때 마다 필요한 만큼만 타입 검사를 수행하는 점이다. 이러한 방법을 실제 루비에 대해 구현하여 오픈 소스 라이브러리에서 실제 타입 오류를 찾아내는 성과를 얻었다고 한다.

8. 맺으며

국제 학회에 올 수 있는 좋은 기회를 주신 이광근 교수님과 허충길 교수님께 감사 드립니다. 또한 학회 참석에 차질 없게 행정적으로 도와주신 최은희 선생님께도 감사 드립니다. 처음으로 국제 학회에 참석하고 발표도 하게 되어 좋은 경험이 되었습니다. 평생 기억에 남을 것 같습니다.



학회 만찬에서 PLDI'16에 오신 교수님들과 학생들 단체 사진