SW 정적분석

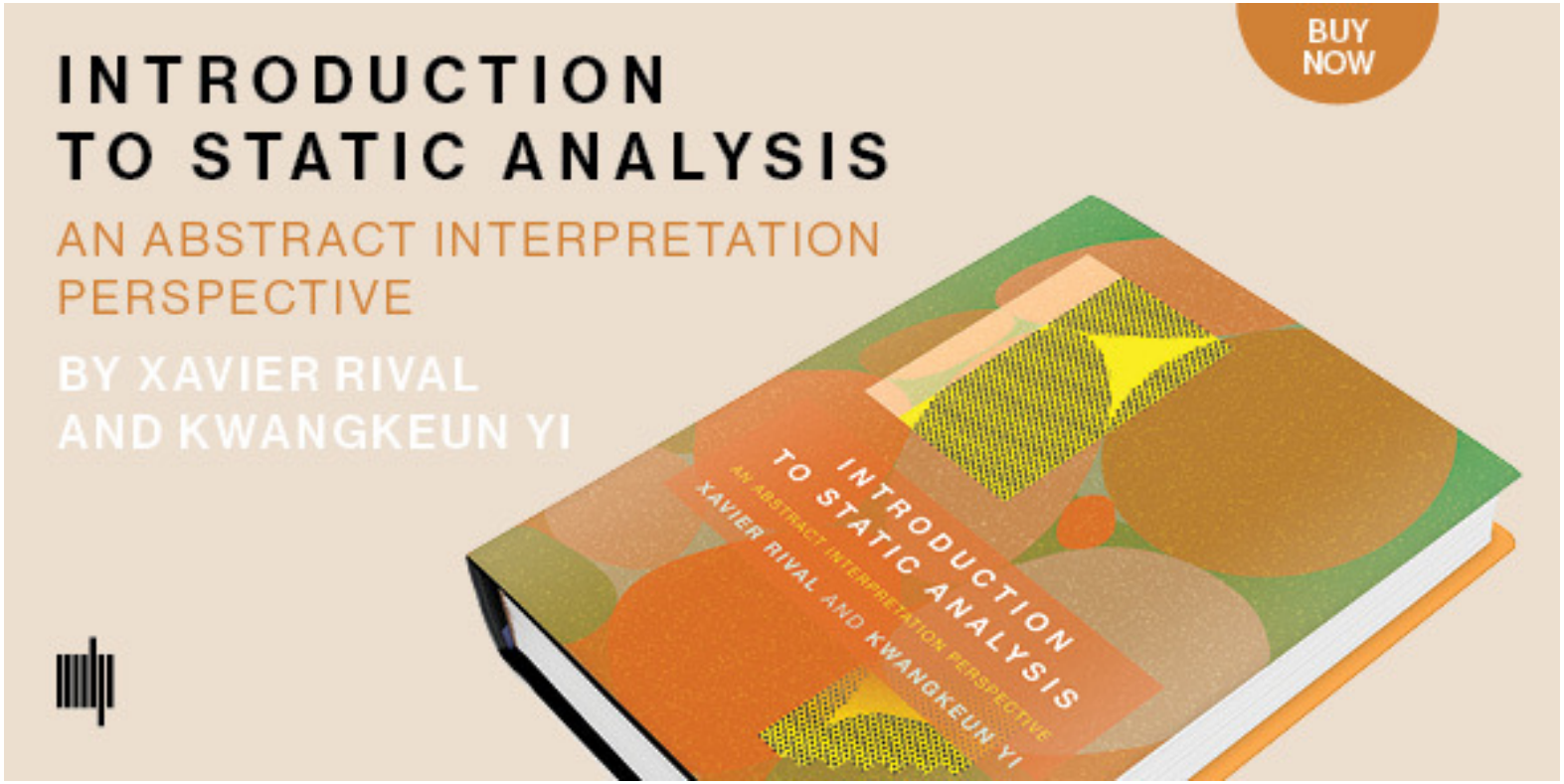
이광근

서울대학교 컴퓨터공학부

kwangkeunyi.snu.ac.kr

12/22/2020

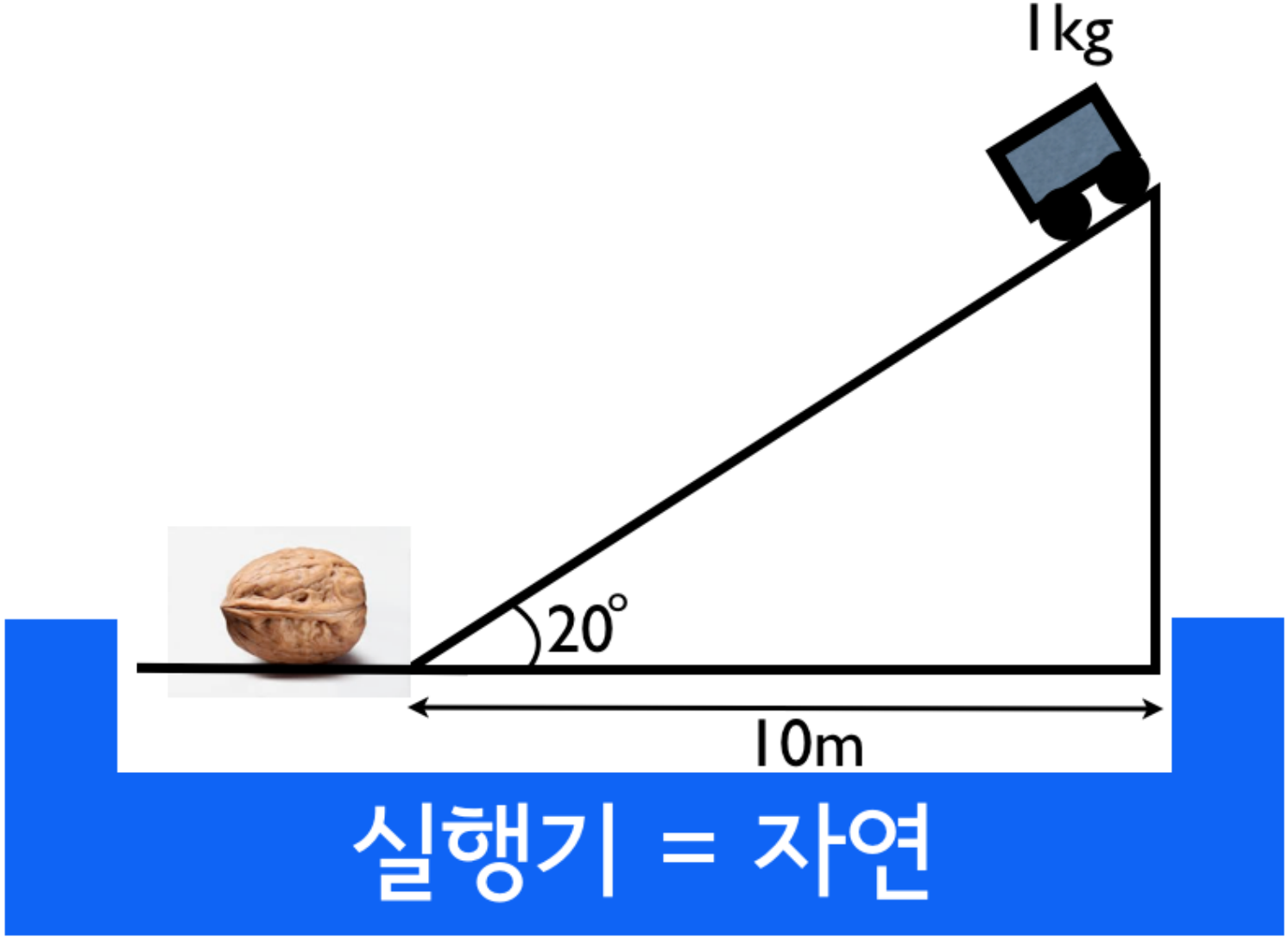
Introduction to Static Analysis, MIT Press, 2020



(저자명: 알파벳순)

SW 정적분석(static analysis)

SW의 실행미래를 빠짐없이 어림잡는 기술


1kg
20°
10m
실행기 = 자연

Software

실행기 = 컴퓨터

다른 분야와 다르지 않은

프로그램	$\Longleftrightarrow$	기계/화공/전기/금융
실행 = 컴퓨터 (언어정의)	$\Longleftrightarrow$	작동 = 자연 (자연법칙)
실행에대한 방정식	$\Longleftrightarrow$	작동에대한 방정식
방정식 풀기	$\Longleftrightarrow$	방정식 풀기
“생각대로 돌겠네”	$\Longleftrightarrow$	“생각대로 작동하겠네”
언어와 논리 이론	$\Longleftrightarrow$	물리-확학법칙, XX방정식
컴퓨터과학	$\Longleftrightarrow$	자연과학

정적분석(static analysis)

프로그램의 실행 성질을

실행전에 자동으로

안전하게 어림잡는

일반적인 방법

정적분석(static analysis)

프로그램의 실행 성질을
실행전에 자동으로
안전하게 어림잡는
일반적인 방법

- 응용: sw 오류검증, sw 관리, sw 테스트, sw 최적화 등 ∼상상력
- 다양한 레벨/목적/이름으로 존재:

일반이론	요약해석(abstract interpretation)
언어/논리	타입시스템(type system), 프로그램논리(program logic)
sw공학/컴파일러	모델체킹, 데이터흐름분석

정적분석

- "실행전": 프로그램을 실행시키지 않고
- "자동으로": 프로그램이 프로그램을 분석
- "안전하게": 모든 가능성을 포섭
- "어림잡는": 실제 이외의 것들이 포함됨
  - 어림잡지 않으면 불가능
- "일반적": 소스언어와 성질에 무제한
- "1:1": 분석기1개당, 1언어 1성질

정적분석 예

$128 \times 22 + (1920 \times -10) + 4$는 어떤 값을 계산합니까?

- 정적분석: "정수입니다."
- 정적분석: "짝수입니다."
- 정적분석: "$-10,0000$과 $1,0000$ 사이의 수입니다."
- 정적분석: "음수입니다."

정적분석 예

$$x = \text{readInt};$$
$$\text{while } (x \leq 99)$$
$$x{+}{+};$$

실행중/후에 변수 x가 가질 값은?

- 정적분석: “정수입니다”
- 정적분석: “양수입니다”
- 정적분석: “100이상 정수입니다”

정적분석기 개발 싸이클

1. 분석할 소스언어 결정:
   - Python, C, Rust, Java, ML, Scala, KVM, binary 등
2. 알고싶은 실행성질 결정:
   - 데이터모양, 메모리 접근범위, 변수 값범위, 변수관계 등
3. (설계; 구현; 테스트)$^{+}$

모든 정적분석은 3스텝

1. "연립방정식"을 세운다
   - 요약된 세계에서
   - 프로그램의 실행에 관한
2. 그 방정식을 푼다
3. 그 해를 가지고 결론을 내린다
   - 있는가 없는가? 같은가 다른가?

정적분석 이론

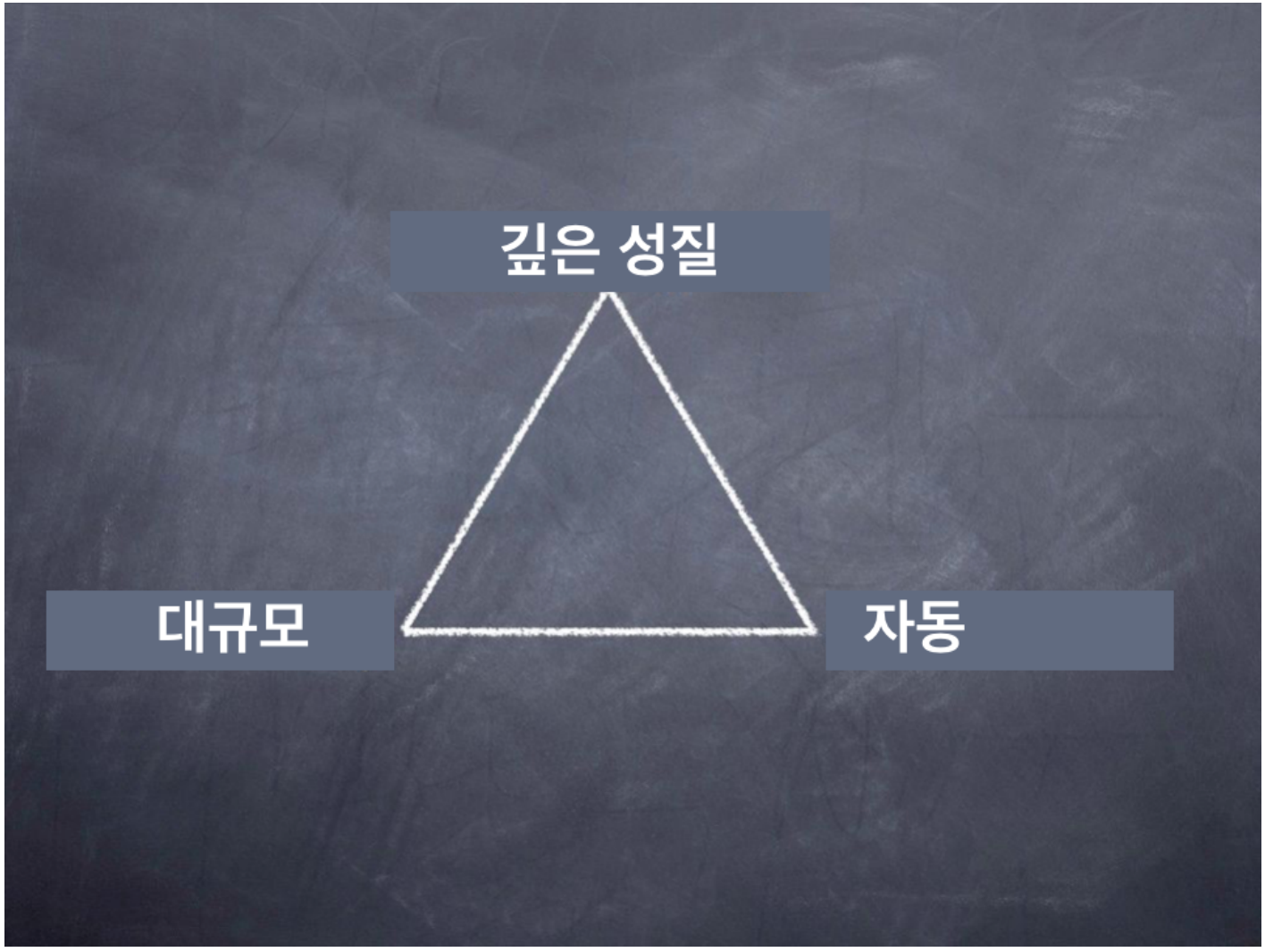
프로그래밍언어 이론과 어휘를 사용

- 올바른 연립방정식을 어떻게 유도하는가?
  - 방정식이 실제 실행의 모든 것을 꼼꼼히 보는가?
- 유도한 연립방정식의 해는 항상 있는가?
  - 그 해를 어떻게 계산하는가?
  - 유한시간내에 해를 구할 수 있는가?
  - 빨리 해를 구할 수 있는가?

정적분석 기술의 한계 I

마의 삼각형

단 둘 만 가능



정적분석 기술의 한계 II

허위정보

- 실행 경우가 아닌것을 포함
- 이론적으로 불가능: 허위정보가 항상 0

우리가보는 틈새 (SW 오류 분석에서)

오류 검출과 무결점 검증

- 오류 검출(bug-finding)
  - 허위 경보가 항상 적음 ($\leq 20\%$)
  - 오류를 모두 찾지 못함
- 무결점 검증(verification)
  - 허위경보가 거의 0
  - 오류를 모두 찾는 것이 보장
  - 특정 SW에 대해서만

오류 검출기(bug-finder)

있는 오류 대부분을 검출

허위 경보율 $\leq 20\%$

무결점 검증기(verifier)

오류가 없으면 없다고 확인

허위 경보율 $\leq 1\%$

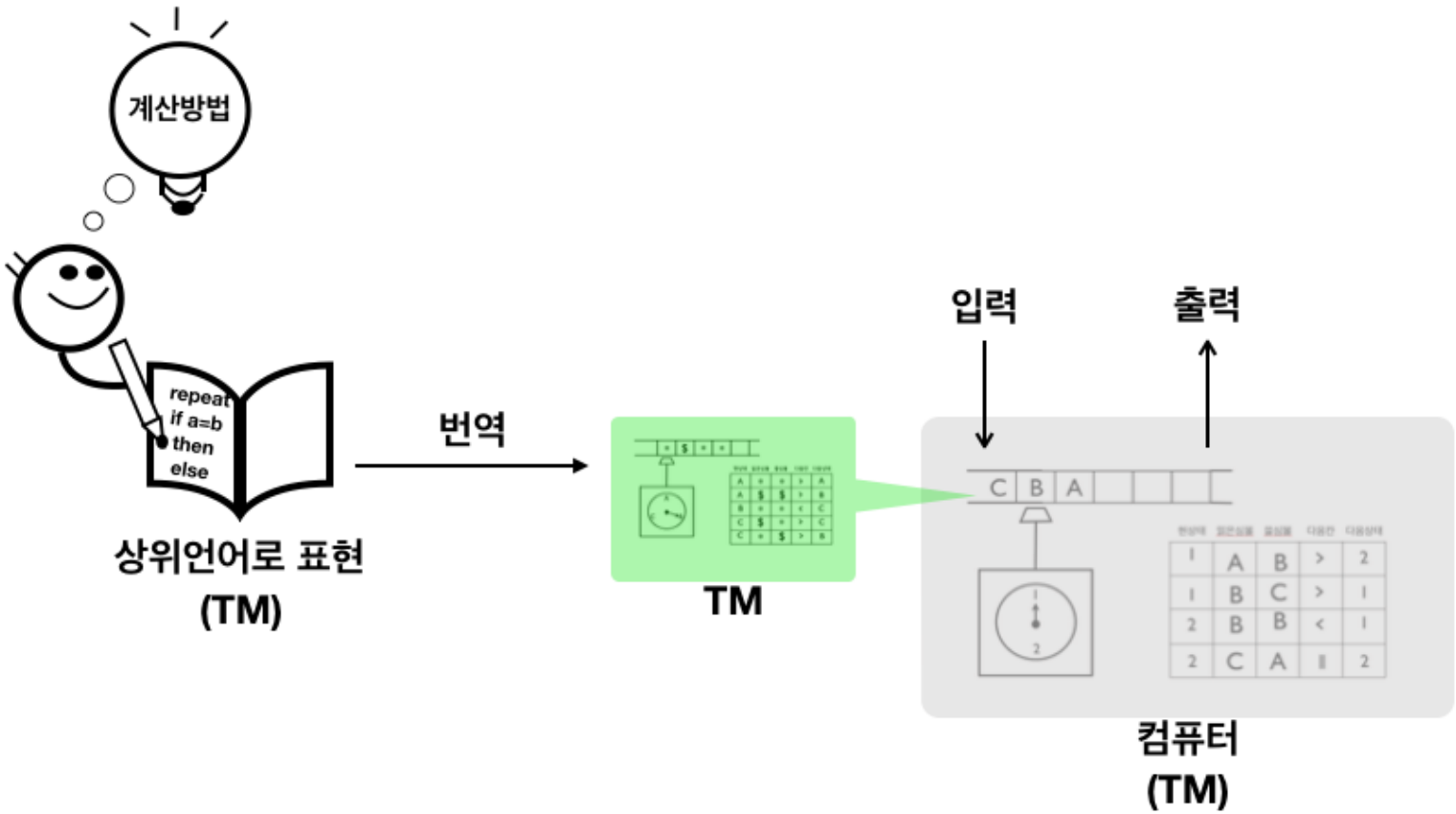
정적분석 기술 제품상황

전세계적으로

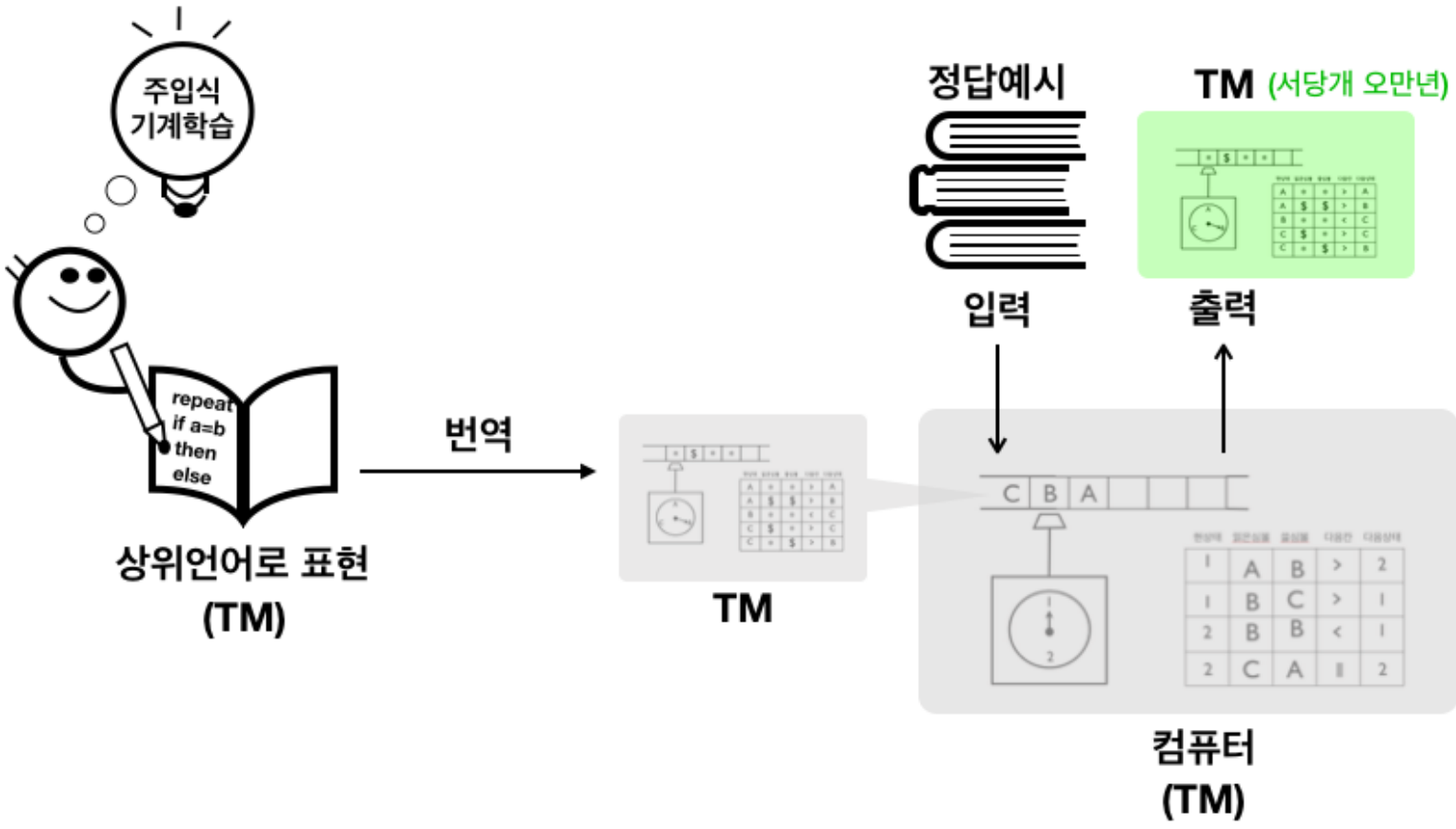
- 무결점 검증기(verification-level analyzer) 개발 업체는 아직.
- 오류 검출기(bug-finder)들이 시장에 있슴
  - 경쟁 잣대: 오류검출력 vs 허위경보율 vs "센스있는" 오류들 vs UI+UX
- 우리 예)

www.sparrowfasoo.com

정적분석의 새로운 도전: 신 SW종족의 출현

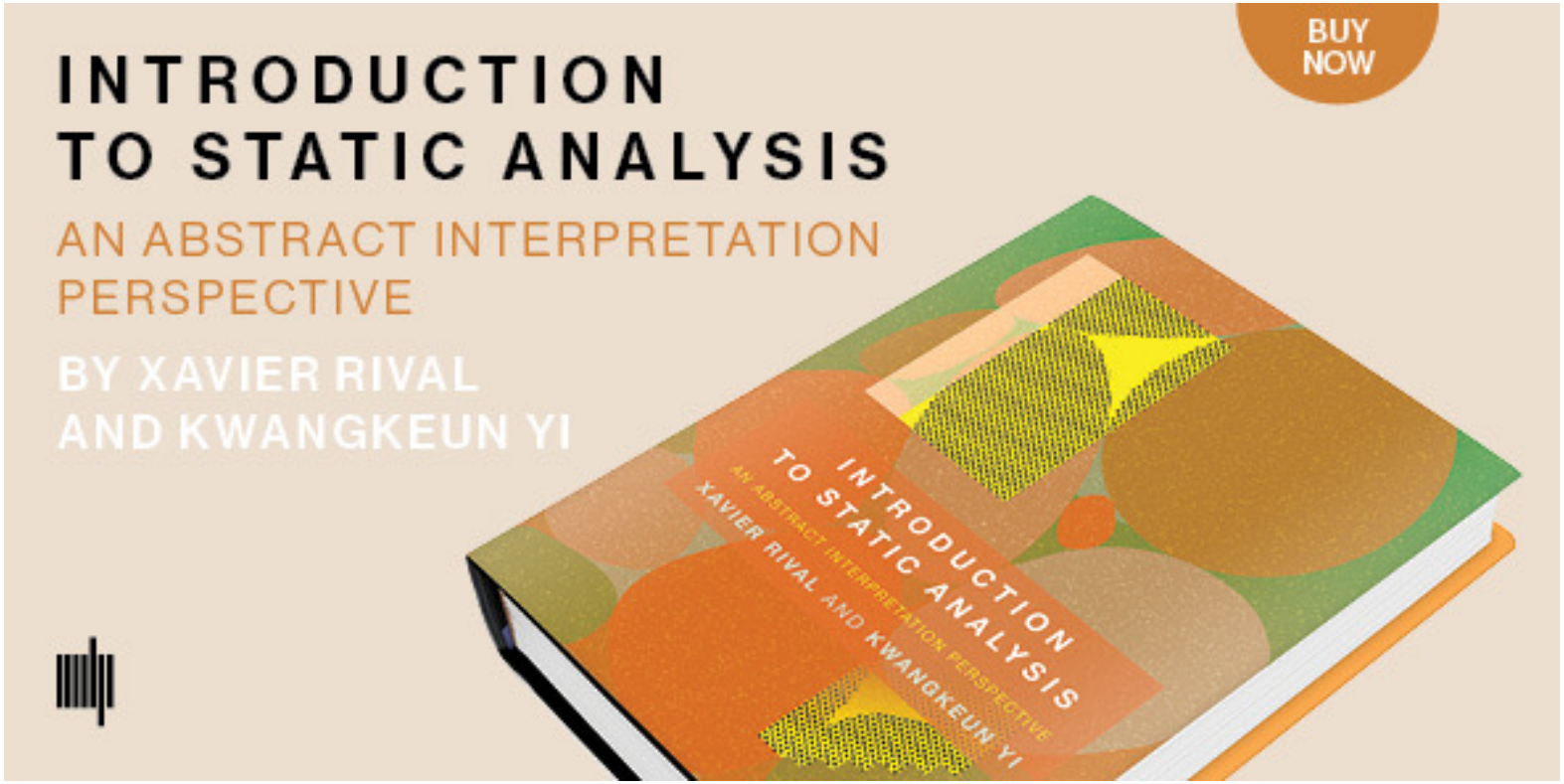


정적분석의 새로운 도전: 신 SW종족의 출현



정적분석: 성숙(이론+실제) & 도전

Introduction to Static Analysis, MIT Press, 2020



(저자명: 알파벳순)

기반: 컴퓨터과학의 원천 아이디어들

(4장: “소프트웨어, 지혜로 짓는 세계”)

컴퓨터과학이 여는 세계, 인사이트, 2015