

Semantic Labels-Aware Transformer Model for Searching over a Large Collection of Lecture-Slides

Supplementary Material

In this supplementary material, we show more analysis of our proposed LecSD dataset, annotation details, synonym based paraphrase sentence generation.

1. Dataset analysis

To analyze the distribution of topics in the dataset, we extracted trigram keywords from the text data of each slide image using KeyBERT [3] and obtained 252K keywords from the dataset. Figure 1 shows the sunburst of common keywords, and the data structure topics such as ‘tree’, ‘list’, ‘heap’, ‘queue’, and ‘stack’ are fairly distributed in the dataset. The inner-circle keywords occur at least 10K times, and the outer-circle keywords occur more than 50 times in the entire dataset.

To analyze the collected slide images, we use two approaches such as a) an Optical Character Recognizer (OCR) with text properties and b) layout-based image segmentation. In the first approach, we utilize off-the-shelf Google lens OCR engine¹ to extract the text information from the slide images. Further, we categorize the text in the slide based on font size. Figure 4(a) shows the distribution of words with various font sizes in the dataset. The font size is the height of a word, and the number of pixels represents the height. As shown in the figure, we divide the histogram into four regions as follows– S: small, M: medium, H: high, and L: large font size with a height of (0, 15], (15, 30], (30, 45], and > 45 pixels respectively and create the word cloud images using the words in each region. Word clouds provide a simple and effective means to communicate the most frequent words of the documents visually. We further analyzed our dataset – LecSD using image properties such as the number of color channels in the image. In the analysis, first, we sorted out the slide images based on its RGB channels and obtained only 6261 grayscale images out of 54K slide images. Generally, grayscale slide images have a less complex layout design than color slide images. Next, we studied the layout components and types of figures used in the slide images. Figure 2 shows the frequency of occurring text layout regions,

¹<https://cloud.google.com/vision/docs/ocr>

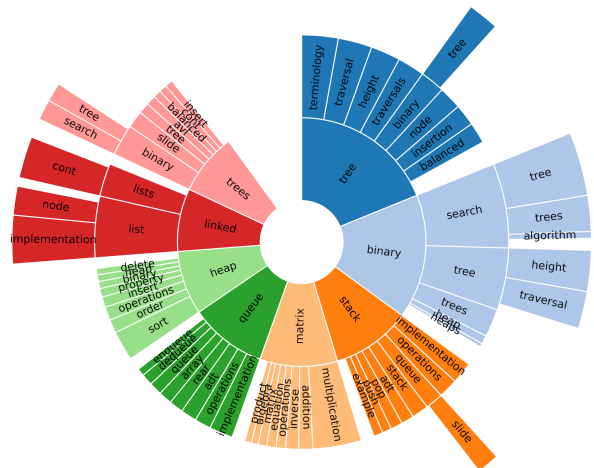


Figure 1. The topics and sub-topics in LecSD dataset visualize using a sunburst of trigram keywords extracted from the text data of each slide images using KeyBERT [3].

and Figure 3 shows the types of figures in the slide dataset. Figure 2 shows that the ‘Enumeration’ region has a higher frequency, and the ‘Footnote’ has the lowest frequency. Figure 3 shows the most frequently occurring document figure is the ‘flowchart’.

Figure 4(b) shows the word cloud of the four regions. The word cloud of regions M, H, and L have common words, and the word cloud image of the S region has different words related to the copyright and publication details. Figure 4(c) shows the typical example of four regions of text appearing in a slide image. The L, H, and M regions are ‘title’, ‘section’, and ‘paragraph/list’ items. Hence, it shares common words on the topic of *Data structures*. However, the small font size text belongs to represent ‘affiliation’, ‘date’, ‘slide number’, ‘press titles’, ‘footnote’, and ‘web-site’. Hence, the words in the S region’s word cloud differ from other regions.

We extract the text information from the slide and combine the text from similar logical regions and plot its word cloud [4] as shown in Figure 5. By analyzing Figure 5,

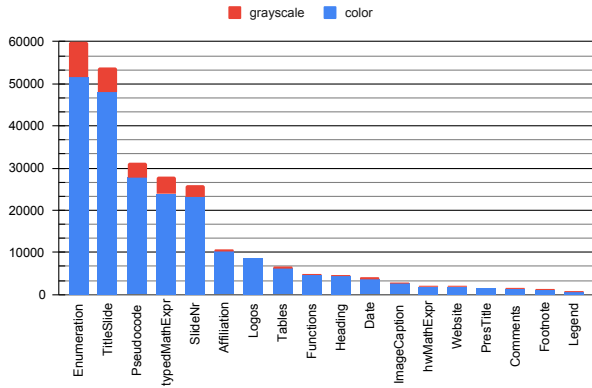


Figure 2. Distribution of text logical labels in LecSD dataset.

the layout segmentation has higher accuracy in segmenting most logical regions except ‘Date’, and ‘Slide number’. The word cloud of ‘Affiliation’, ‘Pres title’, ‘Website’, and ‘Footnote’ have high similarities, indicating a high confusion in segmenting these regions. We also plot the word cloud of all texts in the slide, and it has a high similarity to the word cloud of the ‘Enumeration’ text.

2. Slide Image summary

We used a modified version of VIA annotation software [2] and a screenshot of the web-based annotation tool is shown in Figure 10. Table 2 and 3 shows sample slide images and its manual summary, ChatGPT [1] generated paraphrase, Automatic summary, Synonym based paraphrased sentences.

In the synonym based paraphrase generation, first we identify the slide topic T and semantic regions C . We replace each semantic region with its corresponding synonyms. Some of the sample synonyms are shown in Table 1. We also randomly change structure of the sentence. The half of the paraphrased sentence with structure T explain using C and half with C is used to explain T .

3. Figure bounding box annotation

We manually annotate figure bounding boxes in the slide image. A screenshot of the web-based annotation tool is shown in Figure 11.

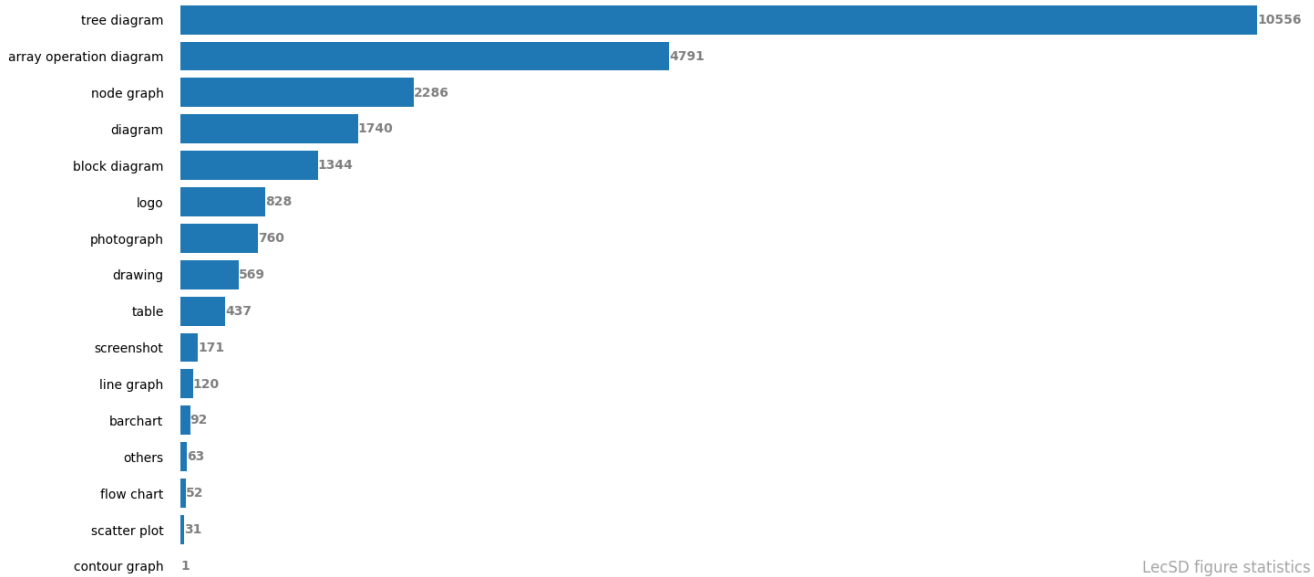
4. Figure sketches drawing

Our annotation team drew the annotated figures on a specially designed A4 paper. A screenshot of the web-based annotation tool is shown in Figure 7.

References

- [1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. 2
- [2] Abhishek Dutta and Andrew Zisserman. The VIA annotation software for images, audio and video. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM ’19, New York, NY, USA, 2019. ACM. 2
- [3] Maarten Grootendorst. Keybert: Minimal keyword extraction with bert., 2020. 1
- [4] Steffen Lohmann, Florian Heimerl, Fabian Bopp, Michael Burch, and Thomas Ertl. Concentri cloud: Word cloud visualization for multiple text documents. In *International Conference on Information Visualisation*, 2015. 1

LecSD figure statistics



LecSD figure statistics

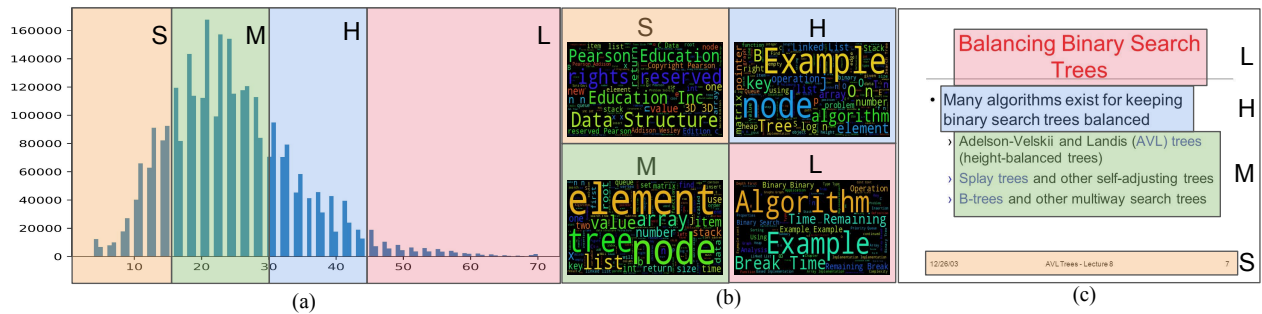


Figure 4. Visualize various font size text in the slide images (a) histogram of words with various font sizes and divide the histogram into four regions, *S*: small, *M*: medium, *H*: high, and *L*: large font size. (b) The word cloud images of words from *S*, *M*, *H*, and *L* regions. (c) A typical example of text with various font sizes in a slide image (best view in color)

block diagram	Diagram	enumeration	equation	Line graph	paragraph	explained using
Chart of relationships	Sketch	List items	Formula	Line chart	Block of text	Depicts using
Conceptual diagram	Artwork	Itemized points	Algebraic expression	Trendline	Section	elucidated by
Schematic diagram	Illustration	Key details	Numeric expression	Graph	text Segment	Used for describing
Structural diagram	Rendering	bullet points	Mathematical formula	Chart	Passage	Employed to elucidate

Table 1. A sample synonyms of commonly occurring words such as 'block diagram', 'sketch', 'enumeration', 'line graph', 'paragraph', and 'used to explain' in the summary of slide image

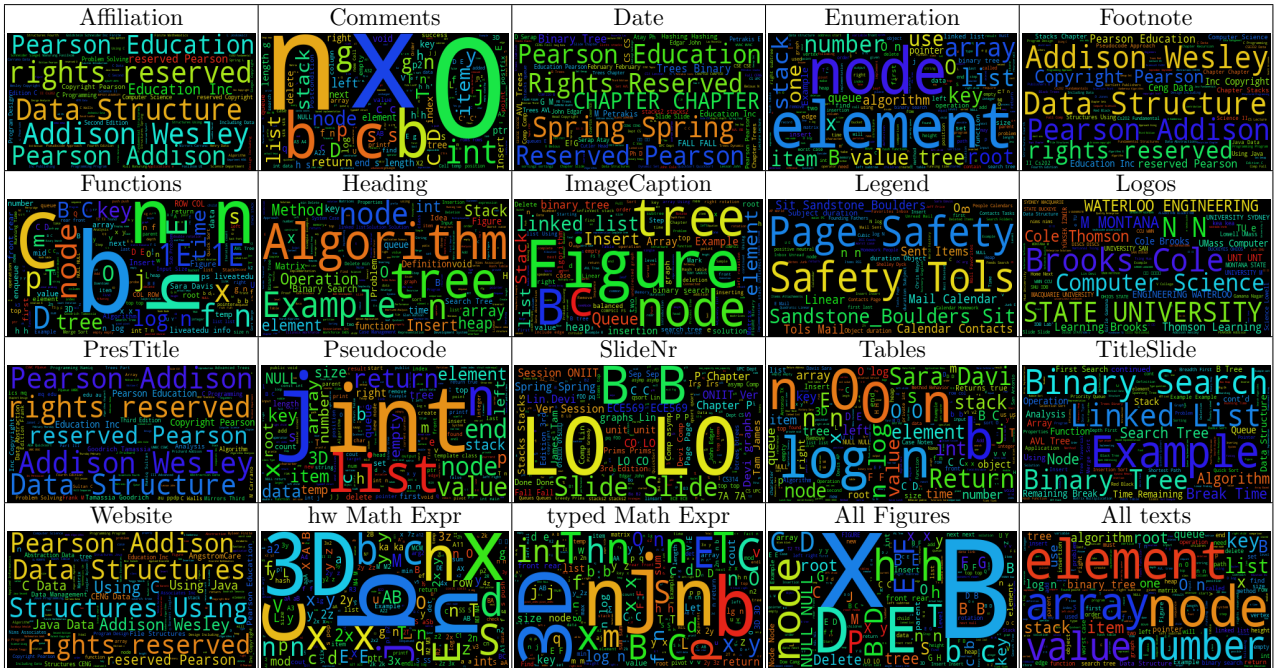


Figure 5. The word cloud image of the text extracts from the various logical regions in the slide dataset. The odd rows indicate the logical region, and the figure below is its word cloud image.

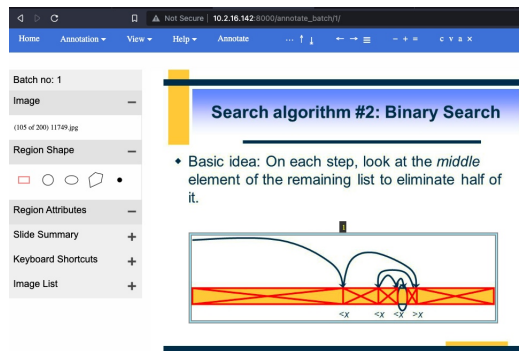
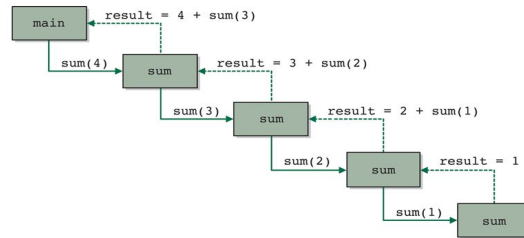


Figure 6. Figure bounding box annotation using a web-based annotation tool.

FIGURE 10.3
Recursive calls to the sum method



19

Manual Summary: A block diagram depict the recursive calls to the sum methods.

ChatGpt: A block diagram illustrates the repetitive invocations of the sum function.

Manual Summary: Recursive calls to the sum method explains with a block diagram

Synonym based paraphrase: Abstract code Used for describing Recursive calls to the sum method.

What is Program

- A Set of Instructions
- Data Structures + Algorithms
- Data Structure = A Container stores Data
- Algorithm = Logic + Control

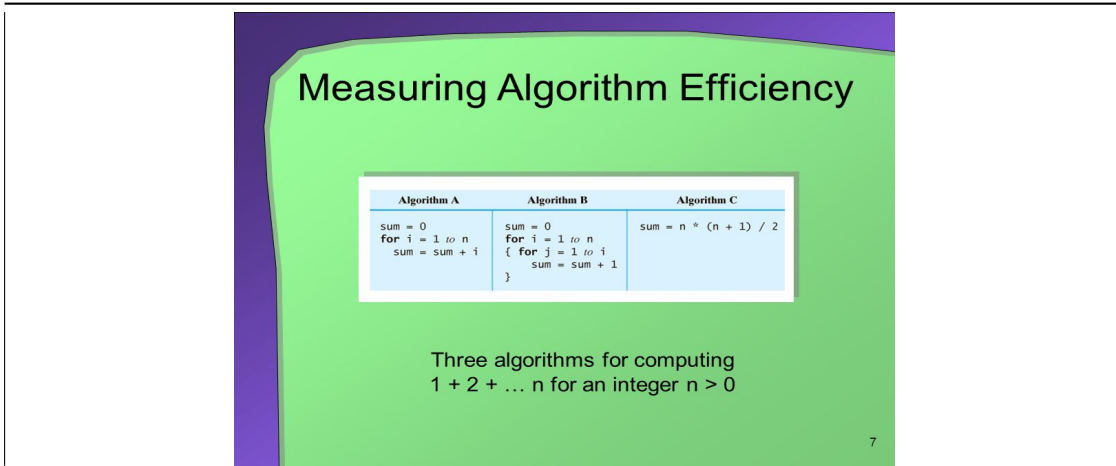
Manual Summary: The definition of program explains using enumeration.

ChatGpt: The program's definition elucidates the utilization of enumeration.

Automatic Summary: What is Program explains using Enumeration

Synonym based paraphrase: Itemized points Employed to elucidate the What is Program

Table 2. Showing sample slide images and its manual summary, ChatGPT generated paraphrase, Automatic summary, Synonym based paraphrased sentences



Manual Summary: three algorithms for calculating sum of n integers given in a table.

ChatGpt: Three methods for computing the sum of a set of n integers provided in a table.

Manual Summary: Measuring Algorithm Efficiency explains using a table and Paragraph

Synonym based paraphrase: Measuring Algorithm Efficiency Describes with tables, and text Segment.

Selection Sort: *C++ Implementation*

```
void selectionSort(DataType a[], int n)
{
  for(int last = n - 1; last >= 1; last--)
  { // set imax to the index of the largest element in a[0 .. last]
    int imax = 0;
    for(int i = 1; i <= last; i++)
      if(a[i] > a[imax]) imax = i;

    // swap a[imax] with a[last]
    DataType temp = a[last];
    a[last] = a[imax];
    a[imax] = temp;

    // invariant: a[last]..a[n-1] are the largest elements in sorted order
  }
}
```

Manual Summary: The code of Selection sort implementing using C++.

ChatGpt: The C++ implementation of Selection Sort code.

Manual Summary: Selection Sort: C++ Implementation explains using a Pseudocode

Synonym based paraphrase: Selection Sort: C++ Implementation Depicts using algorithm.

Table 3. Showing sample slide images and its manual summary, ChatGPT generated paraphrase, Automatic summary, Synonym based paraphrased sentences

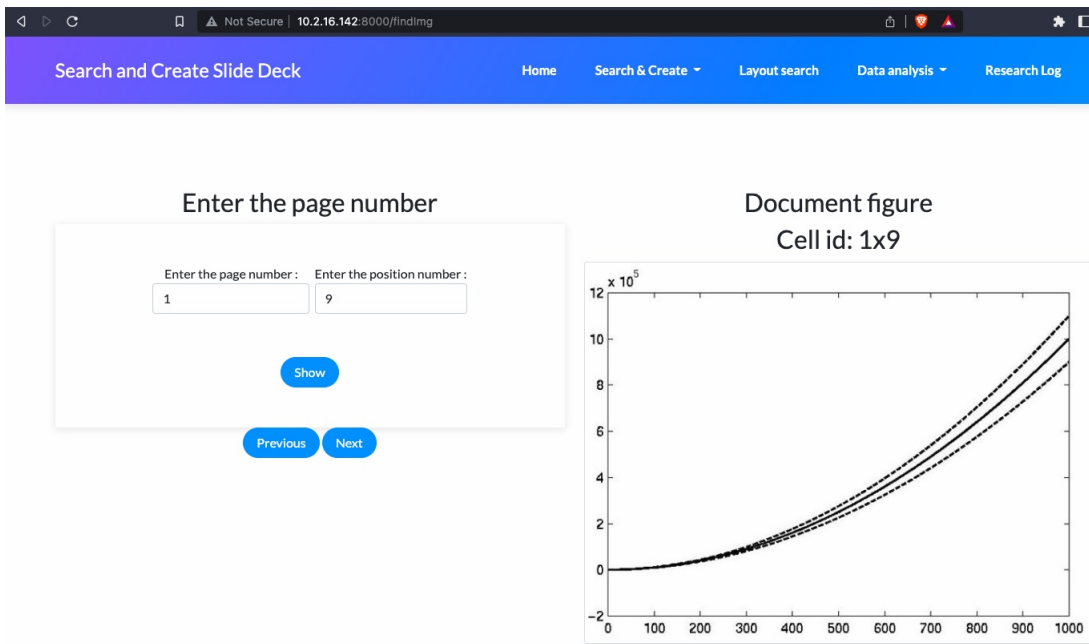


Figure 7. The web-based annotation tool shows the figure and its cell id to sketch the image.

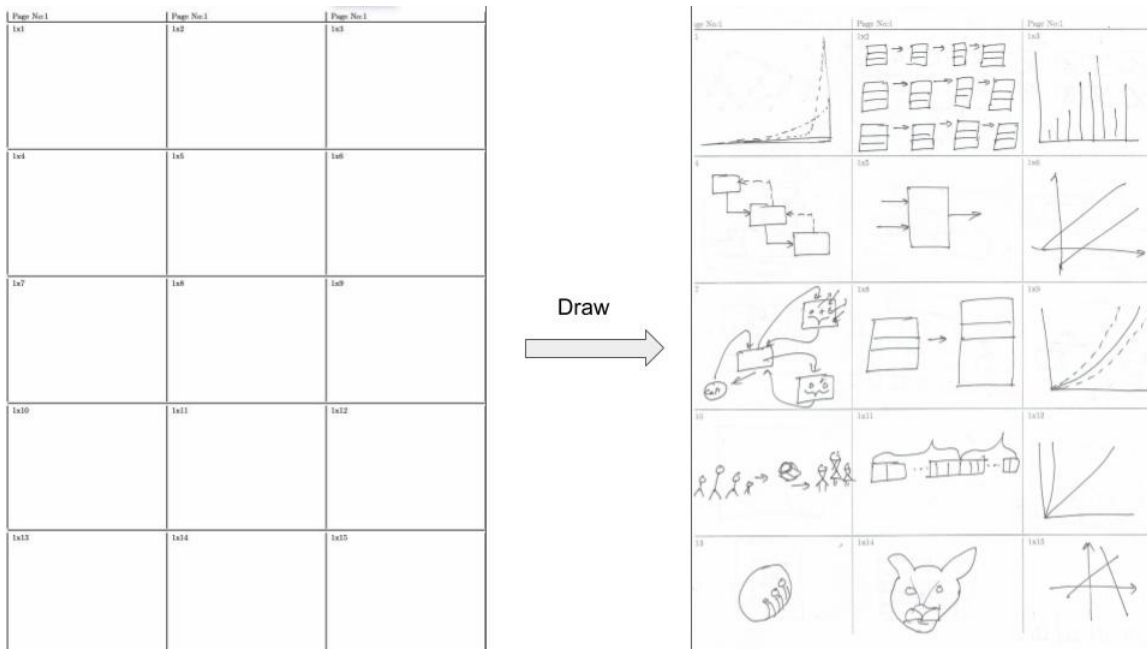


Figure 8. The left side shows the A4 paper canvas with 15 table cells with unique cell id. The right side shows the A4 paper canvas after drawing the figure image.

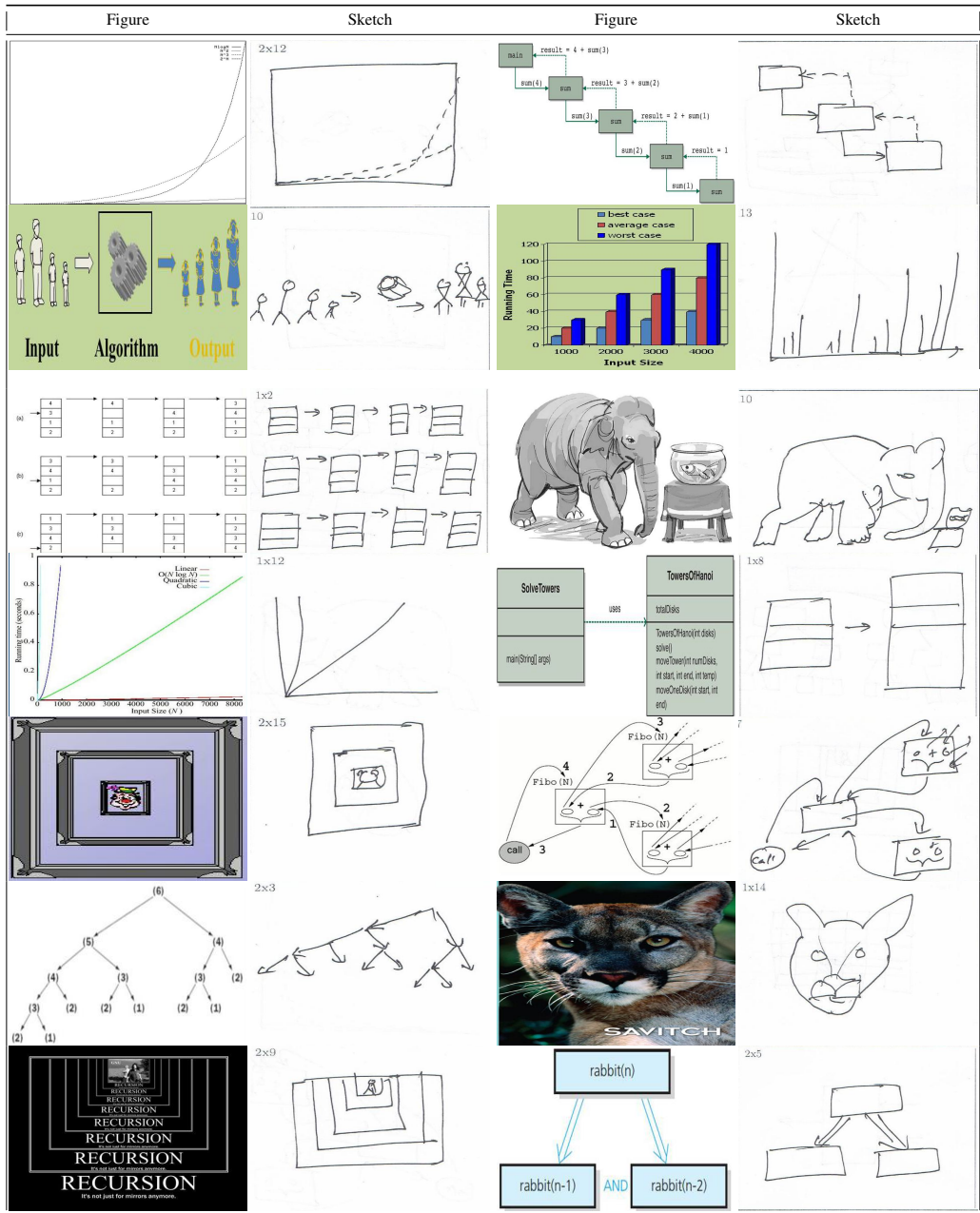


Figure 9. Sample slide image figures and their corresponding drawn sketch of our dataset. The figure columns show the figure annotated from the slide images and the right-side Sketch columns shows its hand-drawn sketches.

- Batch no: 1
- Image -
- (105 of 200) 11749.jpg
- Region Shape +
- Region Attributes +
- Slide Summary -
- Keyboard Shortcuts +
- Image List +

Search algorithm #2: Binary Search

- ◆ Basic idea: On each step, look at the *middle* element of the remaining list to eliminate half of it.

[Load images to start annotation or, see Getting Started.](#)

	summary	[Add New]
11749.jpg	Search algorithm #2: Binary Search explained with enumeration and a diagram ✎	

Figure 10. The slide image summary writing using a web-based annotation tool.

Not Secure | 10.2.16.142:8000/annotate_batch/1/

Home Annotation View Help Annotate ... ↑ ↓ ← → ≡ - + = c v a x

Batch no: 1

Image -
(105 of 200) 11749.jpg

Region Shape -
□ ○ ◌ ◊ ●

Region Attributes -

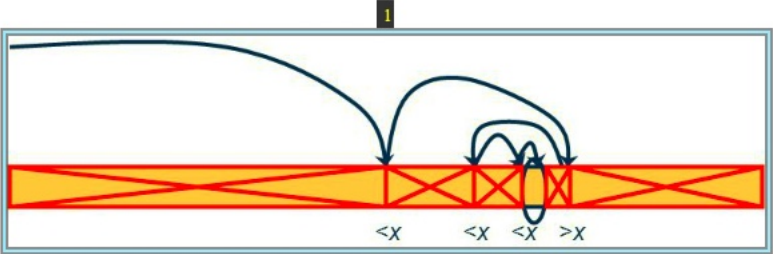
Slide Summary +

Keyboard Shortcuts +

Image List +

Search algorithm #2: Binary Search

- ◆ Basic idea: On each step, look at the *middle* element of the remaining list to eliminate half of it.



The diagram shows a horizontal bar representing a list of elements. The left portion of the bar is shaded yellow and crossed out with a red 'X', indicating it has been eliminated. A blue arrow points from the top to the middle of the remaining red-shaded portion of the bar. Below the bar, four labels are shown: '<X', '<X', '<X', and '>X'. The first three labels are under the red-shaded portion, and the last one is under the unshaded portion. A small '1' in a box is positioned above the bar, indicating the current step in the search process.

Figure 11. Figure bounding box annotation using a web-based annotation tool.