2024 SIGPL 겨울학교 Trip Report

2024/01/31 ~ 2024/02/02 @ KAIST

안중원
2024/02/13

이번 겨울 두번째로 SIGPL에 다녀왔다. 이번에는 포스터 발표도 준비하는 등 첫번째와는 사뭇 의미가 달랐다. 이 글에서는 겨울학교를 다녀오며 기억에 남은 일들을 기록한다.

번개 발표와 포스터 발표

이번에 처음으로 SIGPL에 발표를 하게 되었다. 계속 교수님과 면담을 하면서 하기는 했지만 지금까지의 연구는 그냥 개인적인 호기심으로 한다는 기분이 없잖아 있었다. 포스터 발표에서 다른 연구실 사람들에게도 지금까지의 진행상황을 보여주고 토론하니 내가 학계에 발 담근 채 무언가 기여하고 있구나 하는 것을 느낄 수 있었다.

발표 주제는 지난 반년간 만든 리액트 핵언어의 의미구조였다. 지금까지 주로 의미구조를 예쁘게 정리하는 데에만 집중하다보니 언어의 기술적인 디테일을 있는 그대로 드러내기만 하는 발표가 되어 아쉬웠다. 작년에는 포스터 발표를 보면서도 그냥 '우와~'하고 감상만 했지만, 이번에 발표를 보면서는 연구한 내용을 포스터로 정리하여 발표하는 것도 허투로 볼 일이 아니라는 생각이 들었다. 다음에 발표할 때에는 — 당장 다음 쇼앤텔에서도 — 의미구조도 머릿속에 이미지가 바로바로 들어오게 정리하고 분석기 개발도 진행해서 이 연구를 개괄적으로 보여줄 수 있다면 좋겠다.

다른 포스터들을 돌아보는 것도 두번째 SIGPL에서는 저번과 의미가 달랐다. 저번에는 연구실에 막 들어와서는 최신 연구 동향을 포스터 수십개로 휘리릭 읽어들이는 꼴이다 보니 적혀있는 내용을 읽는 것부터가 버거웠다. 이번에 가보니 저번에 발표했던 연구 이후에 진행된 내용으로 발표하는 것들이 많았다. 아직 두번째지만 슬슬 맥락이 생기는 것 같았다.

다른 포스터들 사이에는 거대 언어 모델을 활용하는 연구가 많았다. 코드 리팩터링에 활용하거나, 프로그램 합성과 결합하는 식이다. 우리 연구실에는 언어 모델을 다루는 연구가 없어서 길건너 불구경하는 듯한 느낌이었지만 PL 관련 연구도 그 거대한 파도를 타고 있는 것을 볼 수 있었다. 좁게는 언어 모델, 넓게는 인공지능 쪽을 계속 주시하고 있어야겠다는 생각이 들었다.

그 외에는 카이스트 김태우 님이 발표한 구간 스킵 리스트가 기억에 남는다. 저번에 PL 학회에 자료구조가 나온 것이 특이하여 기억에 남았는데, 운영체제 과제로 고생을 하고나서 이번에 다시 보니 해결하고자 하는 문제가 이해가 되었다. mmap 할당 수정을 병렬적으로 수행할 수 있도록 리스트에 대해 원자적인 스왑이 가능하게 하는 자료구조였다. mmap 구현에서 전역으로 락을 잡도록 구현한 것이 아쉬움으로 남았는데 역시 쉽지 않은 문제였던 모양이다.

이번에는 포스터 발표 시간에 다양한 캔맥주가 제공되었다. 3일 내내 제공된 성심당 케이터링과 더불어 참 호화로운 겨울학교다. 술이 약해 반캔도 마시지 못한 게 아쉬울 따름이다. 그래도 맥주 덕분에(?) 더 적극적으로 토론에 참여할 수 있었던 것 같다.

초청강연

이번 겨울학교에도 유익한 초청강연이 많았다. 인상적이었던 내용과 강연을 듣고 들었던 생각들을 정리했다.

딥 전이학습 기반 프로그램 자동 수정: 산업 적용 사례를 중심으로 - 김미수 (전남대)

이 강연은 코틀린 코드의 결함 및 버그를 딥러닝으로 자동 수정하는 연구에 대한 것이었다. 코드 패치에는 크게 두가지 방법론, 템플릿 기반과 딥러닝이 있는데 그중 딥러닝을 선택한 것이다. 이 강연에서는 딥러닝 기반 연구의 속내를 들여다볼 수 있었던 점이 기억에 남는다. 제한된 데이터셋을 그대로 써서 기존 모델을 튜닝하려고 하면 한계가 있는데, 이 연구에서는 데이터셋의 분포를 기존 모델이 학습한 데이터셋과 맞추기 위해 코드 내의 이름 등을 조금씩 수정하여 추가로 넣어줬다고 한다. 야매인듯하면서도 그럴듯한 근거를 갖고 효과적인 튜닝을 해낸다는 게 미묘한 재미가 있다.

임베디드 시스템을 위한 컴파일러 디자인 - 허선영 (경희대)

이 강연은 여러 제약이 많은 임베디드 시스템에서 머신러닝 프로그램을 실행하기 위한 컴파일러를 설계하는 연구에 대한 것이었다. 이 컴파일러는 범용 래퍼 함수를 사용하는 대신 각 상황에 최적화된 커널 함수를 직접 호출하는 코드를 생성한다. 정적 메모리를 효과적으로 활용하기 위해 liveness analysis로 메모리 활용 계획을 세우고 하이퍼 파라미터를 상수 전파하는 등의 기법을 적용한다.

강연 후에 전통적인 최적화도 적용하는지에 대한 질문이 나왔는데, 이 컴파일러 설계의 주안점은 IR 수준 최적화이고, 저수준 코드에서도 전통적인 최적화는 그대로 진행된다고 한다. 저수준 코드에서는 손실되는 정보가 많기 때문에 IR에서 특수한 최적화가 이루어진다. 그 이야기를 들으니 소개원실 수업 프로젝트로 최적화 패스를 적용하던 때가 떠올랐다. LLVM 최적화 패스들을 보면 고수준 코드를 다 저수준으로 풀어헤친 IR에서 다시 루프 구조 등을 찾아내 최적화를 진행하는 경우가 있다. 그런 것들을 아예 중간수준 코드에서 최적화를 할 수 있다면 좋았겠다는 생각이 들었다. 러스트 언어에도 여러 단계의 IR이 사용되는 것을 보면 고수준 최적화도 효과적인 것 같다.

LeakPair: 단일페이지 웹 어플리케이션내의 메모리 누수를 고치기 위한 선제적 디버깅 - 김동선 (경북대)

웹 어플리케이션은 내가 잘 아는 분야라 열심히 들은 강연이다. 이 연구에서 다룬 메모리 누수는 대부분 이벤트 핸들러를 해제하지 않아 생기는 간단한 결함들이었고, 패치도 한두줄이면 되는 것들이었다. 흥미로운 부분은 오픈소스 프로젝트에 적용된 패치들이 실제 코드에 반영되기 위한 과정이었다. 오픈소스가 많은 웹 생태계에 특화된 방법론으로 선제적 디버깅을 내세운 것이 인상적이다. 이러한 메모리 누수는 보통은 잘 신경쓰지 않는 사소한 결함인데, 패치의 중요성을 역설하기 위해 패치를 먼저 적용하여 전후의 성능 변화를 수치로 보여주었다고 한다. 필자가 연구중인 리액트 검사기도 eslint 등 기존 툴에 잘 결합할 수 있도록 만들어야겠다는 생각이 들었다.

소프트 오류로부터의 신뢰성 향상을 위한 비트 단위 정적 분석 기법 - 고유선 (연세대)

소프트 오류란 하드웨어 결함으로 비트가 뒤집히는 오류이다. 하드웨어의 신뢰성은 정지 문제와는 또 다른 종류의 "불가능성"인 느낌이라 평소 재밌다고 생각하던 주제다. 이 강연에서는 소프트 오류를 모델링하고 분석하기 위한 기법을 소개했다. 소프트 오류가 프로그램에 실질적인 영향

을 끼치는 과정을 알 수 있어 재미있는 발표였다. 우리 연구실에서 프로그램 합성 싹수 분석에 사용하던 비트열 분석과도 겹치는 부분이 많아 복습(?)도 되었다.

코드 속성 그래프의 부분 그래프 매칭을 통한 대규모 취약점 탐지 - 위성일 (UNIST)

이 강연은 PHP 프로그램에서 그래프를 뽑아내 취약점을 탐지하는 연구에 관한 것이었다. 기존 연구에서도 그래프를 이용한 취약점 탐지 알고리즘을 고안했는데, 그건 너무 일반적이어서 일부 취약점을 놓치는 문제가 있었다. 그렇다고 각 취약점 유형에 대해 개별적으로 알고리즘을 만드는 것도 품이 너무 들어가는 일이다. 이 연구에서 고안한 것은 취약점 보고서 DB에서 각 항목을 요약하는 그래프를 생성하여 부분 그래프 매칭으로 취약점을 탐지하는 것이다.

이 강연에서는 프로그램을 그래프로 나타내 분석하는 것이 흥미로웠다. 생각해보면 CFG도 있고 하니 그렇게 특별한 것은 아닐지 모르지만 이 강연을 듣고 그래프로 생각하는 방식도 있다는 것을 새삼 깨달았다. 주로 코드의 모양을 AST로 시각화하곤 했었는데 의미구조도 그래프로 시각화하여 생각하면 유용한 아이디어들을 "볼" 수 있을 것 같다.

대전 여행

이번 여행에서는 오로지 공부만 하지 않았다. 조별활동 시간이 마련되어 다른 연구실 사람들과 잡담도 나누고 현지인이 추천하는 맛집에서 맛있는 태국 음식도 먹었다. 다른 연구실들은 어떻게 돌아가는지 엿볼 수 있었다.

우리 연구실 식구들과도 친해진 여행이었다. 저번에는 어쩌다보니 숙소가 다 1인실이라 밤은 각자의 시간이었지만, 이번에는 모두 2인실로 예약이 되었다. 아무래도 연구실에서는 연구에 관한 이야기를 많이 하게 되는데 여기에서 시시콜콜한 잡담을 많이 할 수 있었다. 둘째날 저녁에는 한 방에 모여 맥주를 마시며 떠들었다. 설입에 있는 맛집들을 많이 알게 됐다.

셋째날 아침에는 수강신청이 있어 혼자 일찍 PC방에 갔다. SIGPL 행사 시작까지는 시간이 약간 떠서 자전거를 타고 여유롭게 가기로 했다. 가면서 갑천 구경도 하고 카이스트 명물인 오리 호수도 구경했다. 이름은 오리 호수이지만 오리보다도 커다란 거위들이 모여있는 모습이 퍽 위압적이었다.

모든 행사가 끝난 뒤에는 칼국수를 먹고 대전의 최고 핫플레이스인 성심당에 갔다. 수많은 사람들이 줄을 서있고 매장 안에도 사람이 그득그득했다. 하지만 의외로 줄이 빨리 줄어들고 매장 내에서도 금방 빵을 골라 나올 수 있었다. 각자 다양한 빵을 샀는데 나는 그 유명한 튀김소보로와 부추빵, 그리고 케이터링에서 맛있게 먹었던 찹쌀버터빵을 골랐다. 명성만큼 맛있는 빵이었다.