

RECOMENDACIÓN DE VIDEOJUEGOS BASADO EN ANÁLISIS SEMÁNTICO Y MINERÍA DE OPINIÓN

DANIEL YELAMOS

TUTOR: ALEJANDRO BELLOGIN

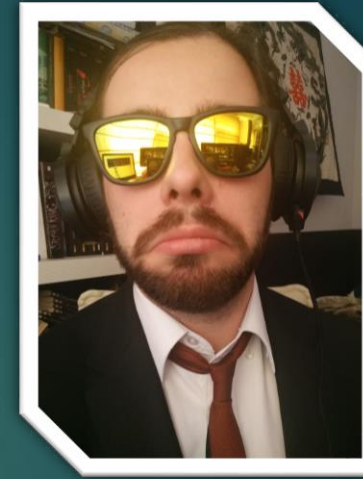
PONENTE: PABLO CASTELLS



GAME RECOMENDATION SYSTEMS BASED ON SEMANTIC ANALYSIS AND OPINION MINING



A short Story: Joe's dilemma



- ▶ Joe wants a game
- ▶ He wants a game on sale (with an active discount)

- ▶ Joe has a problem
 - ▶ 13 pages of 25 games on sale
 - ▶ $13 \times 25 = \underline{325}$ games to choose from

	Deponia STEAMPLAY™	7 Aug. 2012	👍 -90%	49,99€ 4,99€
	The Whispered World Special Edition STEAMPLAY™	6 May, 2014	👍 -90%	49,99€ 4,99€
	Out of the Park Baseball 17 STEAMPLAY™	22 Mar. 2016	👍 -50%	36,99€ 18,49€
	Goblins and Grottos STEAMPLAY™	7 Jul. 2016	👍 -25%	12,99€ 9,74€
	The Red Solstice	9 Jul. 2015	👍 -50%	19,99€ 9,99€
	Project Nimbus STEAMPLAY™	5 Nov. 2014	👍 -25%	14,99€ 11,24€
	Euro Truck Simulator STEAMPLAY™	8 Mar. 2013	👍 -75%	7,99€ 1,99€
	AR-K: The Great Escape STEAMPLAY™	14 Jul. 2015	👍 -60%	7,99€ 3,19€
	Attrition: Nuclear Domination	7 Aug. 2015	👎 -51%	0,99€ 0,49€

showing 1 - 25 of 302

1 2 3 ... 13 >

A short Story: Joe's dilemma

- ▶ Joe is a paranoiac
- ▶ Joe does not trust **professional** reviews

- ▶ Joe likes to check out reviews on Steam so as to see the general opinion of a game.

- ▶ Joe has a big problem
 - ▶ Joe cannot read 4,500 reviews



DEPONIA

In Deponia, the world has degenerated into a vast garbage dump, in which the crotchety Rufus ekes out his sorry existence.

User reviews:
RECENT: **Mostly Positive** (139 reviews)
OVERALL: **Very Positive** (4,536 reviews)

Release Date: 7 Aug, 2012

Popular user-defined tags for this product:
Adventure Point & Click Comedy Indie Puzzle +

Tags you've applied to this product:
+ Add your own tags



A short Story: Joe's dilemma

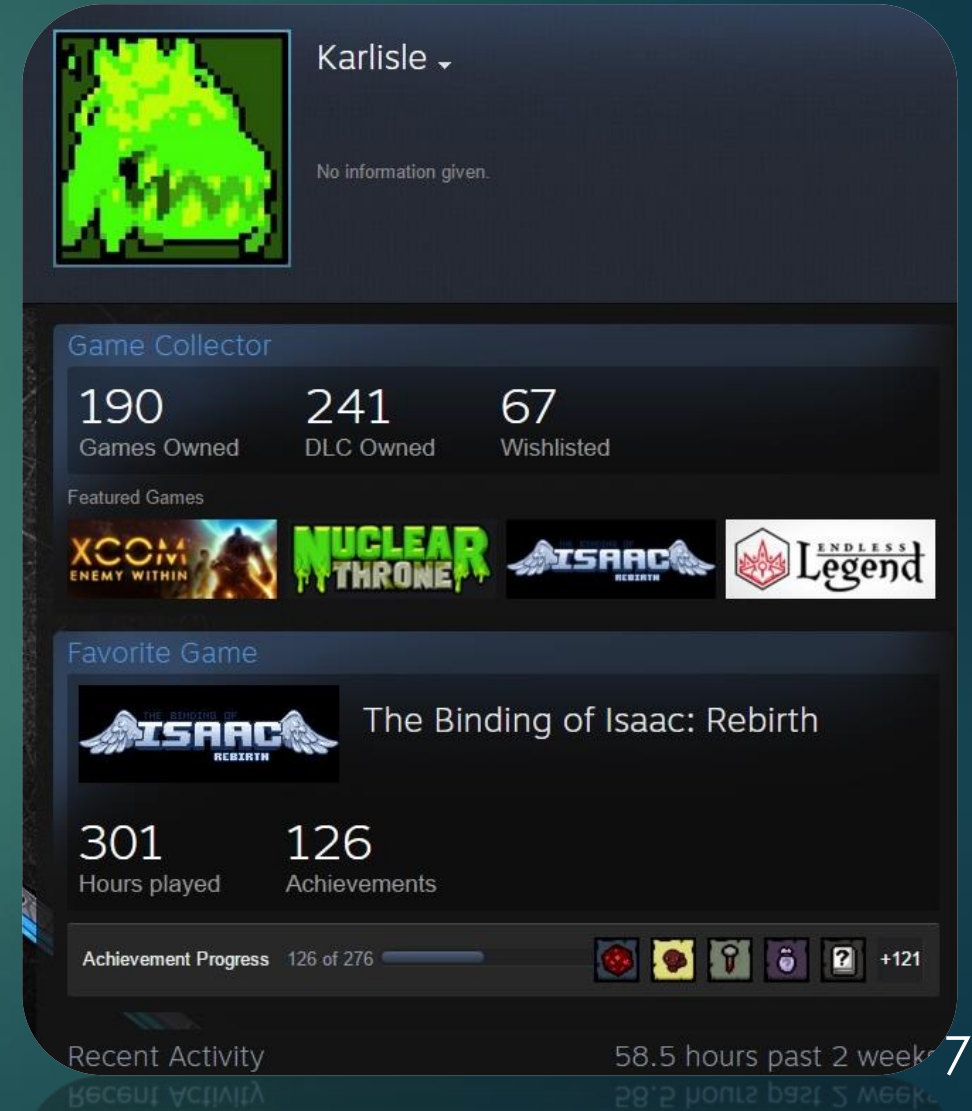
- ▶ Joe would like to ask all of steam's community to recommend him a game.
- ▶ But Joe is smart and he knows that there would be a bias and a huge difference of opinion.
- ▶ Joe needs STEAMIND

Project's purpose:

- ▶ To create a recommender system based solely on the input of Steam's gaming community that can be able to refer a user, once given a definition, no matter how short or long, of what s/he is looking for, to a game or a group of games.

I: Introduction - Personal Motivation

- ▶ I am a “hardcore gamer”
- ▶ I spend a lot of time “browsing” videogames
- ▶ A lot of things to consider
 - ▶ Price
 - ▶ Hours per euro spent
 - ▶ Genre
 - ▶ Originality
 - ▶ Setting
- ▶ Do not trust company input.
- ▶ Users are more reliable.



The screenshot shows a Steam profile for a user named Karlisle. The profile features a pixelated green and yellow avatar. Below the name, it states 'No information given.' The 'Game Collector' section displays statistics: 190 Games Owned, 241 DLC Owned, and 67 Wishlisted. The 'Featured Games' section includes XCOM: Enemy Within, Nuclear Throne, The Binding of Isaac: Rebirth, and Endless Legend. The 'Favorite Game' section highlights 'The Binding of Isaac: Rebirth' with 301 Hours played and 126 Achievements. An achievement progress bar shows 126 of 276 achievements completed, with icons for various achievements and a '+121' indicator. The 'Recent Activity' section at the bottom shows '58.5 hours past 2 weeks'.

Karlisle ▾
No information given.

Game Collector

190 Games Owned 241 DLC Owned 67 Wishlisted

Featured Games

XCOM: ENEMY WITHIN NUCLEAR THRONE THE BINDING OF ISAAC: REBIRTH ENDLESS Legend

Favorite Game

The Binding of Isaac: Rebirth

301 Hours played 126 Achievements

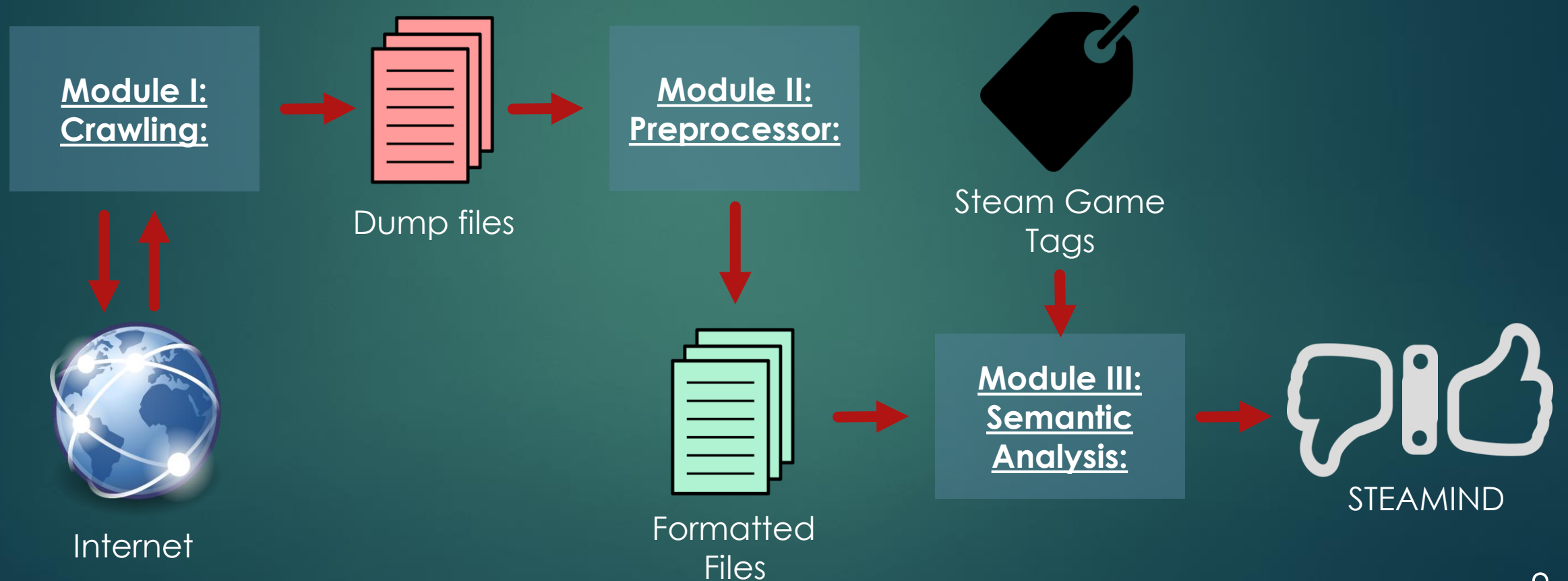
Achievement Progress 126 of 276

Recent Activity 58.5 hours past 2 weeks

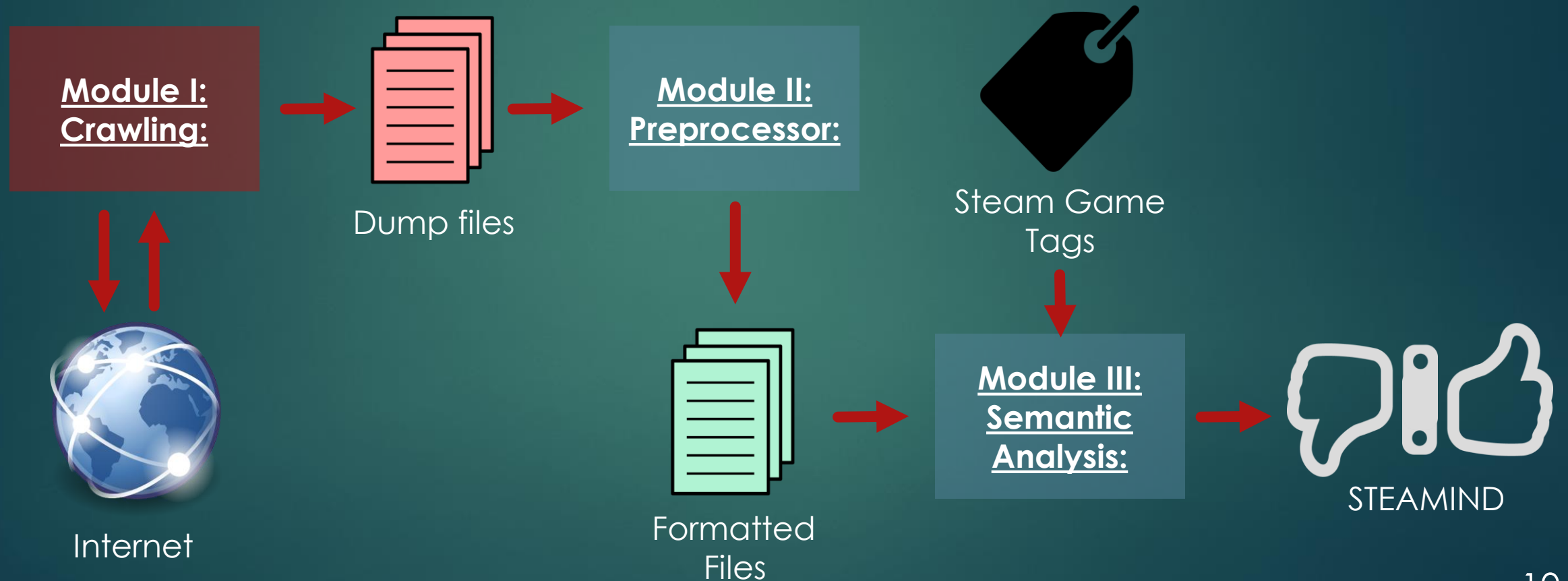
I: Introduction - Professional and technological approach.

- ▶ Semantic analysis.
 - ▶ Unsolved problems.
 - ▶ Yet multiple approaches:
 - ▶ Grammatical
 - ▶ Statistical
- ▶ Big data.
 - ▶ Well known to be a very popular topic amongst IT research nowadays.

III: Design and Development (D&D): Introduction:

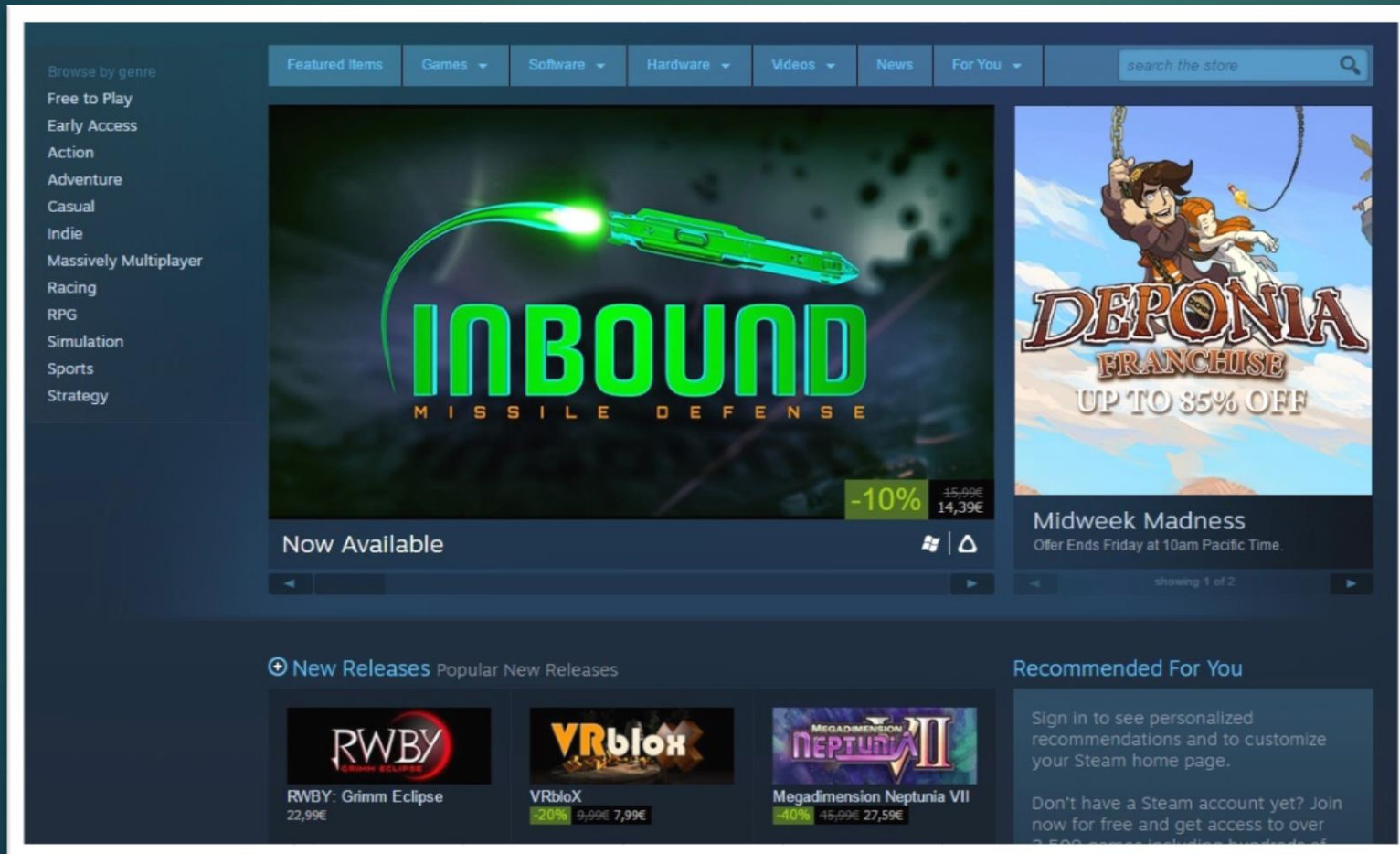
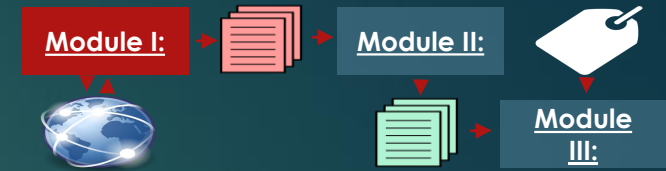


III: Design and Development (D&D): Introduction:

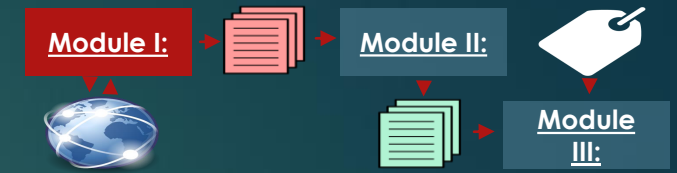


III: D&D - Crawler

Steam: Landing page

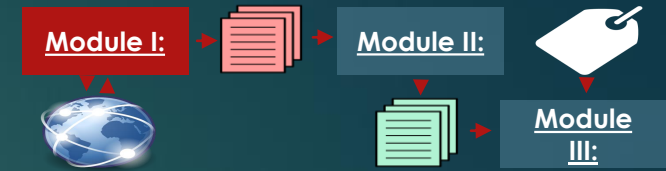


III: D&D - Crawler: Steam: Fetching page



```
mostSoldGamesAppid
appid : ArrayList<String>
constructor()
getAppids() : ArrayList<String>
```

III: D&D - Crawler: Steam: Game Reviews Page



INSIDE Store Page

All Discussions Screenshots Artwork Broadcasts Videos News Guides **Reviews**

SHOW MOST HELPFUL (ALL TIME) (?) ALL LANGUAGE ENGLISH [About Reviews](#)

42 of 70 people (60%) found this review helpful
13 people found this review funny

Recommended
5.1 hrs on record

Posted: 7 July
Product received for free

Standing on the Shoulders of Giants

Full disclosure: review copy provided by the developer/publisher!

When I think "indie games" a few come to mind: Super Meat Boy, Braid, World of Goo, and Limbo. Limbo came out way, WAY back in 2010; back when nighttime television audiences were torn between Team Coco and Team Jayjay, iPads roamed wild and free, and vuvuzelas were declared auditory torture devices. Other stuff happened too, but I've been very successful in repressing any memories of Glee and I don't want those memories rushing back now.

0 of 3 people (0%) found this review helpful
1 person found this review funny

Recommended
1.4 hrs on record

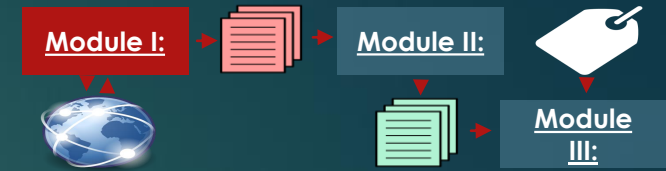
Posted: 7 July
finally

Dr.Ew
73 products in account 0

0 of 3 people (0%) found this review helpful
3 people found this review funny

```
reviewURLMiner
USER_AGENT : String
reviewCount : Integer
reviewURL : ArrayList<String>
mine(appid)
```


III: D&D - Crawler: Steam: Game Review Page



Shadowthane » Reviews » Broforce

78 of 95 people (82%) found this review helpful
72 people found this review funny

Recommended
0.0 hrs last two weeks / 68.0 hrs on record


Posted: 9 Jun @ 2:02pm

Never have I felt more free. Everytime I play this game, I get the sensation that George Washington is looking down on me, stroking his bald eagle, and chuckling with mad glee as I rip apart terrorist, alien and pig. Yes. That just happened.
Boom.

Was this review helpful? Yes No Funny

2 Comments

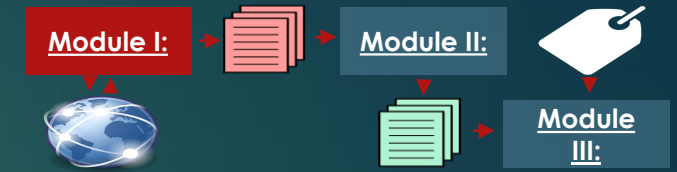
Huffulufugus 2 Jul @ 10:56pm
The part "Bald eagle" could be interpreted in a couple of ways.



- [View Store Page](#)
- [View Community Hub](#)
- [Find More Like This](#)

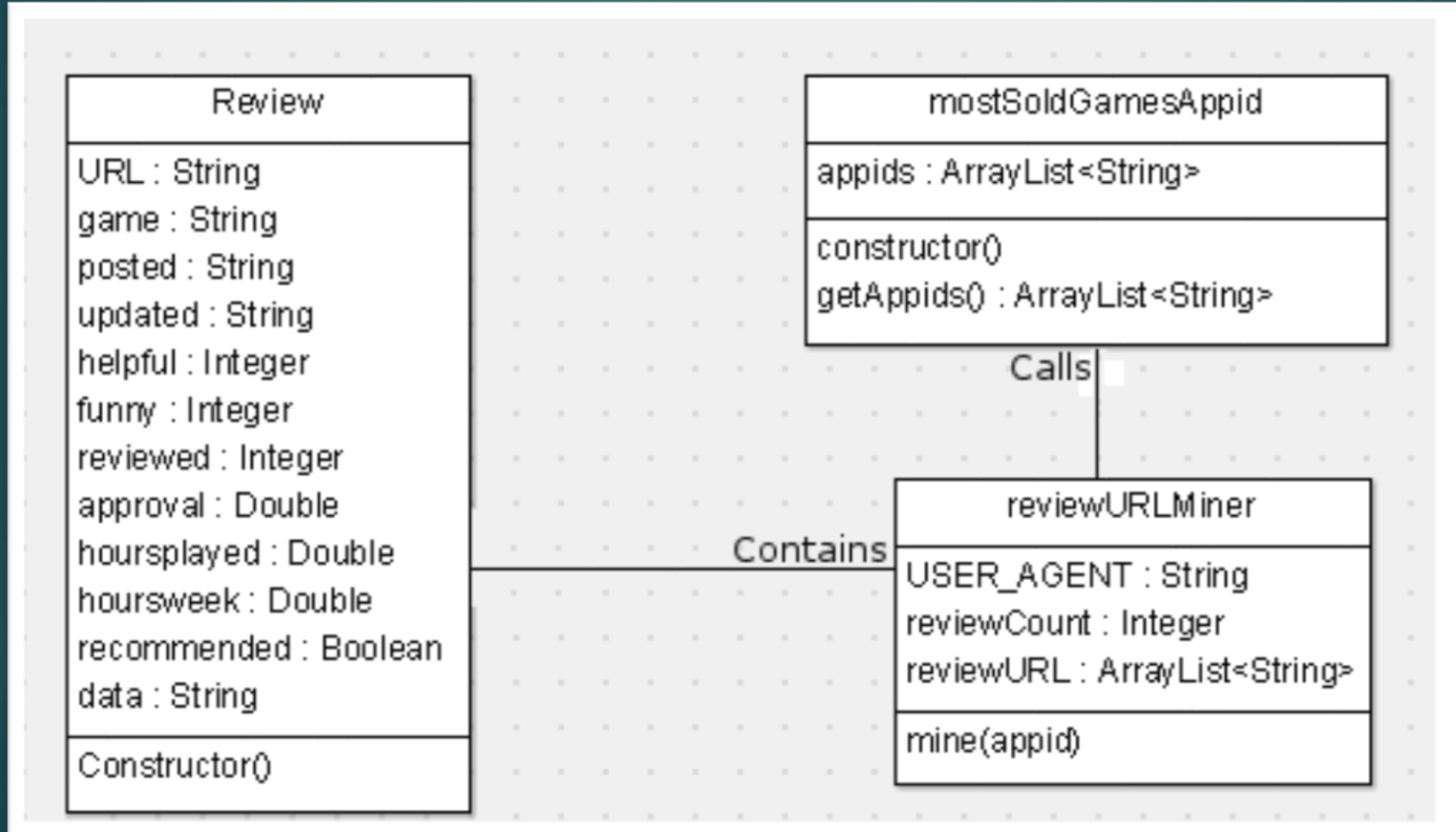
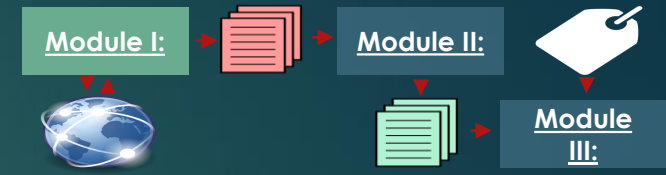
Review
URL : String
game : String
posted : String
updated : String
helpful : Integer
funny : Integer
reviewed : Integer
approval : Double
hoursplayed : Double
hoursweek : Double
recommended : Boolean
data : String
Constructor()

III: D&D - Crawler Requirement Analysis

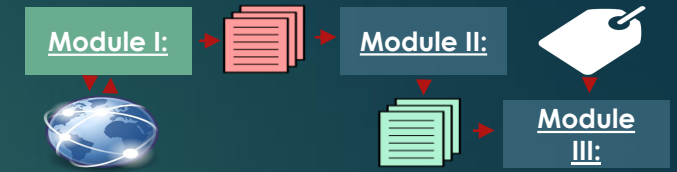


- ▶ Functional Requirements:
 - ▶ Steam Integration
 - ▶ **Infinite Scroll**
 - ▶ Game list collection
 - ▶ URL builder
 - ▶ Game review crawler
 - ▶ Game review data extractor
 - ▶ Lost URL detection
 - ▶ .txt dump file

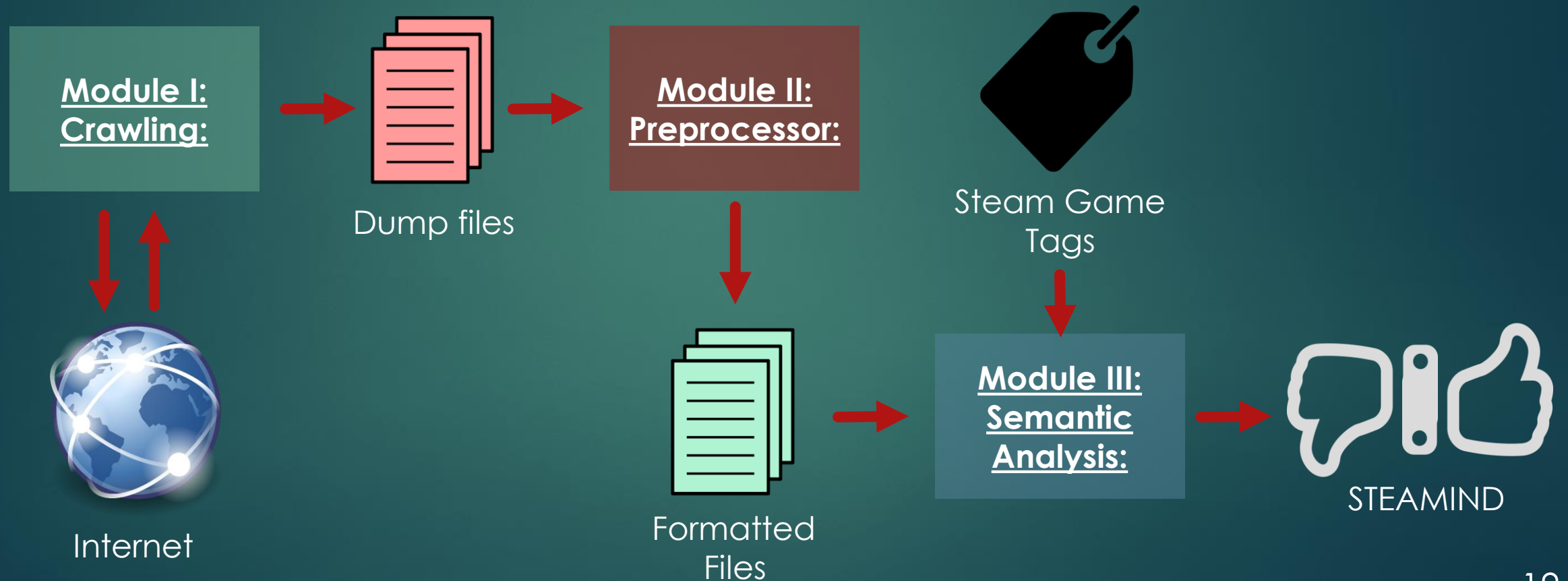
III: D&D - Crawler Package structure.



III: D&D - Crawler Final prototype.

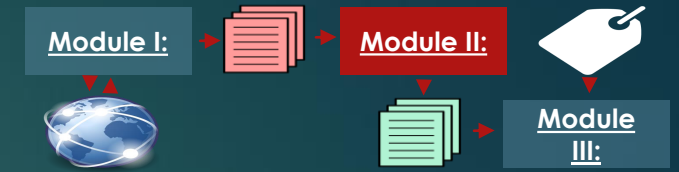


III: Design and Development (D&D): Introduction:

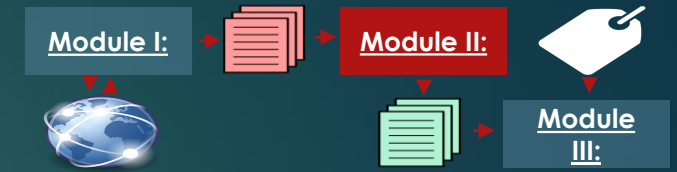


III: D&D - Preprocessor Requirement Analysis

- ▶ Functional Requirements:
 - ▶ Data low level parsing
 - ▶ Review integrity
 - ▶ Review Formatting

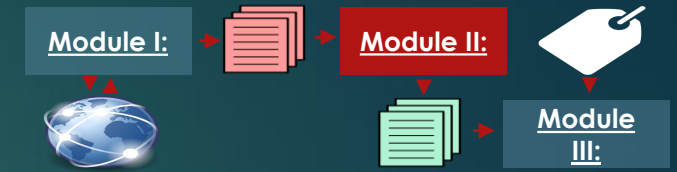


III: D&D - Preprocessor Design: “Review” examples

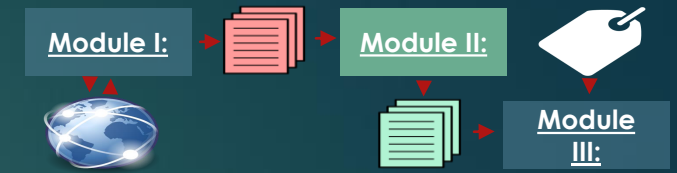


III: D&D - Preprocessor

Preprocessing rules:

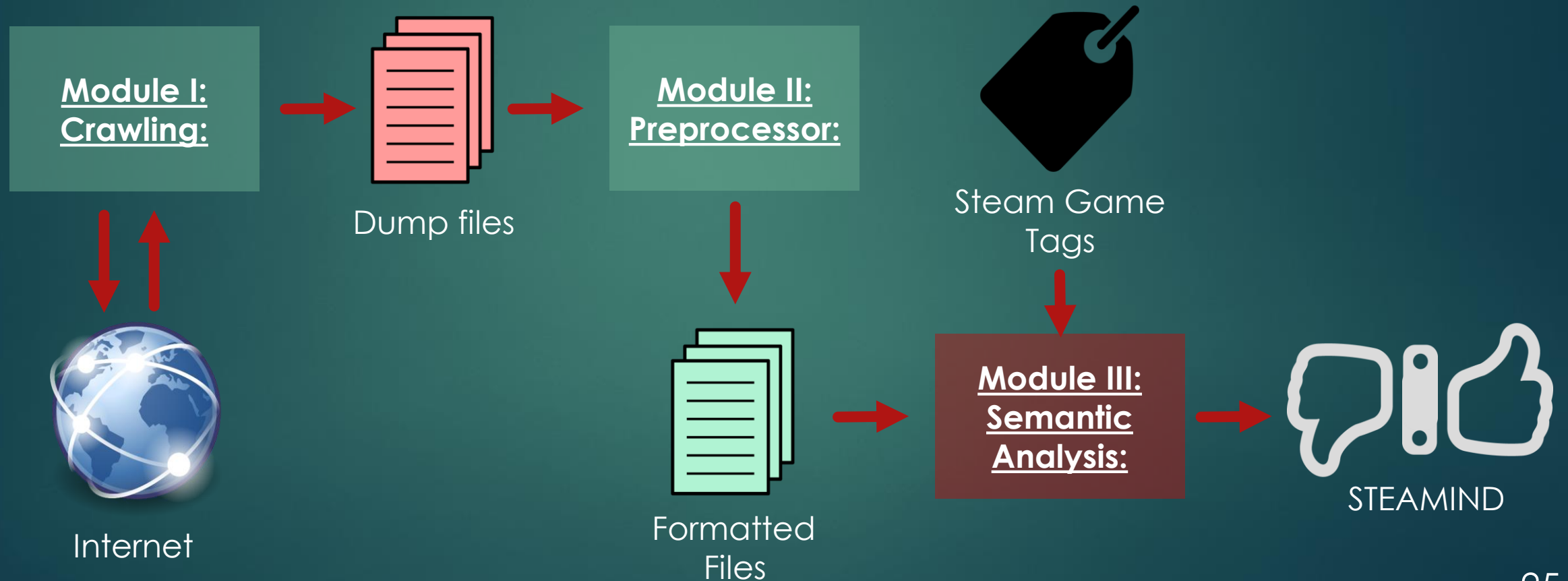


III: D&D - Preprocessor Final Prototype

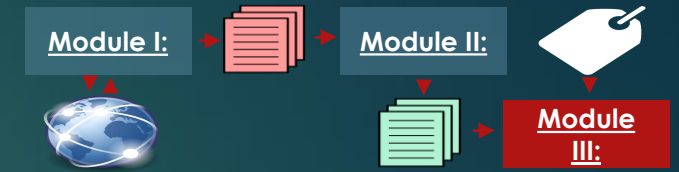


- ▶ Applies the preprocessing rules.
- ▶ Splits the reviews per games.
- ▶ Dumps them in separate files depending on the needs of the third module
 - ▶ Each technology has an specified input format.

III: Design and Development (D&D): Introduction:



III: D&D – Semantic Analysis Requirement Analysis



III: D&D – Semantic Analysis

Final Prototype: Mallet



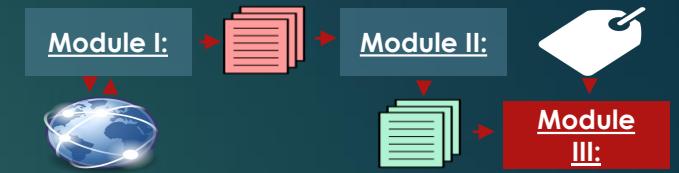
- ▶ What is Mallet?
- ▶ Statistical language processing and document classification.
- ▶ Topic extraction.
- ▶ LSA, Bag of Words.

- ▶ Corpus: Soccer news:
 - ▶ It would know that “Real Madrid, Barcelona... etc”
 - ▶ It won’t know that it is talking about soccer.



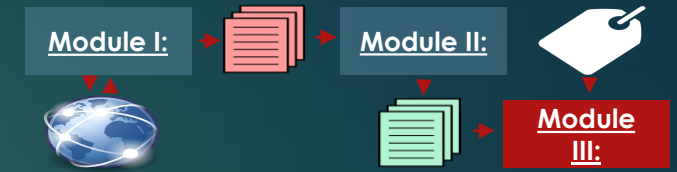
III: D&D - Semantic Analysis

Design: Mallet



III: D&D - Semantic Analysis

Design: Mallet



Topic 0:

witcher
story
world
quest
combat

Topic 1:

Great
fun
ive
good
time

Topic 2:

gta
online
story
rockstar
fun

Topic 3:

space
paradox
stellaris
empire
strategy

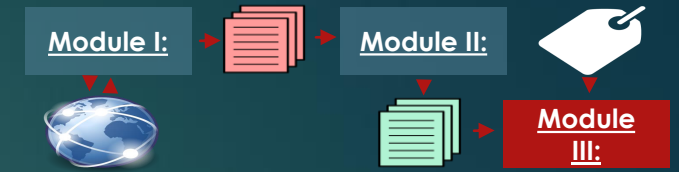
...

Topic 20:

amazing
story
feel
fps
great

III: D&D – Semantic Analysis

Design: Mallet



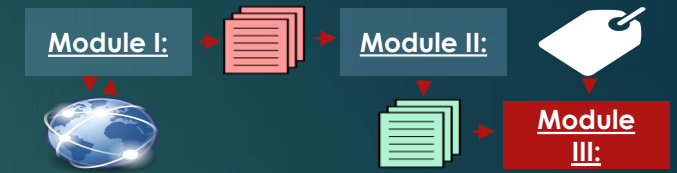
Given this text: “I want a game that is fast, action based, no thinking, shoot shoot shoot.”
What **topic** would it belong to?

Step 1 model:

Given that text, it would belong to **topic 4**.

III: D&D – Semantic Analysis

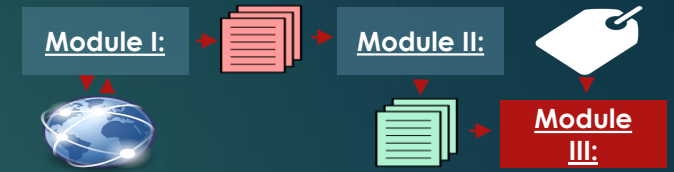
Design: Mallet



- ▶ Step 2: Topics to games.
- ▶ Brief explanation of the tag system:
 - ▶ Users can vote for a tag created by themselves or other users to be assigned to a game. This voting system is more used than the review system.
- ▶ Input: Tags from every single game collected.
- ▶ Output: A map that relates topics to the games, related by tags.
- ▶ Note, **if a topic within the first model is not mapped to any game, it is discarded.**
- ▶ This step answers this question:
 - ▶ Given a series of tags belonging to a game. ¿Which of the topics is more likely to talk about my game?

III: D&D – Semantic Analysis

Design: Mallet



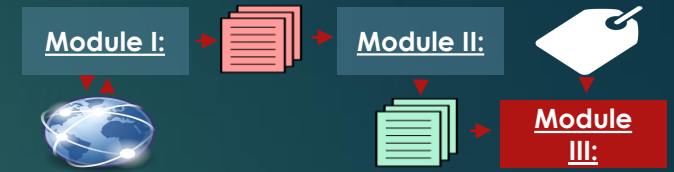
Given my **tags** for a game, which **topic** does it belong to?



Step 1 model:

Given those **tags**, it belongs to **topic 3**.

III: D&D - Module III – SA Design: Mallet



Step 2 model:

Topic 0:
1. Witcher
2. Dark Souls III

witcher
story
world
quest
combat

Topic 1:
1. War for the overworld

dungeon
keeper
campaign
Bugs
original

Topic 2:
1. Stellaris

space
paradox
stellaris
empire
strategy



Topic 17:
1. DOOM

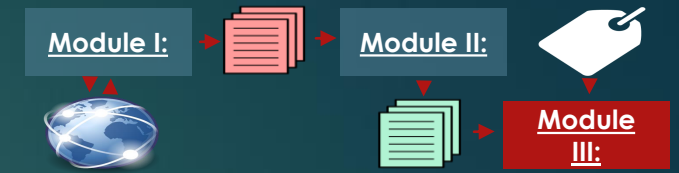
doom
multiplayer
fps
campaign
fast



Popular user-defined tags for this product: (?)

- FPS
- Action
- Gore
- Demons
- Shooter
- First-Person
- Multiplayer
- Horror
- Sci-fi
- Singleplayer
- Classic
- Fast-Paced
- Great Soundtrack

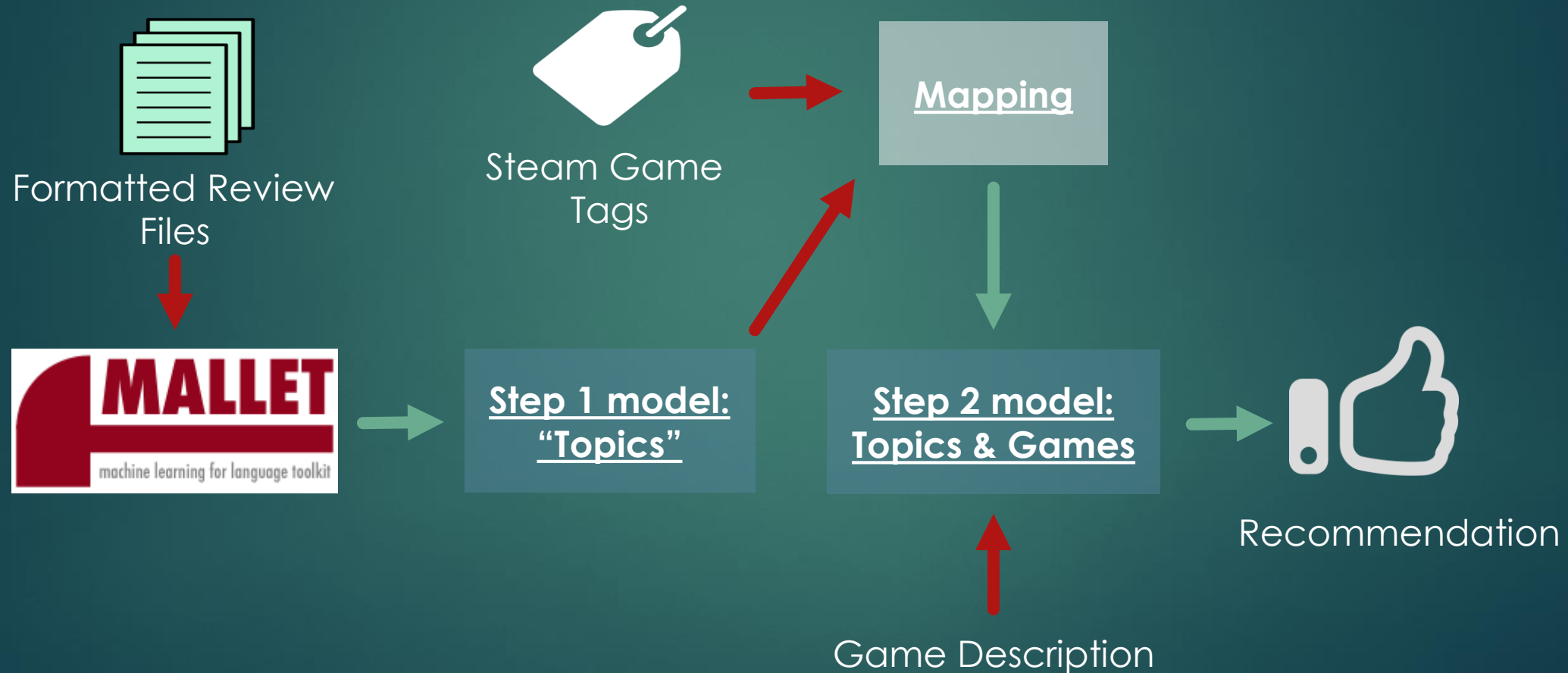
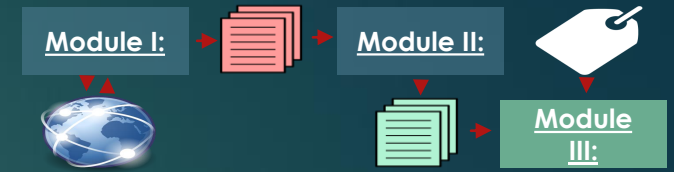
III: D&D - Module III – SA Design: Mallet



- ▶ Step 3: Testing and recommending.
- ▶ A user input is matched to the most probable topic.
- ▶ That topic will be mapped to at least a game.
- ▶ The recommendation is done.

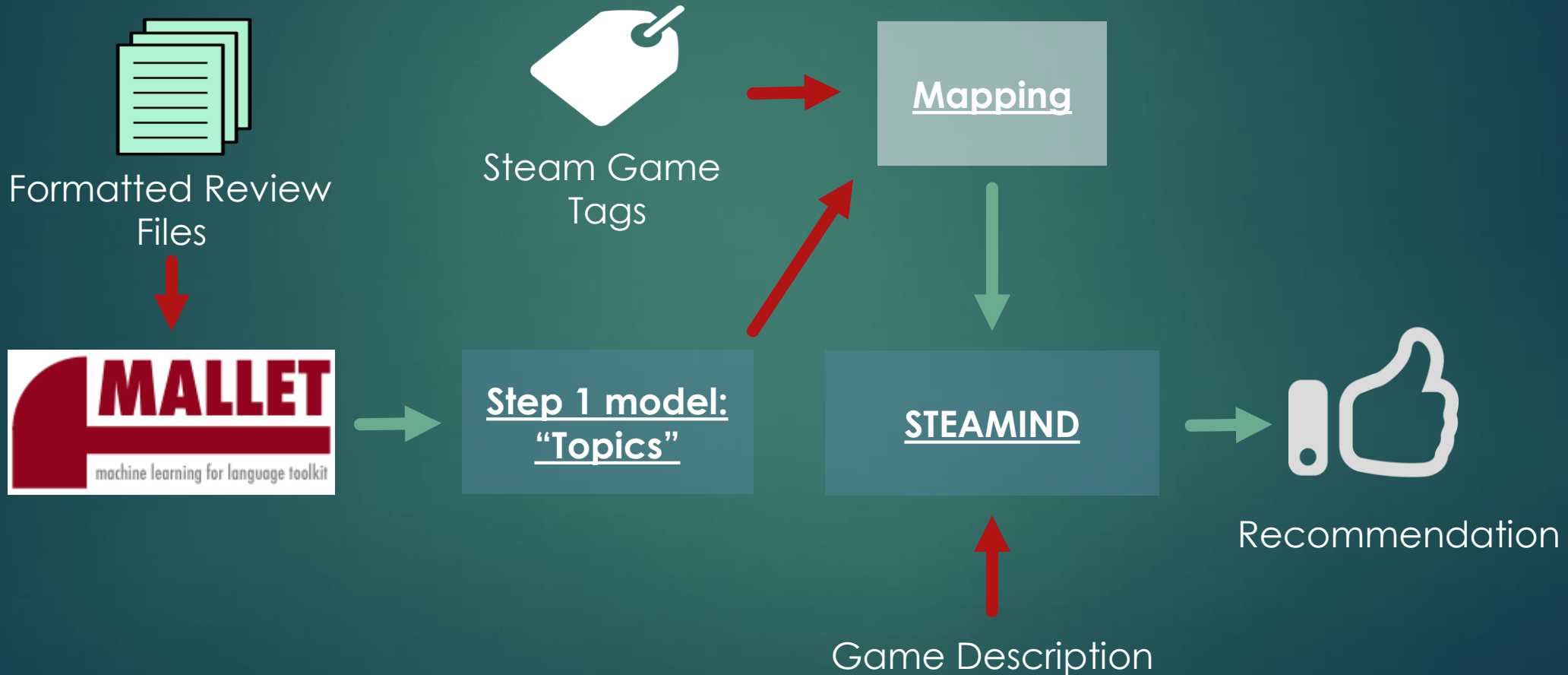
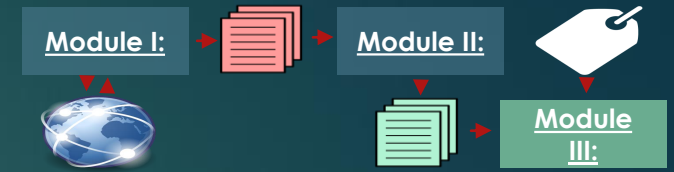
III: D&D - Module III – SA

Final Prototype: Mallet

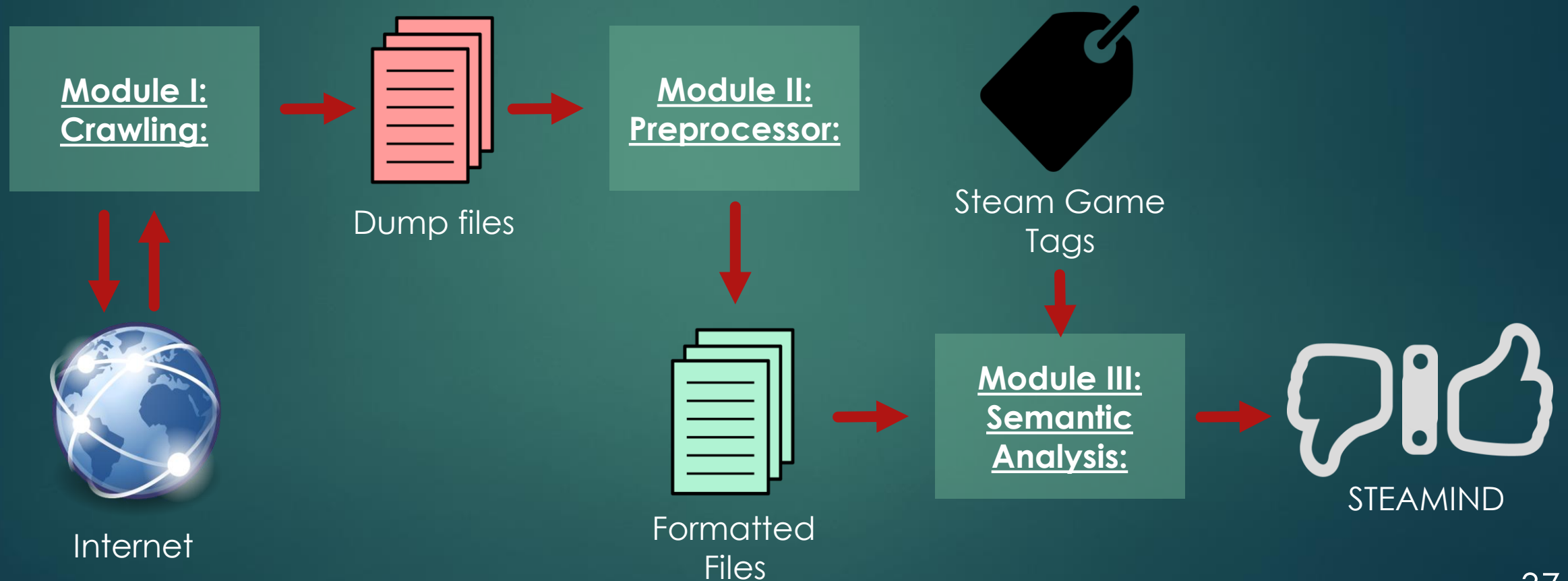


III: D&D - Module III – SA

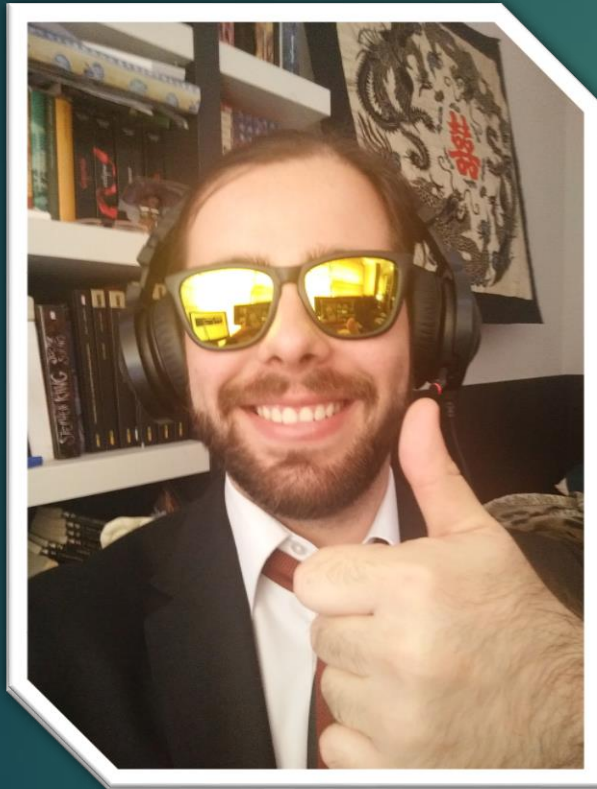
Final Prototype: Mallet



III: Design and Development (D&D): Introduction:



Let's get back to Joe



STEAMIND



"You want to play"

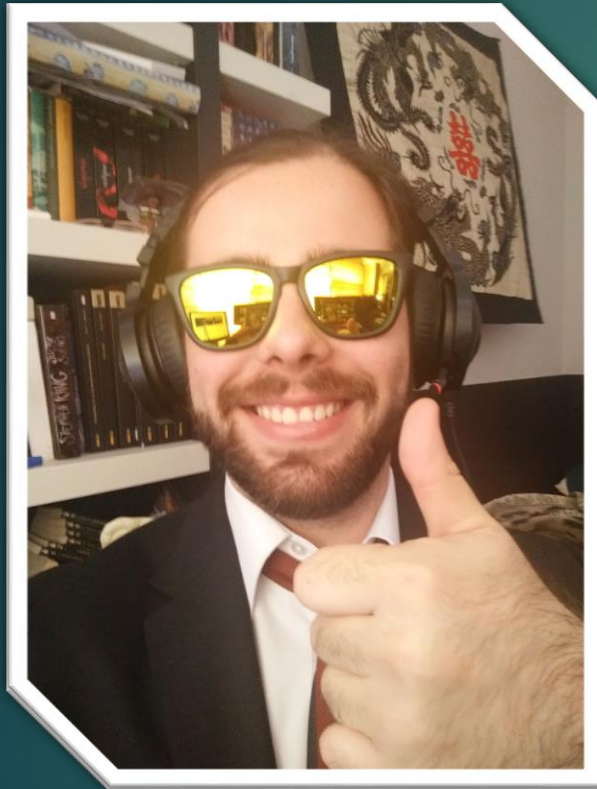
Topic 4:

WITCHER III
DARK SOULS III

"I want to kill dragons"



Let's get back to Joe



STEAMIND



"You want to play"

Topic 20:

Counter Strike
Insurgency

"I want a competitive shooter fast paced, realistic, that has complex class system that is balanced and requires a lot of playing to get used to"

IV: Testing and Results.

Introduction

- ▶ How can this be tested.
- ▶ User based testing.
 - ▶ Long time.
 - ▶ Lots of users.
 - ▶ Hard to evaluate.
- ▶ Automatic partitioning testing.
 - ▶ Divide the corpus in two partitions. Test, Train.
 - ▶ See if the model recommends the game the test partition was originally talking about.
- ▶ Second partition (out of 3) results shown.

IV: Testing and Results.

Testing

This review from my test partition belongs to Witcher 3:

“ ... ”

Would the model be able to know?

Step 2 model:

It belongs to The Witcher 3

It belongs to DOOM

IV: Testing and Results.

Introduction

- ▶ From the crawler, 18 games were mined for about 90,000 reviews.
- ▶ Not every game had the same amount of reviews.
- ▶ For review integrity purpose, 3 games were discarded.
- ▶ Their reviews confused the model more than contribute to it.

IV: Testing and Results.

Population

- ▶ From the crawler, 18 games were mined for about 90,000 reviews.
- ▶ Not every game had the same amount of reviews.

- ▶ For review integrity purpose, 3 games were discarded.
- ▶ Their reviews confused the model more than contribute to it.

- ▶ Final population:
 - ▶ 15 games
 - ▶ 80,000 reviews

IV: Testing and Results.

Game list



Second Partition: Step 2 model

Topic 0:
1. War for the
overworld

dungeon
keeper
campaign
bugs
original

Topic 16:
1. Rocket
League

fun
rocket
cars
league
friends

Topic 17:
1. DOOM

doom
multiplayer
fps
campaign
fast

Topic 18:
1. Insurgency
2. Vermintide
3. Counter
Strike

fps realistic
insurgency
team
tactical

Topic 2:
1. Stellaris

space
paradox
stellaris
empire
strategy

Topic 3:
1. Metal
Gear V

metal
gear mgs
kojima
missions

Topic 19:
1. GTA V

gta
online
rockstar
fun
friends

Topic 5:
1. ARK

server
survival
dinos
early
dinosaurs

Topic 8:
1. Dark
Souls III
2. Witcher
III

witcher
world
quests
combat
rpg

Topic 11:
1. Total War
Warhammer

war total
warhammer
units
campaign

Topic 12:
1. Battleborn

characters
battleborn
moba
borderlands
pvp

Topic 15:
1. Metal
Gear Solid
Revengeance

metal
gear
boss
action
fun

IV: Testing and Results.

Second partition: Successful maps

Table 4.4: Successfully Mapped Games: Partition II		13/15
Game Title	Hit rate	Group assigned
War_for_the_Overworld	97%	War_for_the_Overworld
Battleborn	83%	Battleborn
The_Witcher_3	73%	DARK_SOULS_III, The_Witcher_3
ARK_Survival_Evolved	75%	ARK_Survival_Evolved
Rocket_League	73%	Rocket_League
Insurgency	85%	Insurgency, Counter_Strike
GTA_V	80%	GTA_V
Stellaris	81%	Stellaris
DOOM	85%	DOOM
Total_War_WARHAMMER	81%	Total_War_WARHAMMER
METAL_GEAR_RISING_REVENGEANCE	77%	METAL_GEAR_RISING_REVENGEANCE
Counter_Strike	29%	Insurgency, Counter_Strike
METAL_GEAR_SOLID_V	78%	METAL_GEAR_SOLID_V

IV: Testing and Results.

Second partition: !Successful maps

Table 4.5: Unsuccessfully Mapped Games: Partition II

2/15

Game Title	Hit rate	Group assigned
DARK_SOULS_III	33%	War_for_the_Overworld
Warhammer_Vermintide	24%	War_for_the_Overworld

IV: Testing and Results. Comparison

Table 4.8: Successfully Mapped Games: Overall			
Game Title	Hit rate Partition I	Hit rate Partition II	Hit Rate Partition III
War_for_the_Overworld	91%	97%	91%
Battleborn	88%	83%	85%
The_Witcher_3	75%	73%	65%
ARK_Survival_Evolved	73%	75%	77%
Rocket_League	77%	73%	73%
Insurgency	87%	85%	84%
GTA_V	82%	80%	85%
Stellaris	80%	81%	86%
DOOM	85%	85%	87%
Total_War_WARHAMMER	84%	81%	85%
METAL_GEAR_RISING REVENGEANCE	85%	77%	81%
Counter_Strike	29%	29%	-----
METAL_GEAR_SOLID_V	-----	78%	91%
OVERALL	72%	76%	75%

V: Future work and conclusions.

Future Improvements

- ▶ Create more precise recommendation systems.
 - ▶ Using individual models that classify better a certain group of games.
- ▶ Upgrade filesystem.
 - ▶ Database needed.

V: Future work and conclusions.

Knowledge acquired

- ▶ Certain games are easier to map than others.
 - ▶ Average age of the gamer
 - ▶ Community
 - ▶ Competitive spirit & Difficulty
- ▶ LSA/bag of words VS contextual analysis

V: Future work and conclusions.

How can this be used?

- ▶ Improve Steam's current recommendation system.
- ▶ Predict if comments from a certain source are relevant without needing user active input.
- ▶ This system could be expanded to anything that holds as much information.
 - ▶ Amazon
 - ▶ Restaurants
 - ▶ Movies
 - ▶ All have a very small amount of comments(500-1000)
 - ▶ Youtube videos.

VI: ¿Questions?

VI: Aux: Word2Vec

- ▶ Why didn't it work?
- ▶ It was far too new
- ▶ The Paravec Approach was not mint enough.

- ▶ <https://github.com/deeplearning4j/deeplearning4j/issues/1657>
- ▶ <https://github.com/deeplearning4j/deeplearning4j/issues/1623>

HTTP:GET petition

- ▶ String url = "http://steamcommunity.com/app/" + appid + "/homecontent/?"
- ▶ + "userreviewsoffset=" + count
- ▶ + "&p=" + counter
- ▶ + "&itemspage=2" + counter
- ▶ + "&screenshotspage=" + counter
- ▶ + "&videospage=" + counter
- ▶ + "&artpage=" + counter
- ▶ + "&allguidepage=" + counter
- ▶ + "&webguidepage=" + counter
- ▶ + "&integratedguidepage=" + counter
- ▶ + "&discussionpage=" + counter
- ▶ + "&l=english" + "&appHubSubSection=10" + "&browsefilter=toprated"
- ▶ + "&filterLanguage=default&searchText" + "&forceanon=1";
- ▶