

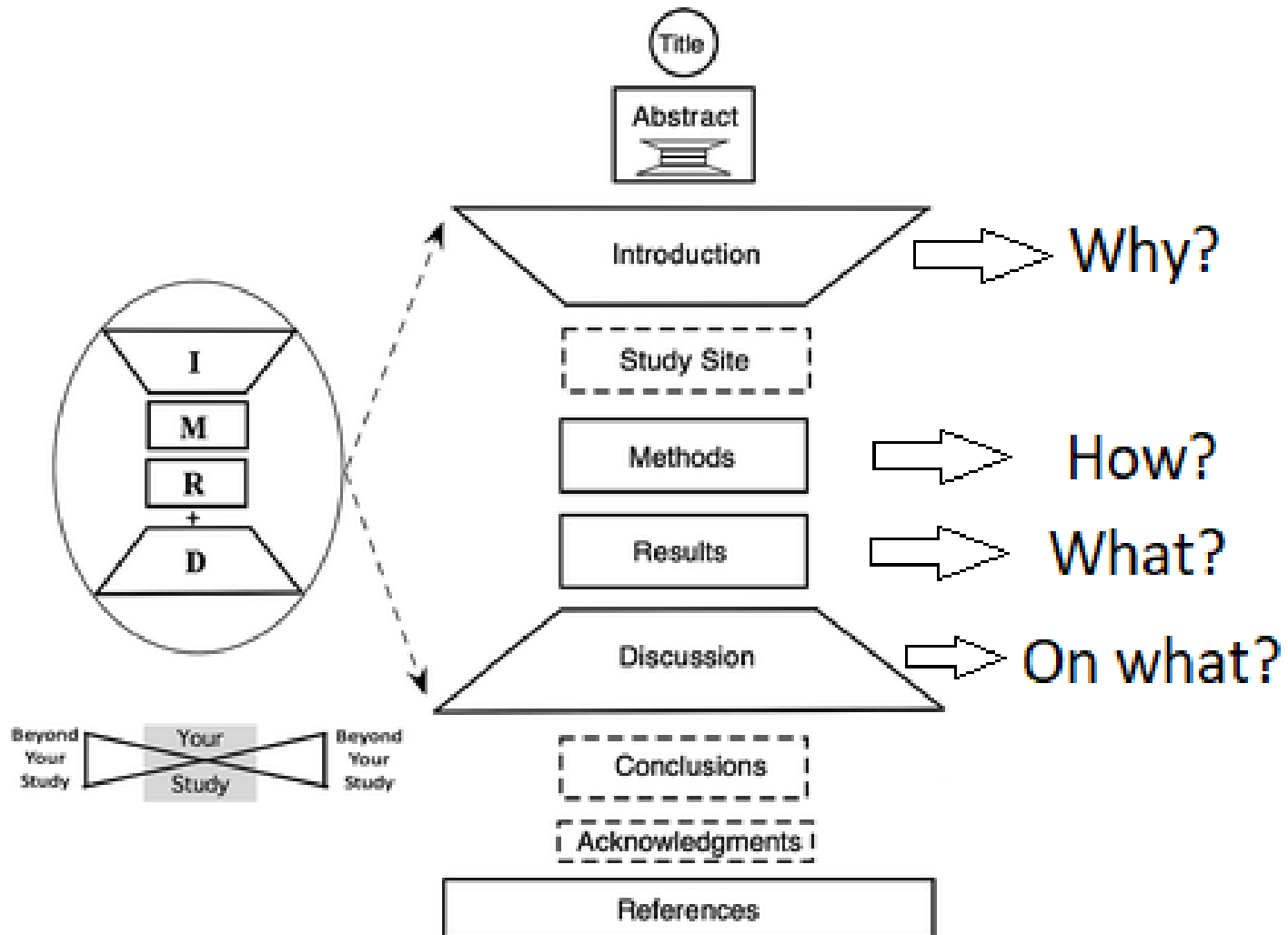
Научная статья

Лекция 2

Структура исследовательской статьи

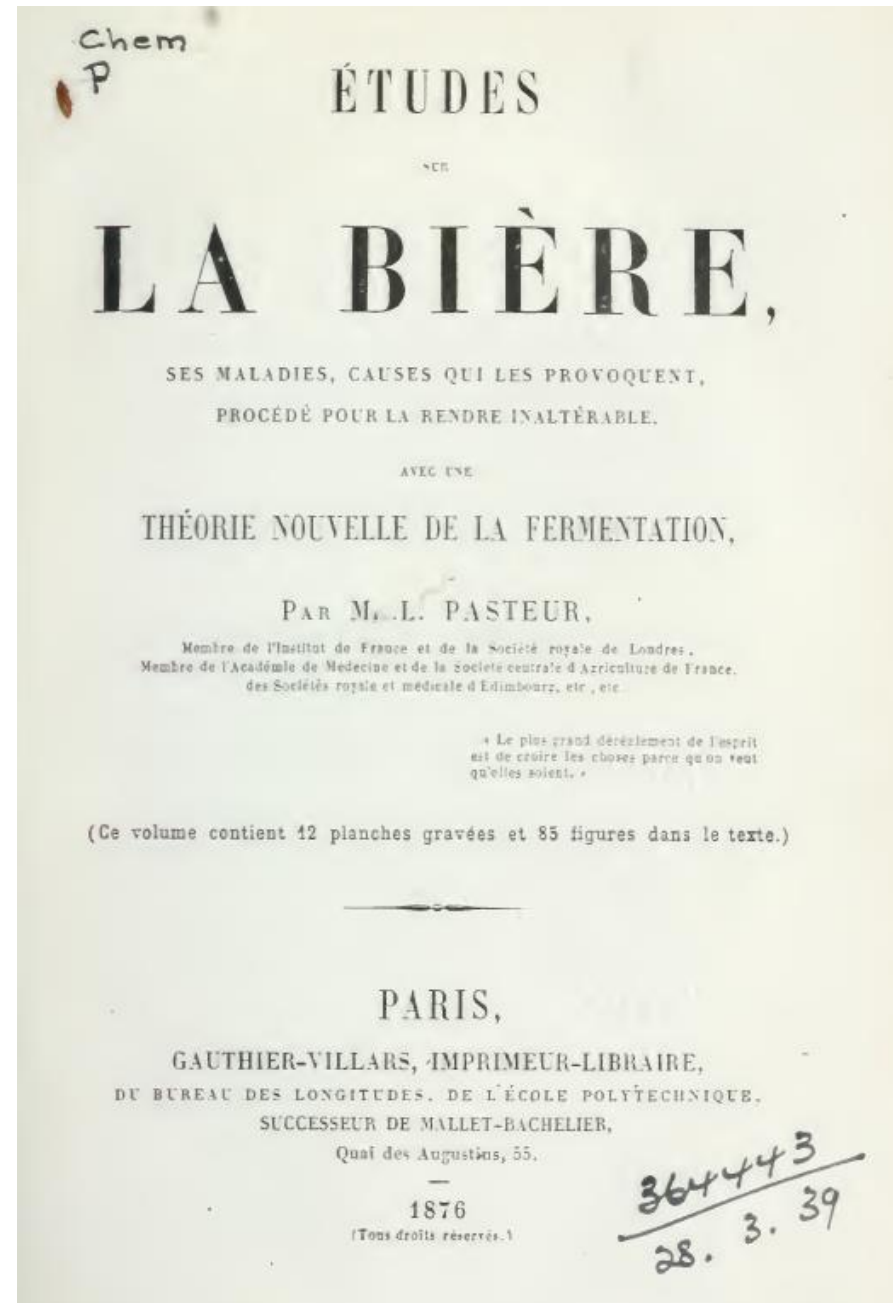
- Название (Title)
- Авторы (Authors)
- Аннотация (Abstract)
- Ключевые слова (Keywords)
- Введение (Introduction)
- Обзор предыдущих работ (Previous Work, Related Work)
- Методы (теоретические основы) (Materials and Methods)
- Результаты (Эксперименты) (Results, Experiments)
- Обсуждение (Discussion)
- Заключение (Conclusion)
- Благодарности (Acknowledgements)
- Список литературы (References)
- Приложения (Appendixes)

IMRAD – Introduction, Methods, Results, and Discussion



IMRAD

- Первая научная работа в стиле IMRAD:
Louis Pasteur
– Etudes sur la Biere (1876)
- 1972 – стандарт подготовки научных статей и устных докладов ANSI Z39.16-1972
- Структура для оригинальных исследований, не для обзорных статей



Примеры

Pennington J., Socher R., Manning C.D. GloVe: Global Vectors for Word Representation (2014)

- Conference proceedings (12 pages, 21 396 citations Scopus)
- Structure:
 - Introduction
 - Related Work
 - The GloVe Model
 - Experiments (results and discussion)
 - Conclusion

Примеры

Luong M.-T., Pham H., Manning C.D. Effective Approaches to Attention-based Neural Machine Translation (2015)

- Conference proceedings (10 pages, 3 540 citations Scopus)
- Structure:
 - Introduction (+related work)
 - Neural Machine Translation
 - Attention-based Models
 - Experiments (results)
 - Analysis (discussion)
 - Conclusion

Примеры

Prabowo R., Thelwall M. Sentiment analysis: A combined approach (2009)

- Journal of Informetrics (15 pages, 540 citations Scopus)
- Structure:
 - Introduction
 - Automatic document classification
 - Existing work in sentiment analysis
 - Different classification approaches used
 - Experiment (results + discussion)
 - Conclusion

Примеры

Zhang L., Wang S., Liu B. Deep learning for sentiment analysis: A survey (2018)

- Survey article!
- Data Mining and Knowledge Discovery journal (25 pages, 942 citations Scopus)
- Structure:
 - Introduction
 - Neural Networks
 - Deep Learning
 - Word Embedding
 - Autoencoder and denoising autoencoder
 - Convolutional Neural Network
 - Recurrent Neural Network
 - Long Short-Term Memory Network
 - Attention Mechanism with RNN
 - Memory Network
 - Recursive Neural Network
 - Sentiment Analysis Tasks
 - Document-Level Sentiment Classification
 - Sentence-Level Sentiment Classification
 - Aspect-Level Sentiment Classification
 - Aspect Extraction and Categorization
 - Opinion Expression Extraction
 - Sentiment Composition
 - Opinion Holder Extraction
 - Temporal Opinion Mining
 - Sentiment Analysis with Word Embedding
 - Sarcasm Analysis
 - Emotion Analysis
 - Multimodal Data for Sentiment Analysis
 - Resource-poor Language and Multilingual Sentiment Analysis
 - Other Related Tasks
 - Conclusion

Название статьи

- Цель – привлечение внимания
- Наполненность, значимость и сочетаемость
- Название не должно быть слишком длинным или слишком коротким (3-15 слов)
 - Короткое название: «Studies of Brucella»
 - Длинное название: «On the addition to the method of microscopic research by a new way of producing colour-contrast between an object and its background or between definite parts of the object itself» (Rheinberg, J.R. Microsc. Soc. 1896:373).

Название статьи

Не использовать «мусорные» слова:

- “Studies on ...” (Изучение ...)
- “Observations on ...” (Наблюдение ...)
- “On the ...” (К вопросу о ...)
- “A study of ...” (Изучение ...)
- “Research on ...” (Исследование ...)
- “Report on ...” (Отчет о ...)
- “Regarding ...” (К вопросу о ...)
- “Use of ...” (Использование ...)

Название статьи

Пример:

- «Classification of images»
- «Research on classification of certain images with several machine learning methods»
- «ImageNet classification with neural networks»
- «ImageNet classification with deep convolutional neural networks»

Название статьи

- Название статьи – это комбинация из наименьшего количества слов, которая адекватно описывает содержание статьи
 - «Всё следует упрощать до тех пор, пока это возможно, но не более того» (А. Эйнштейн)
- Как правило, название статьи не является полным предложением (повествовательным или вопросительным)
 - Но есть исключения: “Do climatic attractors exist?” (Nature, 1986)
- Название придумывается в конце
 - Но имеет смысл сформулировать рабочее название

Авторы

- Включать ли руководителя?
- Есть раздел «Благодарности»
- Место работы (Affiliation)
- Автор для переписки (Corresponding author)

Аннотация (Abstract)

- Цель – отразить содержание и заинтересовать читателя
- Аннотация – мини-версия статьи и включает все основные разделы в сокращенном виде:
 - цель исследования
 - методы
 - результаты
 - выводы
- Нет ссылок на источники (как правило)
 - Часто бывают ссылки на исходный код (Xu et al., 2021)
- Обычно 150-250 слов

Ключевые слова (Keywords)

- Цель – облегчить поиск релевантных статей
- Ключевые слова должны дополнять, расширять и перефразировать название статьи, но при этом оставаться специфичными ее содержанию

Введение

- Во Введении содержится информация, которая позволит читателю понять и оценить результаты представленного исследования без дополнительного обращения к другим литературным источникам
- Также должно содержаться обоснование необходимости и актуальности исследования
- Информация должна быть организована по принципу «от общего к частному»

Введение – структура

- Описание темы, задачи, предметной области, с которой связано исследование или установление научного контекста (о чем речь?)
- Актуальность темы (применение, почему решение задачи важно?)
- Обзор: сейчас задача решается таким образом... (краткий обзор литературы, связанной с исследованием)
- Проблема: описание белых пятен в задаче или того, что еще не сделано (пробелы, недостатки)
- Формулировка цели (и, возможно, задач) исследования
- Вклад исследования и оценка его важности
 - Краткое описание решения
- Описание структуры публикации

Введение — контекст и актуальность

- Описание более широкой проблемы
- Описание темы, задачи, предметной области
- Определения
- Актуальность: где применяется решение задачи?
 - Приложения
 - Области
- См. примеры статей:
 - [Prabowo, Thelwall, 2009]
 - [Pennington et al., 2014]

Введение – обзор

- Как задача решается?
 - Краткий обзор литературы
 - Ссылки на публикации, связанные с исследованием
 - При наличии отдельного раздела «Обзор предыдущих работ», этот пункт может быть небольшим
 - Подробнее про обзор – ниже
- Примеры:
 - [Luong et al., 2015]
 - [Xu et al., 2021]

Введение – проблема исследования

- Описание белых пятен в задаче или того, что еще не сделано (пробелы, недостатки)
- Примеры:
 - [Xu et al., 2021]
 - [Pennington et al., 2014]

Введение – вклад

- Необходимо явно выделить вклад в науку данного исследования
- Формула научной новизны:
 - основано на...
 - отличается от существующих...
 - позволяет...
- Примеры:
 - Krishna K., Wieting J., Iyyer M. Reformulating Unsupervised Style Transfer as Paraphrase Generation (2020)
 - Peyrard M. A Simple Theoretical Model of Importance for Summarization (2019)

Постановка задачи

- Где?
 - «Введение»?
 - Отдельный раздел?
 - «Методы»?
- Примеры:
 - Sebastiani F. Machine Learning in Automated Text Categorization // ACM Computing Surveys. 2002. Vol. 43(1). P. 1–47.
 - Pp. 2–5
 - Liu B. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
 - Pp. 17–21

Обзор предыдущих работ – подходы

- Основные подходы к решению задачи
 - Классификация
 - Преимущества и недостатки
- Примеры:
 - Hemmatian F., Sohrabi M.K. A survey on classification techniques for opinion mining and sentiment analysis
// Artificial Intelligence Review. 2017. Vol. 52(3). P. 1495–1545.
 - Pp. 12, 38–39
 - Prabowo R., Thelwall M. Sentiment analysis:
A combined approach // Journal of Informetrics. 2009. Vol. 3(2). P. 143–157.
 - P. 144

Обзор предыдущих работ – методы

- Описание метода:
 - История
 - Математическое описание
 - Временная сложность
 - Примеры применения (ссылки на статьи)
- Пример:
 - Hemmatian F., Sohrabi M.K. A survey on classification techniques for opinion mining and sentiment analysis // Artificial Intelligence Review. 2017. Vol. 52(3). P. 1495–1545.
 - P. 17 – K-nearest neighbor (KNN)

Ссылки на источники

Авторитетные:

- рецензируемые издания:
 - монографии, учебники (?), энциклопедии
- рецензируемые научные статьи:
 - журналы, материалы конференций
- веб-сайты производителей

Неавторитетные:

- Википедия
- Научно-популярные статьи
- Блоги (?), форумы, социальные сети

Ссылки на источники

- Если описывается алгоритм, то нужно ссылаться на оригинальную статью, а не на пересказ
- Если в источнике более 20-30 стр. – указывать страницу

Ссылки на источники

- Оформление – множество стилей
 - APA, IEEE, Chicago, ...
 - ГОСТ 7.1-2003
 - <https://en.wikipedia.org/wiki/Citation>
 - <https://www.scribbr.com/citing-sources/citation-styles/>
 - Определяется требованиями издания
- Виды внутритекстовых ссылок:
 - Текст¹
 - Текст [1], Текст (1)
 - Текст [автор, год], Текст (автор, год)

Ссылки на источники

- 1 автор:
 - [Иванов, 2012] → [1]
 - [Иванов, 2012, с. 100] → [1, с. 100]
 - [Иванов, 2012, р. 100] → [1, р. 100]
- 2 автора:
 - [Иванов, Петров, 2011] → [5]
- 3 и более авторов:
 - [Иванов и др., 2014] → [6]
 - [Ivanov et al., 2014] → [7]
- Перекрестные ссылки в Word

Описание метода

План рассмотрения метода:

- История
- Математическое описание
- Алгоритм
 - Блок-схема
 - Псевдокод
- Временная сложность
- Примеры
- Применение (ссылки на статьи)

Математическое описание

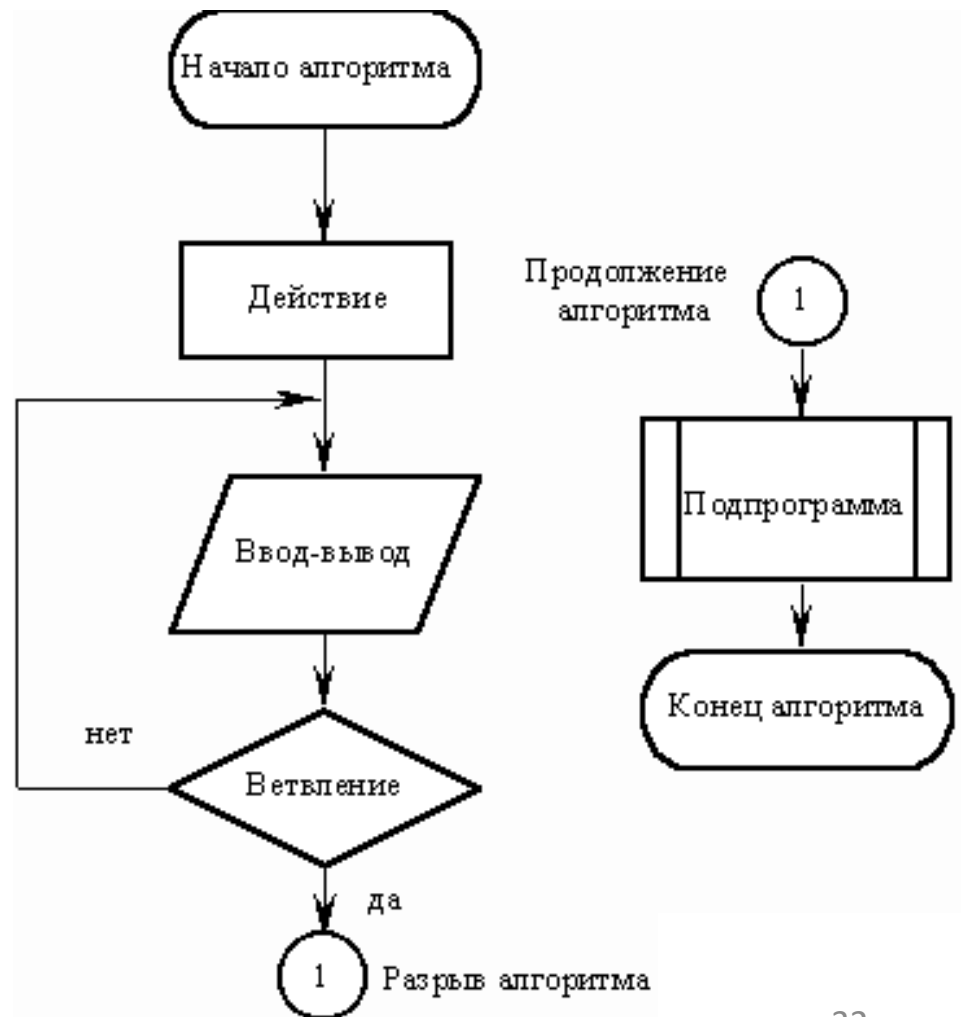
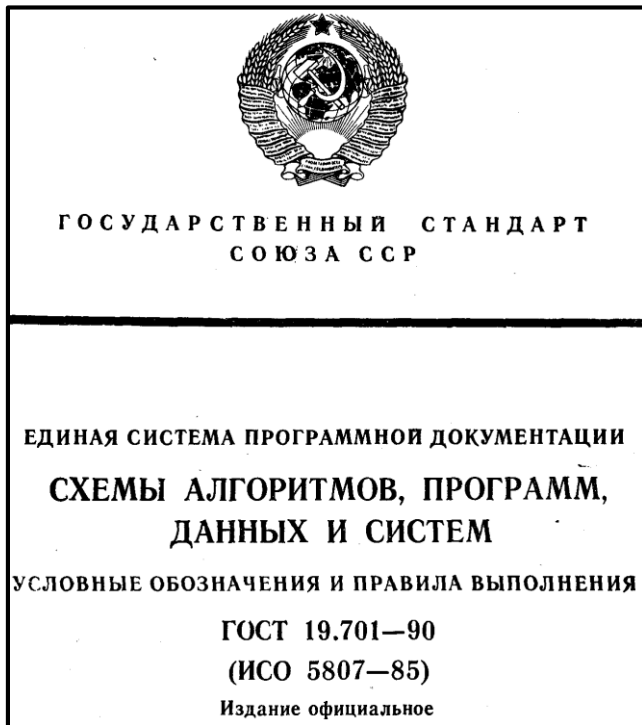
- Ввести все обозначения
- Строгость и корректность
- Нумерация формул

Алгоритмы

- Блок-схемы (ГОСТ, UML)
- Псевдокод (нет стандартов)

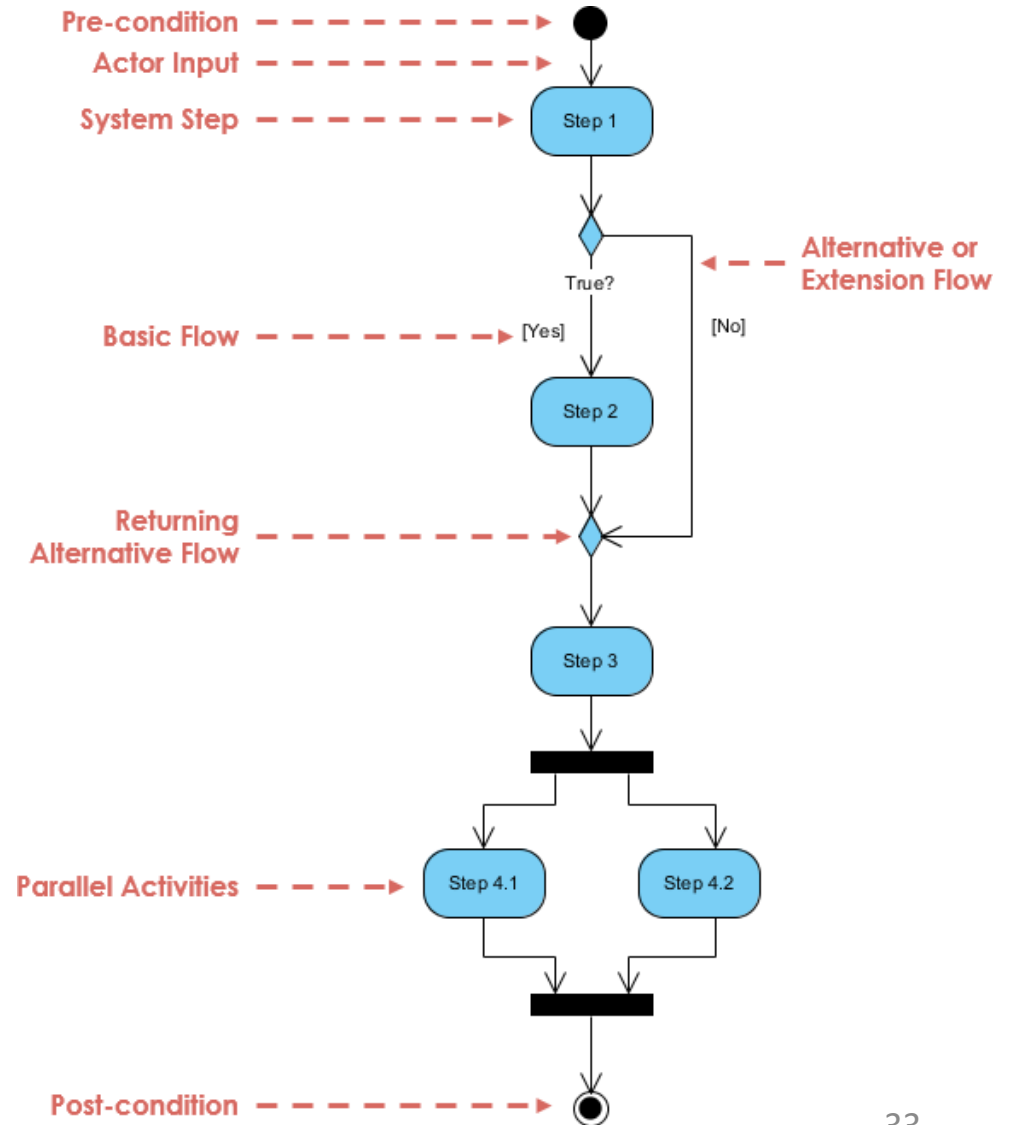
Алгоритм – блок-схемы

- ГОСТ 19.701-90



Алгоритм – UML-диаграммы

- Диаграмма деятельности (activity diagram)



Псевдокод: [Ахо, Хопкрофт, Ульман, 2003]

```
procedure INSERT (x: elementtype; p: position; var L: LIST );  
  { INSERT вставляет элемент x в позицию p в списке L }  
  var  
    q: position;  
  begin  
    if L.last >= maxlength then  
      error('Список полон')  
    else if (p > L.last + 1) or (p < 1) then  
      error('Такой позиции не существует')  
    else begin  
      for q:= L.last downto p do  
        { перемещение элементов из позиций p, p+1, ... на  
          одну позицию к концу списка }  
        L.elements[q+1] := L.elements[q];  
      L.last := L.last + 1;  
      L.elements[p] := x  
    end  
  end; { INSERT }
```

Псевдокод:

[Кормен, Лейзерсон, Ривест, Штайн, 2013]

INSERTION-SORT(A)

```
1  for  $j = 2$  to  $A.length$ 
2       $key = A[j]$ 
3      // Вставка  $A[j]$  в отсортированную
        последовательность  $A[1..j-1]$ .
4       $i = j - 1$ 
5      while  $i > 0$  и  $A[i] > key$ 
6           $A[i+1] = A[i]$ 
7           $i = i - 1$ 
8       $A[i+1] = key$ 
```

Псевдокод:

[Dasgupta, Papadimitriou, Vazirani, 2006]

Figure 4.3 Breadth-first search.

procedure `bfs`(G, s)

Input: Graph $G = (V, E)$, directed or undirected; vertex $s \in V$

Output: For all vertices u reachable from s , `dist`(u) is set to the distance from s to u .

for all $u \in V$:

`dist`(u) = ∞

`dist`(s) = 0

$Q = [s]$ (queue containing just s)

while Q is not empty:

$u = \text{eject}(Q)$

 for all edges $(u, v) \in E$:

 if `dist`(v) = ∞ :

`inject`(Q, v)

`dist`(v) = `dist`(u) + 1

Псевдокод:

[Дасгупта, Пападимитриу, Вазирани, 2014]

Рис. 4.3. Поиск в ширину.

процедура $\text{BFS}(G, s)$

{Вход: граф $G(V, E)$, вершина $s \in V$.}

{Выход: для всех вершин u , достижимых из s ,
 $\text{dist}[u]$ будет равно расстоянию от s до u .}

для всех вершин $u \in V$:

$\text{dist}[u] = \infty$

$\text{dist}[s] = 0$

$Q = \{s\}$ {очередь из одного элемента}

пока Q не пусто:

$u = \text{EJECT}(Q)$

 для всех рёбер $(u, v) \in E$:

 если $\text{dist}[v] = \infty$:

$\text{INJECT}(Q, v)$

$\text{dist}[v] = \text{dist}[u] + 1$

Псевдокод: [Flach, 2012]

Algorithm 11.3: $\text{Boosting}(D, T, \mathcal{A})$ – train an ensemble of binary classifiers from reweighted training sets.

Input : data set D ; ensemble size T ; learning algorithm \mathcal{A} .

Output : weighted ensemble of models.

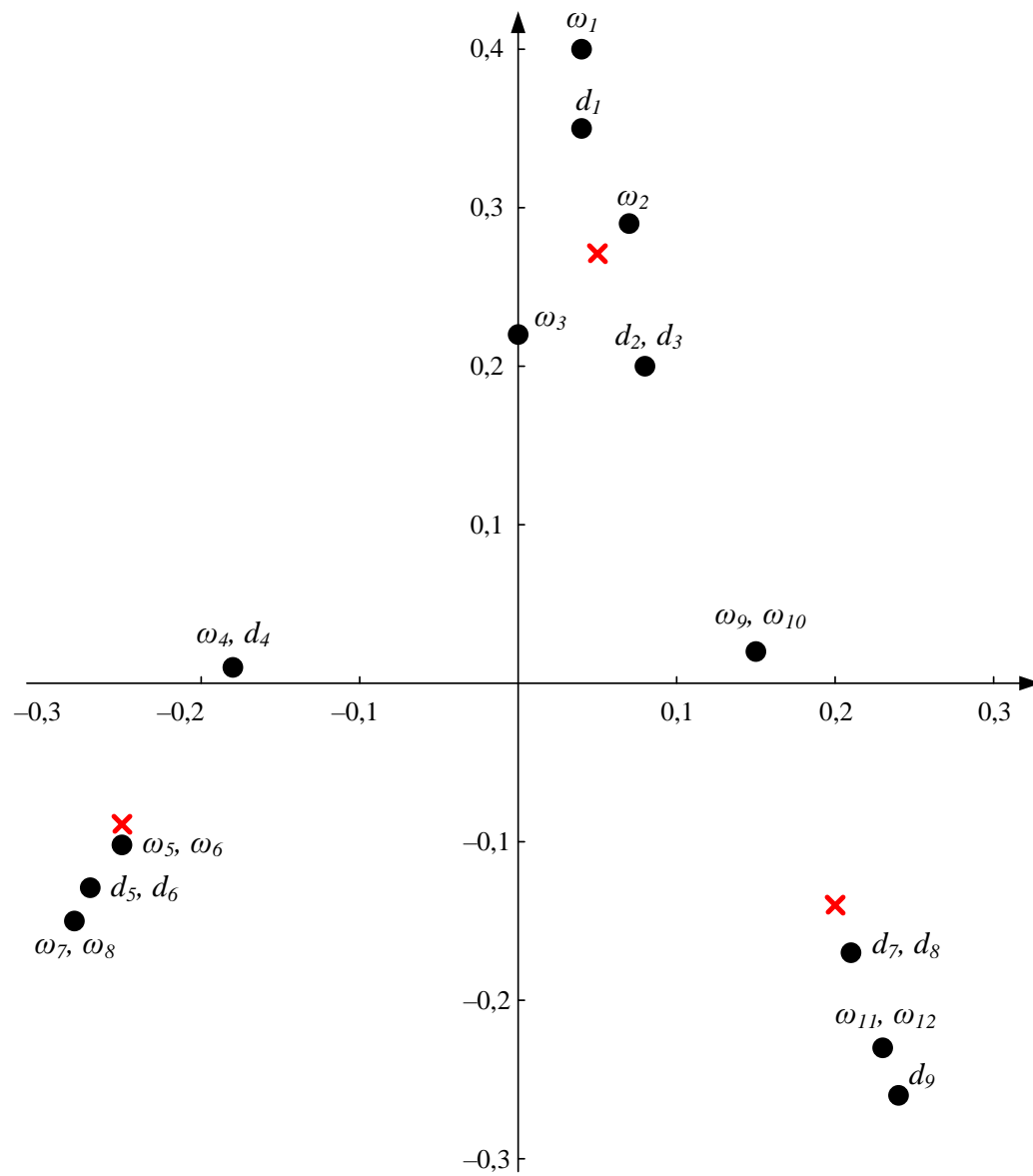
```
1  $w_{1i} \leftarrow 1/|D|$  for all  $x_i \in D$  ; // start with uniform weights
2 for  $t = 1$  to  $T$  do
3   run  $\mathcal{A}$  on  $D$  with weights  $w_{ti}$  to produce a model  $M_t$ ;
4   calculate weighted error  $\epsilon_t$ ;
5   if  $\epsilon_t \geq 1/2$  then
6     | set  $T \leftarrow t - 1$  and break
7   end
8    $\alpha_t \leftarrow \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$  ; // confidence for this model
9    $w_{(t+1)i} \leftarrow \frac{w_{ti}}{2\epsilon_t}$  for misclassified instances  $x_i \in D$  ; // increase weight
10   $w_{(t+1)j} \leftarrow \frac{w_{tj}}{2(1-\epsilon_t)}$  for correctly classified instances  $x_j \in D$  ; // decrease weight
11 end
12 return  $M(x) = \sum_{t=1}^T \alpha_t M_t(x)$ 
```

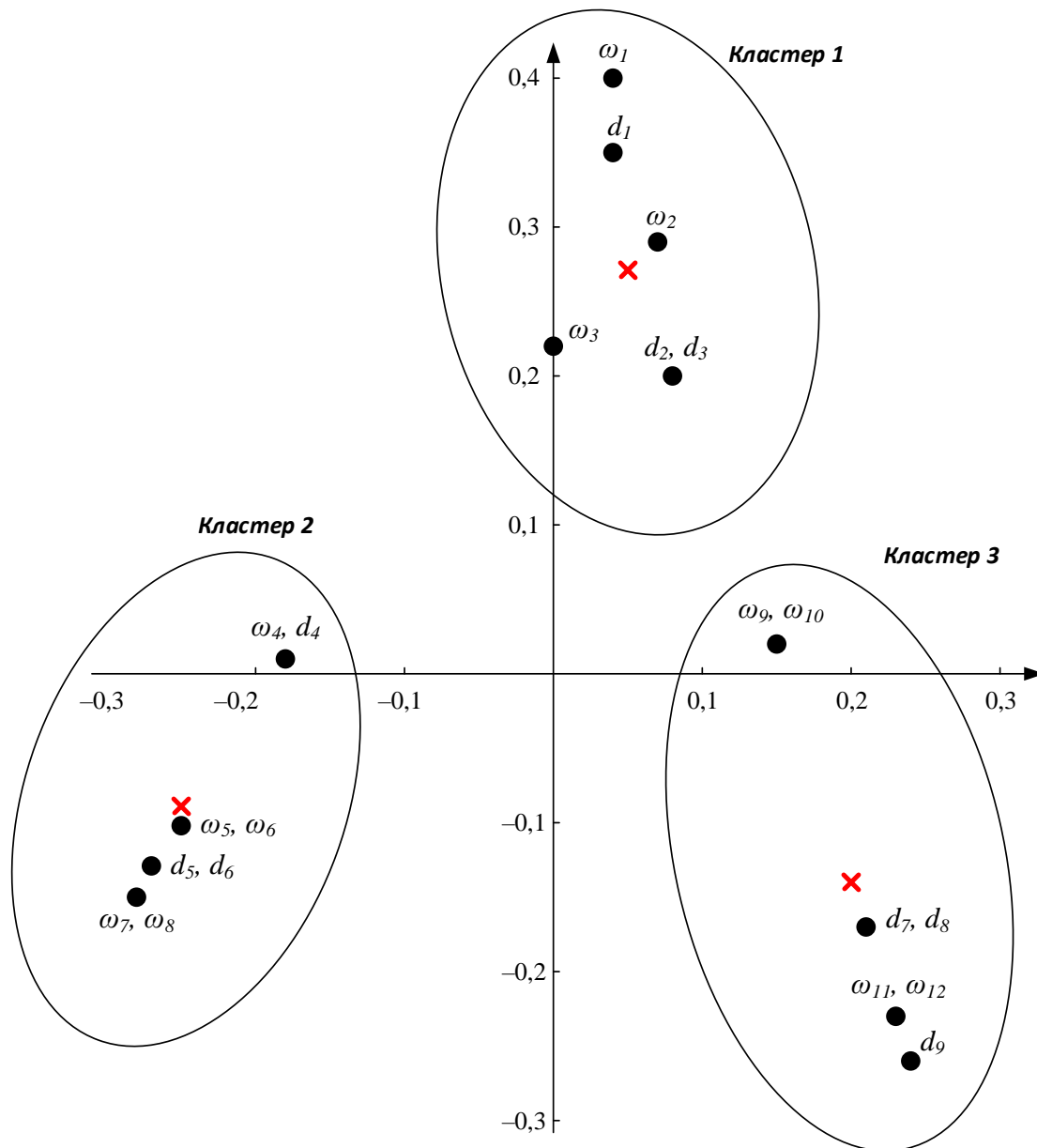
Временная сложность

- $O(C)$ – константная
- $O(\log n)$ – логарифмическая
- $O(n)$ – линейная
- $O(n^2)$ – квадратичная
- $O(n^p)$ – полиномиальная степени p
- $O(2^n)$ – экспоненциальная
- ...

Примеры

- Простые
- С картинками
- В динамике





Эксперименты и обсуждение

План раздела «Эксперименты»:

- описание данных
 - предобработка
- методология проведения эксперимента
 - методы для сравнения (baselines)
 - меры качества
- реализация (инструменты)
 - параметры моделей/инструментов
- результаты
 - таблицы и/или графики
 - выделение лучших
 - проверка статистической значимости
- анализ результатов
 - часто объединяется с обсуждением результатов
 - анализ ошибок (примеры)
 - ablation study

Эксперименты и обсуждение

Примеры:

- Pang B., Lee L., Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques (2002)
- Prabowo R., Thelwall M. Sentiment Analysis: a Combined Approach (2009)
- Pennington J., Socher R., Manning C.D. GloVe: Global Vectors for Word Representation (2014)
- Peyrard M. A Simple Theoretical Model of Importance for Summarization (2019)

Рецензирование: критерии

- Соответствие профилю издания
- Новизна, вклад и значимость
- Формулировка проблемы исследования
- Качество обзора предыдущих работ
- Качество методологии
- Качество эмпирического исследования
- Качество анализа и обсуждения результатов
- Качество текста, рисунков и представления результатов
- Оценка в целом
- Уверенность в оценке/уровень экспертизы

Задания

- Сформулировать неформальную и формальную постановку задачи по своей теме
- Прорецензировать одну из предложенных статей по указанным выше критериям
 - Указать в чате, какую статью выбрали