

SPLASH 2017 Trip Report

Vancouver, Canada

고려대학교 전민석



나의 첫 논문인 “Data-Driven Context Sensitivity for Points to Analysis”가 OOPSLA2017에 accept되어서, 2017.10.22-2017.10.28 동안 캐나다 밴쿠버에서 열린 SPLASH 2017에 참석하였다. SPLASH는 OOPSLA를 중심으로 하여, GPCE 등 여러 학회와 PLMW과 같은 워크샵이 함께 열리는 학술대회이다. 학회 기간 동안 경험했던 일들을 공유하고자 한다.

● Poster session

논문의 내용으로 포스터 세션에도 참가하였다. “Data Driven Program Analysis”가 무엇인지 정확히 전달하고자 하였다. 질문은 다양한 것들이 있었는데 그중 기억에 남는 것들은

1. 구한 formula에 대한 설명.
2. Feature에 대한 설명.

3. 성능이 굉장히 좋은 특별한 이유가 있는가.

4. Genetic Algorithm 을 써도 잘 될 꺼 같은가.

걱정을 많이 했는데, 나름 괜찮게 했던 것 같다. 질문들도 거의 비슷해서(위 4 개) 어렵지 않게 설명해주었다. 포스터를 준비할 때에는 사람들이 Detail 에 대해선 관심이 없을 줄 알았는데, 대부분이 궁금해 하고 Detail 에 대해 자세히 물어봤다.

대부분의 학생들은 “Data-Driven Approach” 를 잘 모르는 것 같았다. 다행히 손으로 Heuristic 을 만드는 것이 어렵다는 것은 모두 공감했었다. 그 부분이 공감이 되니 나머지가 자연스럽게 이야기가 되었다. 공통적으로 결과가 좋다고 칭찬을 많이 해줬다. 한가지 아쉬운 점은, 바로 옆에 Yanniss 가 있었는데 한마디도 못 했다는것.....

● PLMW

학생들 대상멘토들이 강의 해주는, PLMW(PL Mentoring Workshop)에 참석 하였다. 모든 강의를 집중해서 듣지는 못하였지만 인상깊게 들었던 강의 몇가지를 요약하면,

1. Research is a social process - Katheryn S McKinley.

연구를 할 때, Network 의 중요성을 강조를 하셨다. 처음에는 1 인 Network 로 시작해서 울타리를 넓혀갈 것, PLMW 에 참여한 학생들과 계속 만나게 될 것이니 PLMW 동안 많은 대화를 가질 것을 권장하였다. 3 분동안 주변 학생들과 자신의 연구분야에 대해 대화를 나누는 시간을 가지기도 하였다.

2. Why do a Ph.D. and how to pick an area? – Yaniss Smaragdakis

박사란 무엇인가, 박사과정이란 무엇인가에 대한 Talk 이었다. **박사과정의 기간은 중요치 않다는 것을 강조하셨다. 박사과정의 유일한 장점은 자신의 운명을 결정할 수 있는 상태라는 것이라 하셨다.** 연구분야를 결정하는 팁도 얘기를 했는데, 자신의 경우 처음엔 네트워크를 하셨었다고 한다. 수학을 좋아해서 진로를 바꾸셨다고 한다. 좋아하는 분야를 능동적으로 찾을 것을 강조하였다. 고작 1 년차지만 나는 나름 능동적으로 골랐다고 생각한다. 발표가 굉장히 인상적이었는데, STAR WARS 영화를 활용해 발표를 이끌어 갔다.

3. Choosing your mentor and learning to present your idea clearly – Ana Milanova

지도교수님은 한 명이지만, 여러 교수님들과 연구를 해볼 것을 권하였다.

우리 논문 “Data Driven Context Sensitivity for Points-to Analysis”를 읽어 보셨고 좋다고 하셨다. **Parameterized Object Sensitivity for Points to Analysis**를 2002년에 처음 제안 하셨는데(Invocation마다 다른 Depth를 주는 것이 가능하다) 15년동안 이에 대한 후속 연구가 없었다. 우리가 제대로 realize를 한 것 같아(0~k Depth를 모두 활용) 뿌듯했다.

우리 논문을 어떤 분이 리뷰 했는지는 정확히 알 수 없지만, OOPSLA가 시작하기도 전에 읽어 보셨다고 하시고, 좋다고 하셔서 우리 논문의 reviewer였음을 확신했다. 정말 좋다고 하셨던 것으로 보아 accept을 주신 분이라는 확신도 들었다.

사실 이번 PLMW 멘토들은, 내가 보고 공부했던 주요 논문들의 저자들이다. 실제로 보고 들을 수 있는 것은 특별한 경험이다.

● OOPSLA

인상깊은 발표들을 소개하면,

1. A Simple Soundness Proof for Dependent Object Types

DOT(dependent object type)이 type 분야에서 굉장히 hot한 주제인 듯 하다. 이번 OOPSLA에서 여러 발표들이 DOT에 대해 다루고 있다. DOT에 대해서 Soundness

증명이 굉장히 어렵다고 한다. 이 논문은 증명의 어려움을 문제점으로 지목하고, 쉽게 증명할 수 있는 framework(Proof Recipe)를 제시하였다. 증명 Framework 를 제시하는 것이 멋진 Contribution 이 될 수 있음을 느꼈다.

2. Effective Interactive Resolution of Static Analysis Alarms

이 논문은 유저와의 interaction 을 통해 false positive 를 줄이려 했던 논문이다. Xin Zhang 이 발표를 하였는데, sound analysis 와 unsound analysis 를 combine 한 approach 라고 설명을 하였다. 이 논문 역시 Datalog 의 Proof tree 를 이용하여 최대한 적게 물어보고 최대한 많은 false alarm 을 줄이는 방법(optimization problem)을 고안해냈다. Client 는 Data race detection 을 사용하였다. 약 74 퍼센트의 false 알람을 줄였다고 한다. Datalog 의 특징을 항상 잘 살려서 연구를 하시는 것 같다. 나도 Datalog 로 구현된 분석기(Doop)을 사용하고 있는데, 현재는 black box 처럼 사용하고있다. Proof tree 를 적극적으로 활용한 Data Driven Program Analysis 를 생각해보아야 할 것 같다.

3. Learning to Blame: Localizing Novice Type Errors with Data-Driven Diagnosis

이 논문은 Learning 을 통해 error localize 를 하려 했던 논문이다. 발표를 굉장히 잘 하셔서 기억에 남는다. 컴파일러를 modify 하여 유저와의 conversation 들을 기록해 놓고 이를 활용하여 학습을 하였다. 프로그램을 Semantic structure vector 로 표현하고 에러 코드와 첫번째 correct 코드를 가지고 학습을 하였다. 학습은 Supervised learning 을 하였고 대학원생들을 통해 User study 를 하였다고 한다. PRL 에서 할 수 있을법한 주제였기에 먼저 하지못한것에 대해 약간 아쉽기도 하였다. 요즘 hot 한 주제 인듯?(Learning 을 통한 error localize)

4. Heap Don't Lie: Countering Unsoundness with Heap Snapshots

이 논문은 동적 분석을 통해 정적 분석이 놓치는 Heap 들을 기록해 놓고 정적 분석이 이를 놓치지 않고 분석할 수 있도록 하여, Soundness 를 끌어올린 논문이다. Yanniss 연구실의 Neville 의 연구이다. 프로그램 실행을 통해 Static analysis 가 잡아내지 못하는 Heap 들을 detect 하는 것 뿐만 아니라 Context 정보 또한 계산해서 정적 분석기에게 넘겨준다. 내 연구가 precision 과 cost 의 균형을 맞춘 연구라고 한다면, 이 연구는 soundness 를 최대한 끌어올리면서 cost 증가량을 최대한 낮추고자 했던 연구. 실험 결과를 보니 기존의 Java 포인터 분석에서 많은 Heap 들을 놓치고 있다는 것을 알 수 있었다. 포인터분석을 하는 입장에서 중요한 이슈인 것 같다. 포인터 분석의 soundness 올리는 논문이 최근에 많이 보인다. (Tian tan 도 최근 soundness 높이는 연구를 한적이 있음)

5. P/Taint: Unified Points-to and Taint Analysis

Points to 분석과 Taint 분석이 비슷함을 이용한 논문이다. 기존의 포인터 분석룰에 아주 조금 룰을 더 추가해서 Points to 분석과 Taint 분석을 동시에 수행하는 룰을 만들었다. 놀라운 점은 두가지 분석을 동시에 하는데도 불구하고 분석 Cost 는 Points to 분석만 하는 것에 비해 크게 증가하지 않았다는 것. 두 종류의 분석을 Unify 하는 분석 룰을 제시한 논문이다. Contribution 이 굉장히 특이하다는 느낌을 받았다.

● Keynote-How the languages we speak shape the ways we think

Keynote 3 개가 있었는데 마지막 날의 keynote 가 가장 기억에 남아 내용을 공유하고자 한다. 어떤 언어를 사용하느냐에 따라 생각하는 방향이 달라진다고 한다. 한국어를 예시로 들었는데,

He ate poison.

위 문장에는 말하는 이가 어떻게 정보를 입수했는지는 나와있지 않다. 하지만 위 문장을 한국말로 하면 정보를 입수한 경위까지 말하게 된다고 한다.

-그 친구 독을 먹었다더라(전해 들음)

-그 친구가 독을 먹던데?(직접 봄)

반대로 영어는 독을 먹은 사람이 남자인지 여자인지를 자연스럽게 전달을 한다.(He She)

2 가지 이상의 언어를 구사하는 사람은, 2 가지 이상의 영혼을 가진 것과 같다고 한다. 그렇다면, 사용하는 programming language 에 따라서도 성격이 다를까? 설문 조사중이라고 한다.

● 후기

개인적으로 좋으면서 많이 아쉬웠던 학회였다. 좋았던 점은 지금까지 읽었던 모든 논문의 저자가 거의 다 모인 학회였다는 것이다.

Yanniss Smaragdakis, Ana Milanova, Ondrei Lhotak, Mayur Naik, Isill Dillig, Sumit Gulwani, ...

아쉬웠던 점은 커뮤니케이션 능력이 부족해서 위 사람들과는 거의 대화를 못해봤다는 것이다. 학회 기간 동안 외국 학생들이랑은 대화가 어느 정도 가능 했다(질문이 쉬움). 근데 저자들 또는 교수님들과는 대화가 거의 안되었다. 일단 알아듣기가 너무 힘들었는데 영어가 거의 랩처럼 들렸다. 결국 뭘 물어보는지 모르니 동문서답을 하게 되고, 서로 못 알아듣는 과정을 반복하다 보면 결국 서로의 논문을 보여주게 된다.

이번 연구가 성공적으로 끝날 수 있었던 가장 큰 이유는 연구 미팅이었다고 생각한다. 연구 시작부터 끝까지 모든 미팅이 의미 있고 즐거웠다. 언제나 생산적이고 즐거운

미팅으로 나의 연구를 이끌어 주셨던 오학주 교수님께 감사하다고 전해드리고 싶다. 서로의 부족함을 메꿔주며 연구했던 나의 research mate 세훈이 형에게도 가장 고맙다 전하고 싶다. Writing 에서 큰 도움 주신 차성덕 교수님께도 감사하다고 전해드리고 싶다.

한번 더 Context Sensitivity 를 주제로 연구를 할 것이다. 자신감도 생긴 김에 이번엔 더 높은 곳에 도전해보고자 한다.

이쯤에서 나의 첫번째 trip report 를 마무리 하려 한다.