

QCon[上海站]

全球软件开发大会 2016

新媒体 深度学习

SPEAKER

孙子荀

International Software
Development Conference

主办方 **Geekbang** > **InfoQ**
极客邦科技



促进软件开发领域知识与创新的传播



关注InfoQ官方信息
及时获取QCon软件开发者
大会演讲视频信息



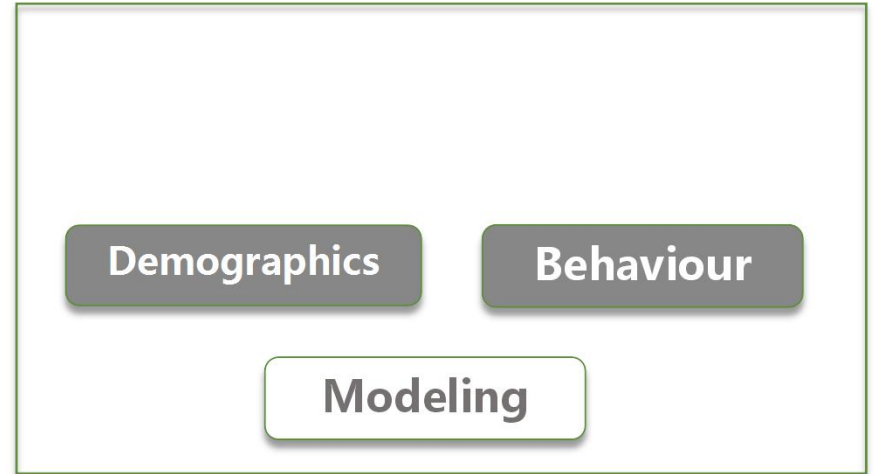
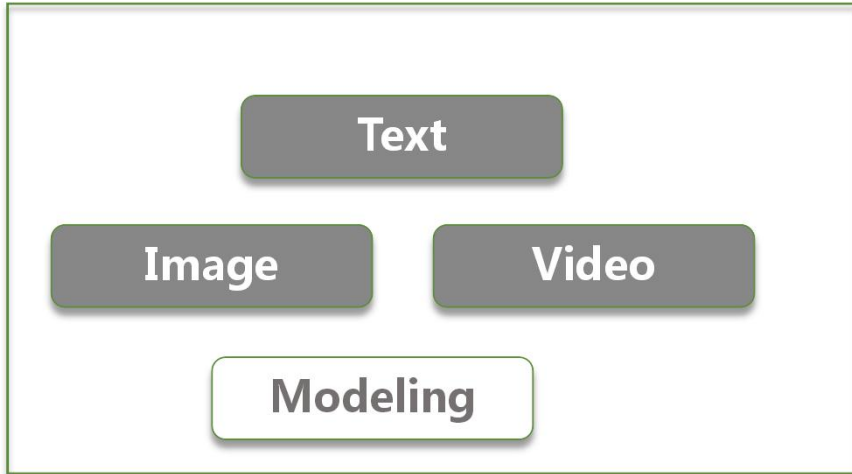
[北京站] 2016年12月2日-3日
咨询热线: 010-89880682



[北京站] 2017年4月16日-18日
咨询热线: 010-64738142

新媒体 深度学习

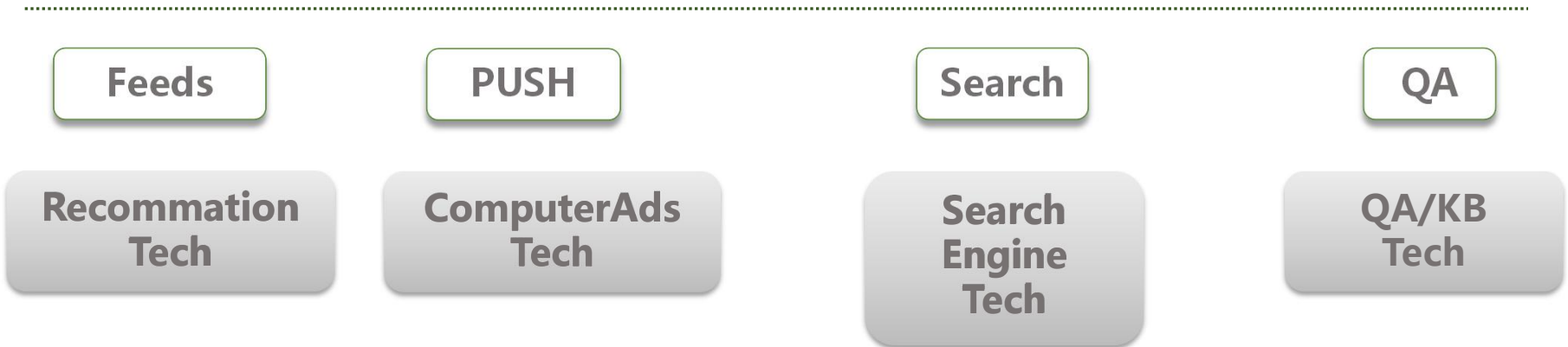




Content



USRE



PUSH



User

- How to Choose Push Time

Feeds

- How to Recommend Image

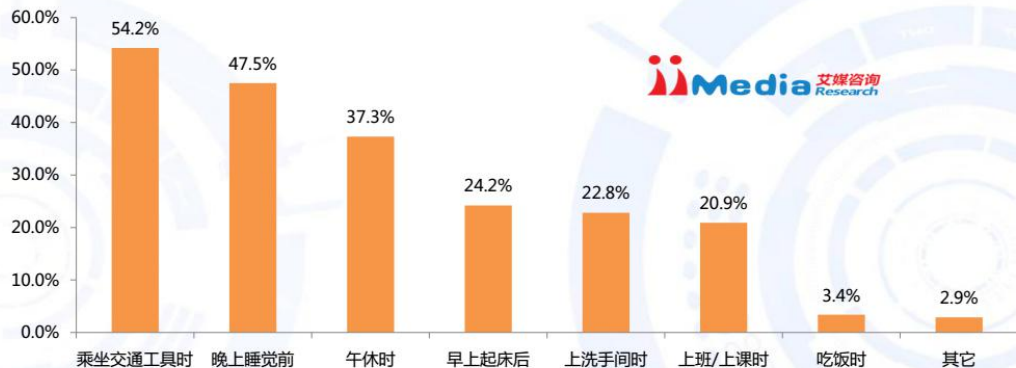
Content

- How to Estimate Article Rating
- How to Select Cover Picture



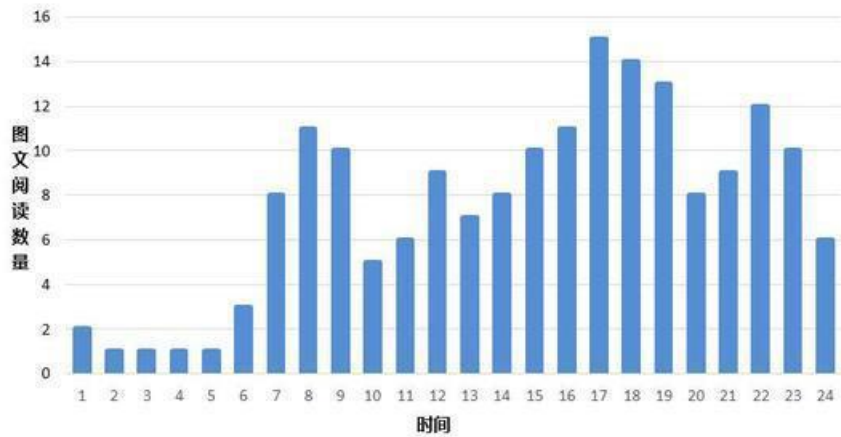
新浪新闻	20
网易新闻	3
虎扑体育	60
天天快报	6
搜狐新闻	2
腾讯新闻	1
一点资讯	8
今日头条	14
凤凰新闻	3

- 乘坐交通工具、睡觉前、午休时是**用户集中的“无聊时间”**，但又不能进行其他娱乐活动，是新闻客户端使用的高峰期。



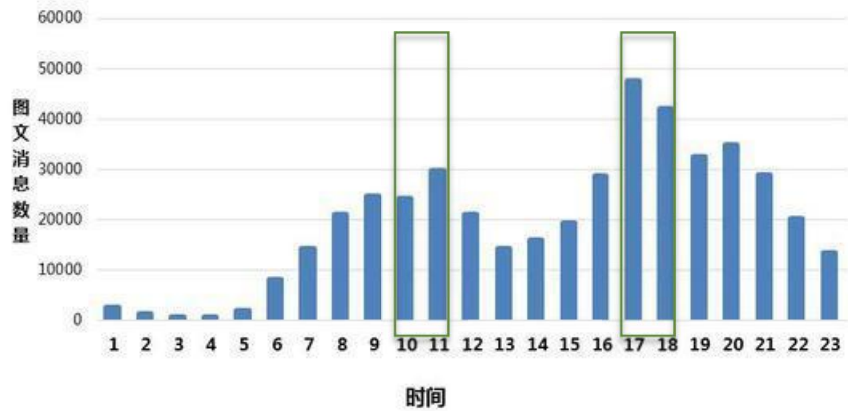
达观数据

公众号用户阅读时间分布



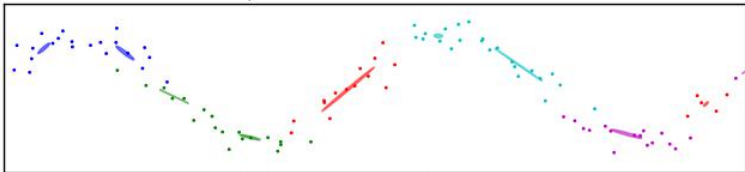
USRE

公众号每日推送时间分布

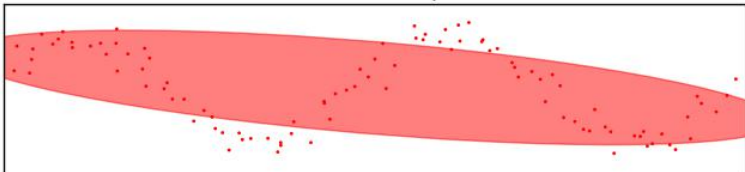


$$(v_i \mid \tilde{\mu}_i) \sim N(\tilde{\mu}_i, \sigma^2)$$
$$\tilde{\mu}_i \sim G$$
$$G \sim \text{DP}(H(\lambda), \alpha)$$

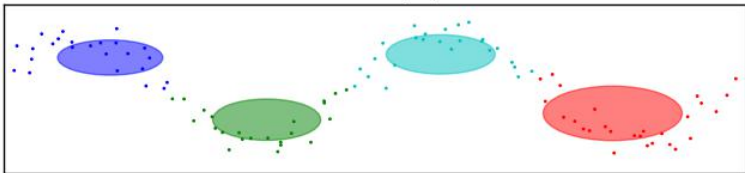
Expectation-maximization



Dirichlet Process, alpha=0.01



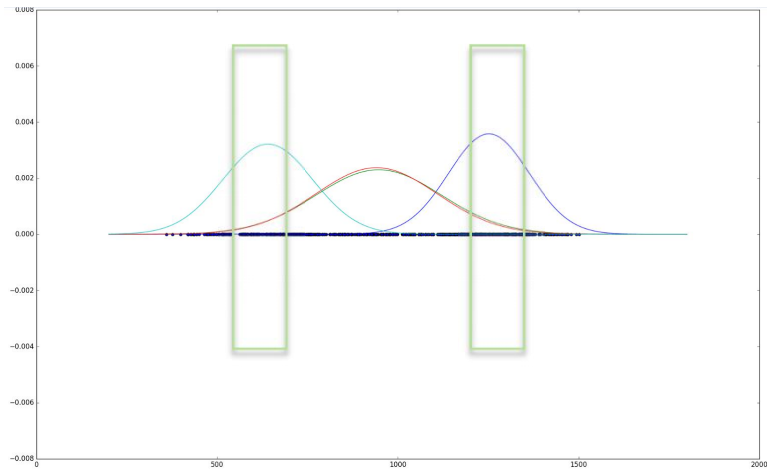
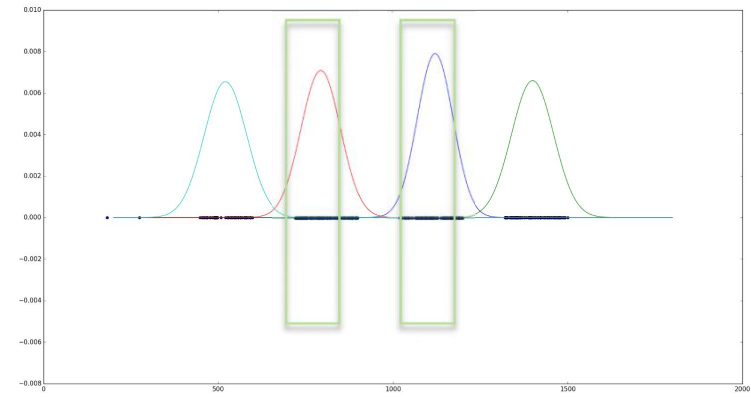
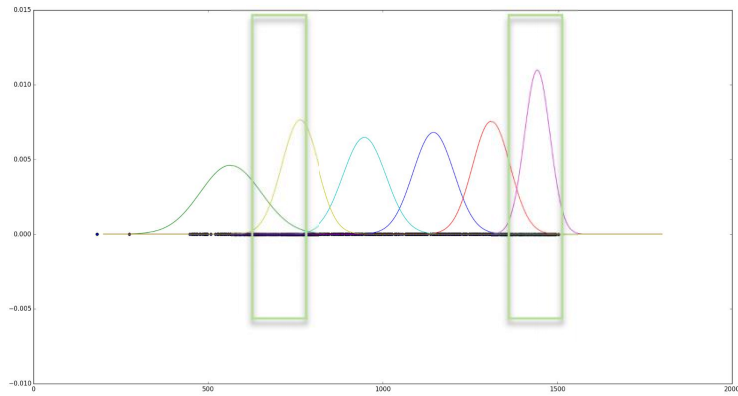
Dirichlet Process, alpha=100.



2016091906
2016091906
2016091917
2016091917
2016092006
2016092006
2016092009
2016092009
2016092012
2016092012
2016092013
2016092013
2016092015
2016092015
2016092106
2016092106
2016092109
2016092109
2016092205
2016092205
2016092209
2016092209
2016092217
2016092217

7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0
7	0

2016-09-19 06:29:55	109
2016-09-19 06:29:55	109
2016-09-19 17:15:58	109
2016-09-19 17:15:58	109
2016-09-20 06:08:48	109
2016-09-20 06:08:48	109
2016-09-20 09:02:31	109
2016-09-20 09:02:31	109
2016-09-20 12:13:21	109
2016-09-20 12:13:21	109
2016-09-20 13:35:27	109
2016-09-20 13:35:27	109
2016-09-20 15:16:47	109
2016-09-20 15:16:47	109
2016-09-21 06:37:12	109
2016-09-21 06:37:12	109
2016-09-21 09:16:23	109
2016-09-21 09:16:23	109
2016-09-22 05:49:13	109
2016-09-22 05:49:13	109
2016-09-22 09:30:07	109
2016-09-22 09:30:07	109
2016-09-22 17:54:43	109
2016-09-22 17:54:43	109



- D:1-5, 6-7; H:23-4
- OverSampling
- Top, Random

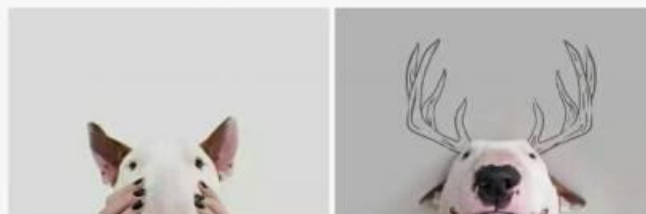
Content¹



- More than one Person
- Non Human



当牛头梗遇上这样的摄影师，
的节奏啊



唯美创意家居 1分钟前

大雪天就不要带短腿出来散步了，心疼！



宠上天

6评 (x)

Image Segmentation

主动轮廓线 Snake Model

Mean shift

Ncuts

Graph cut



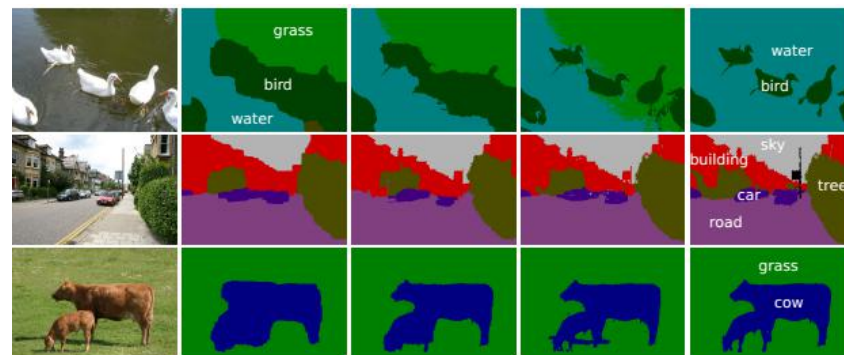
Semantic Segmentation

Hammersley-Clifford

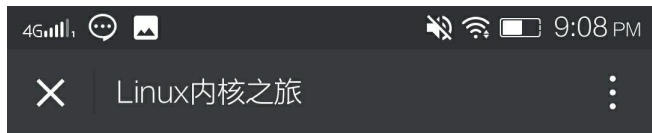
MRF & Gibbs Energy

$$P(X|I) = \frac{1}{Z(I)} \exp(-\sum_{c \in C_S} \phi_c(X_c | I))$$

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j)$$

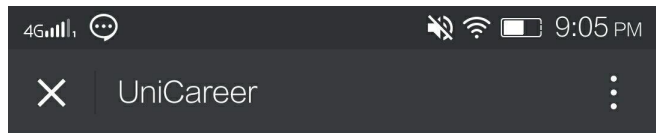


Content Modeling¹



很多操作系统教科书将正在CPU上执行的进程定义为RUNNING状态、而将可执行但是尚未被调度执行的进程定义为READY状态，这两种状态在linux下统一为 TASK_RUNNING状态。

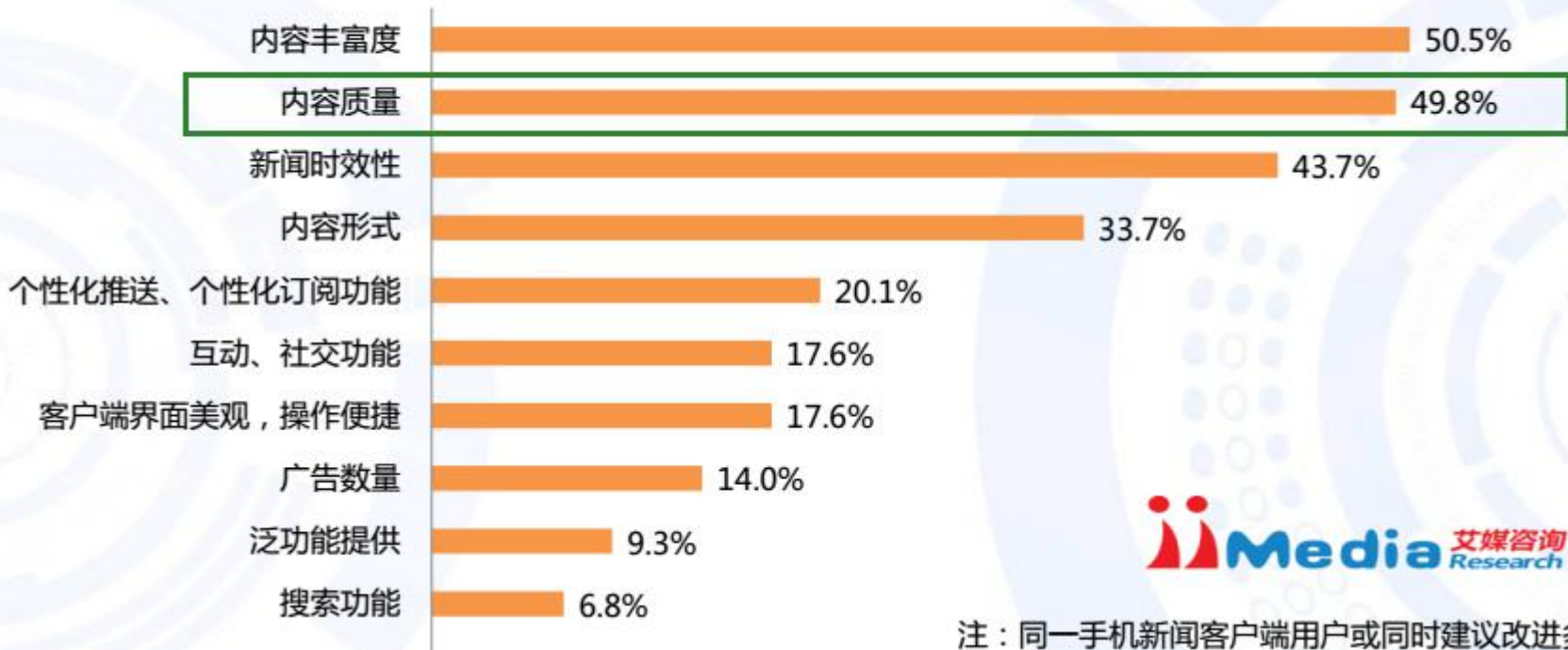
S(TASK_INTERRUPTIBLE)，可中断的睡眠状态。



当向量出现在防御象限时，意味着企业应该集中精力克服内部弱点并回避外部威胁，防御型战略包括紧缩、剥离、结业清算和集中多元化经营等。

当向量出现在竞争象限时，表明企业应该采取竞争性战略，包括后向一体化、前向一体化、市场渗透、市

2016上半年中国手机新闻客户端用户对产品的改进建议



注：同一手机新闻客户端用户或同时建议改进多个方面

Content Modeling²

**Content
Rating**



layout

Word
Count

Image
Rating

Topic
temper

pCTR

pCVR

Recomm

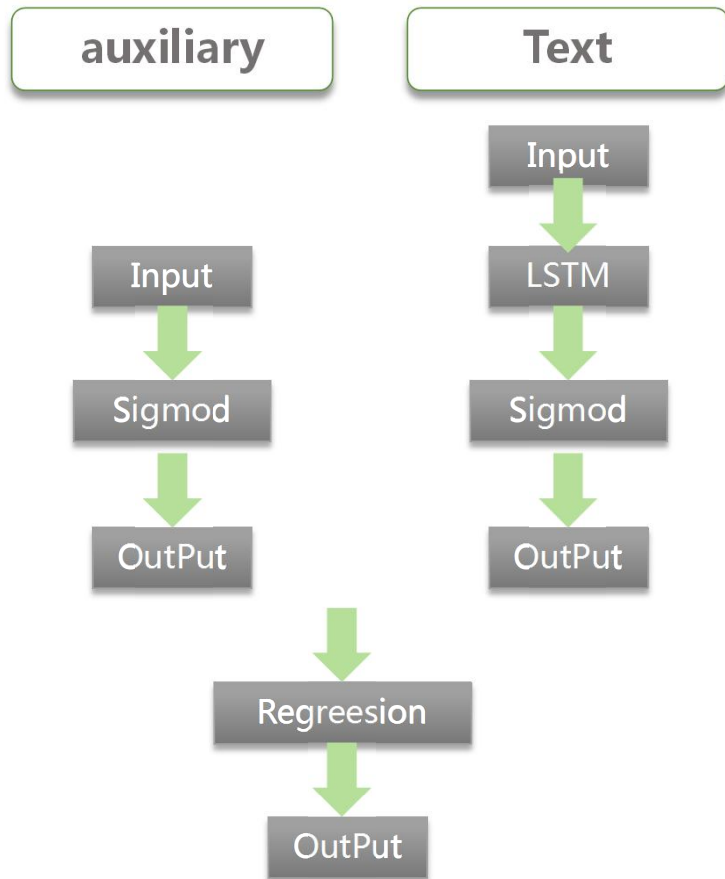
Content



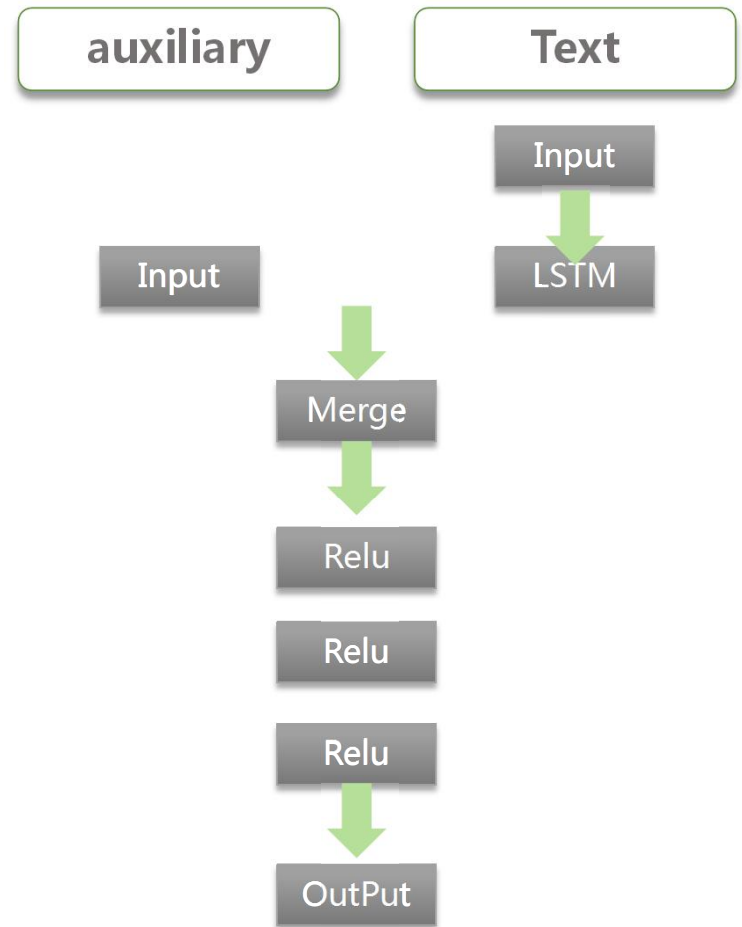
Read thoroughly
Detailed reading
Long stay on Page

- PV > 500
- Image -> [Image][size]
- Title -> [|]
- Word Embedding

Content Modeling³

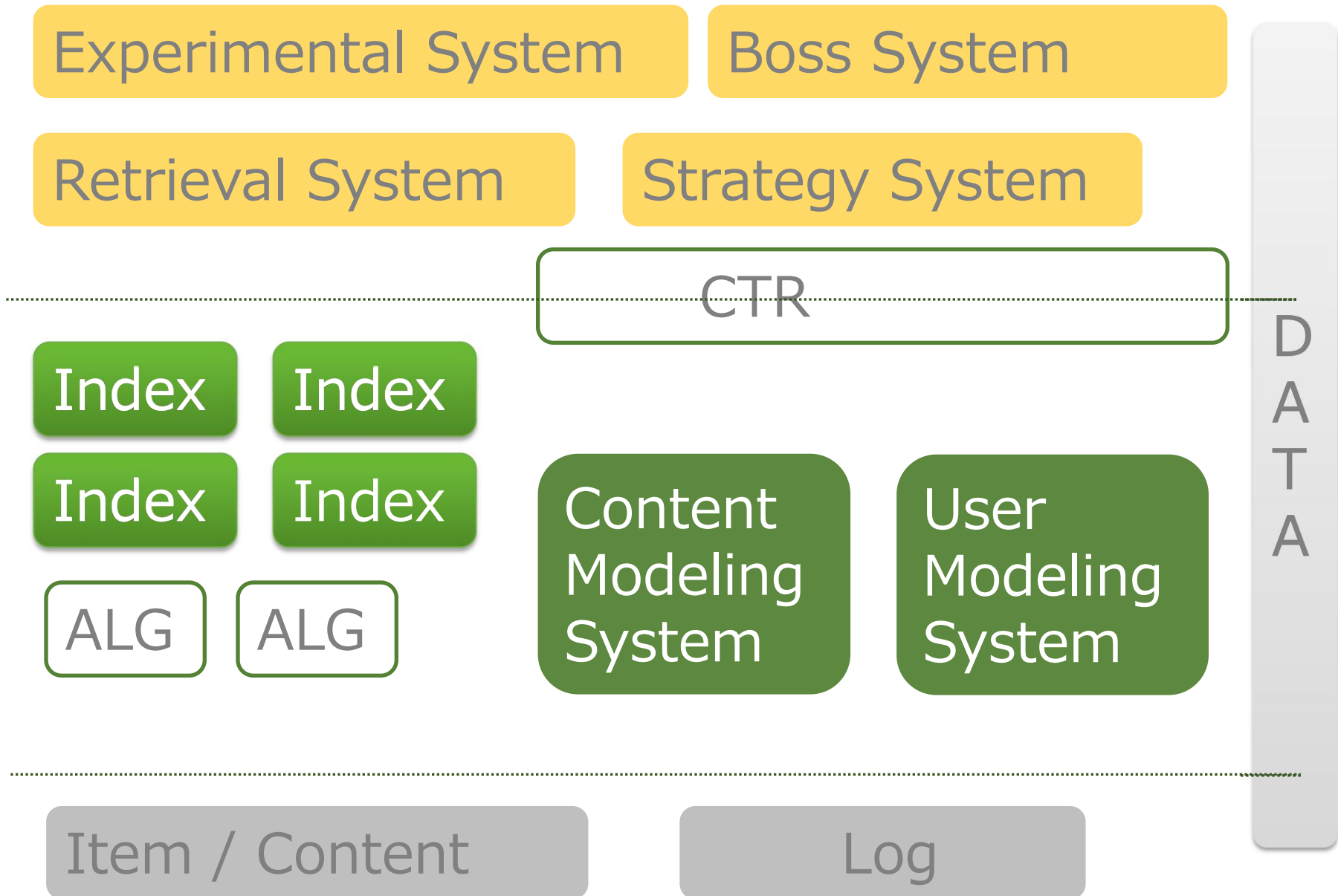


accuracy
67%



accuracy
74%

Image¹



Content Modeling

Similarity Recommendation

Text

Text
Feature

Photo

Image
Feature

Query

Word
Sentence
C G

SIFT/S
U F T
B I N

Cluster

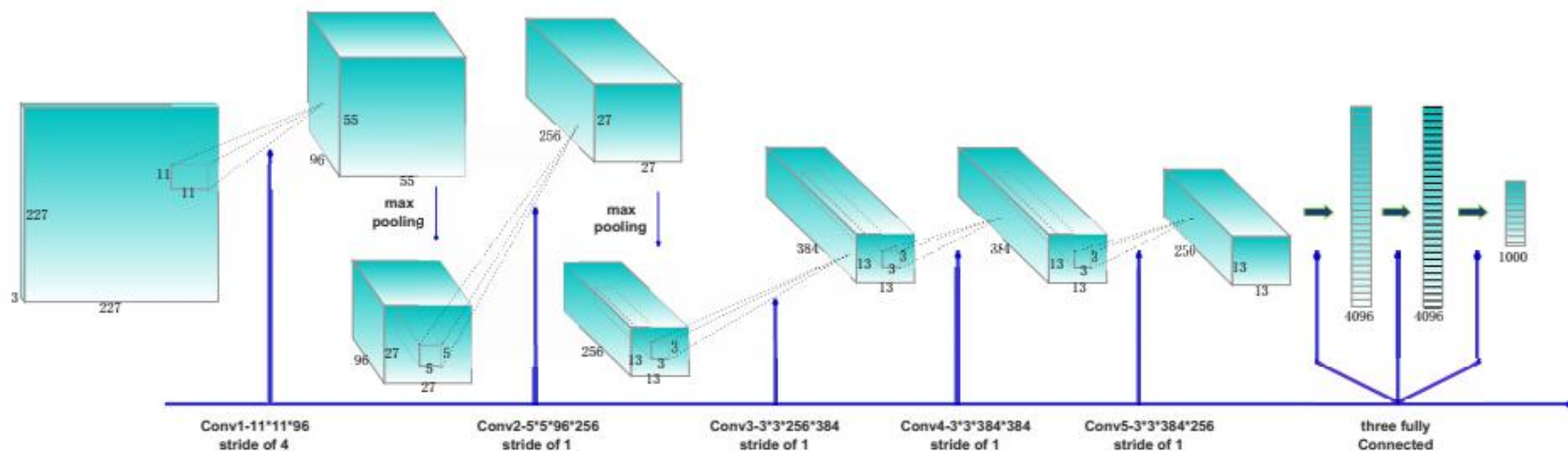
N Baye/D Tree

PLSA/LDA

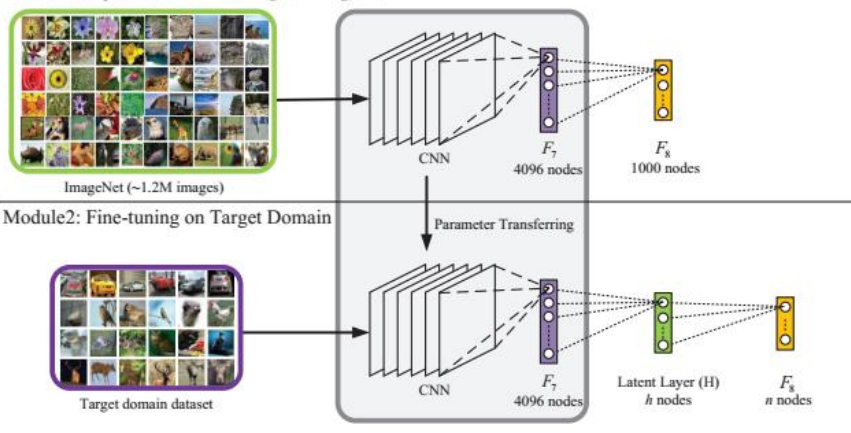
Classifcaton

Bag of Words

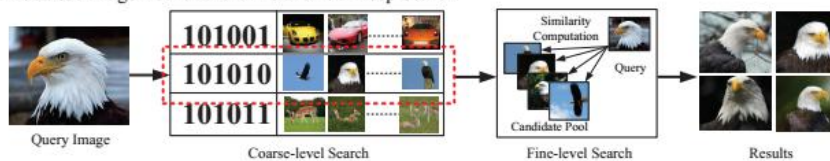
Content Modeling



Module1: Supervised Pre-Training on ImageNet

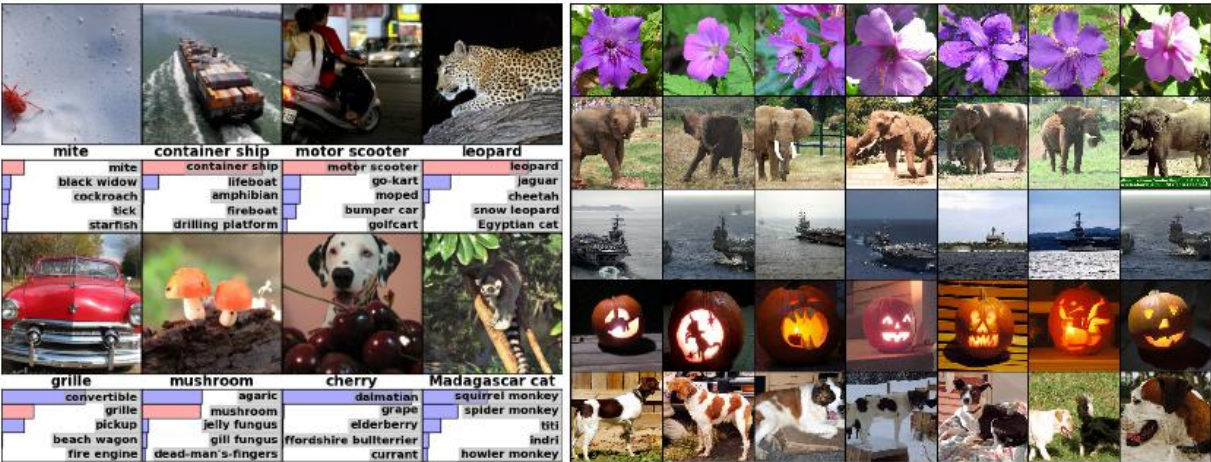


Module3: Image Retrieval via Hierarchical Deep Search



Content Modeling

Neural art



C N N

Object

C N N

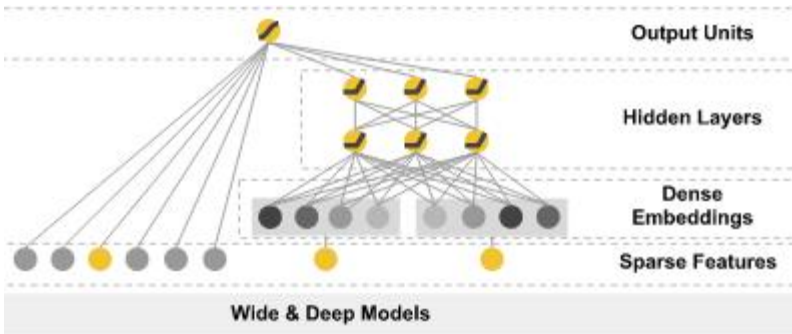
Style

W i d e & D e e p

Text+Image

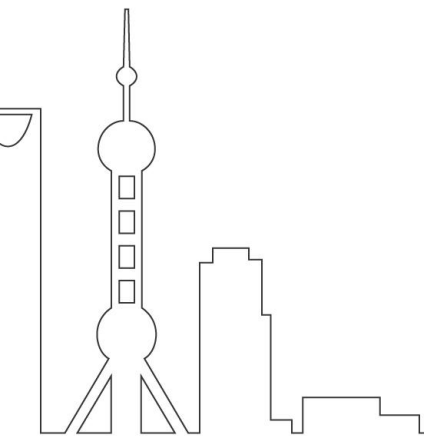
L R

Story



Content Modeling

	<p>[r1]人间圣地，</p> <p>10月22日</p>		<p>[r1]大厨的秘密 解馋好吃</p> <p>10月22日</p>		<p>[r1]旧影集07:1899年的上海.1942年的德军.</p> <p>10月22日</p>
	<p>[r1]荒郊野岭惊</p> <p>10月22日</p>		<p>[r4]又香又辣</p> <p>10月15日</p>		<p>[r1]近200年来世界历史上的经典照片，最后一张最可笑</p> <p>10月12日</p>
	<p>[r1]空中俯瞰宁</p> <p>10月22日</p>		<p>[r3]麻辣鲜香</p> <p>9月9日</p>		<p>[r1]世界历史上的九大军事将领，中国的这位让世人憎恨</p> <p>9月28日</p>
	<p>[r1]青山新江滩</p> <p>10月22日</p>		<p>[r3]中原人特 到嘴里直流泪</p> <p>10月22日</p>		<p>[r3]除了好色和当皇帝，袁世凯还做了些什么</p> <p>7月26日</p>
	<p>[r1]朝鲜的金刚 景的雄奇险秀</p> <p>10月22日</p>		<p>[r3]排骨您这 钙助长高</p> <p>9月20日</p>		<p>[r4]老照片回顾 老上海百年前的十里洋场</p> <p>10月15日</p>
	<p>[r1]旅行中这些 吗?</p> <p>10月22日</p>		<p>[r4]鲜虾加入 都爱吃!</p> <p>10月16日</p>		<p>[r3]上甘岭一战成名，曾身兼党政军数职，毛主席提名他做中央副主席。</p> <p>7月14日</p>



Thanks!

International Software Development Conference