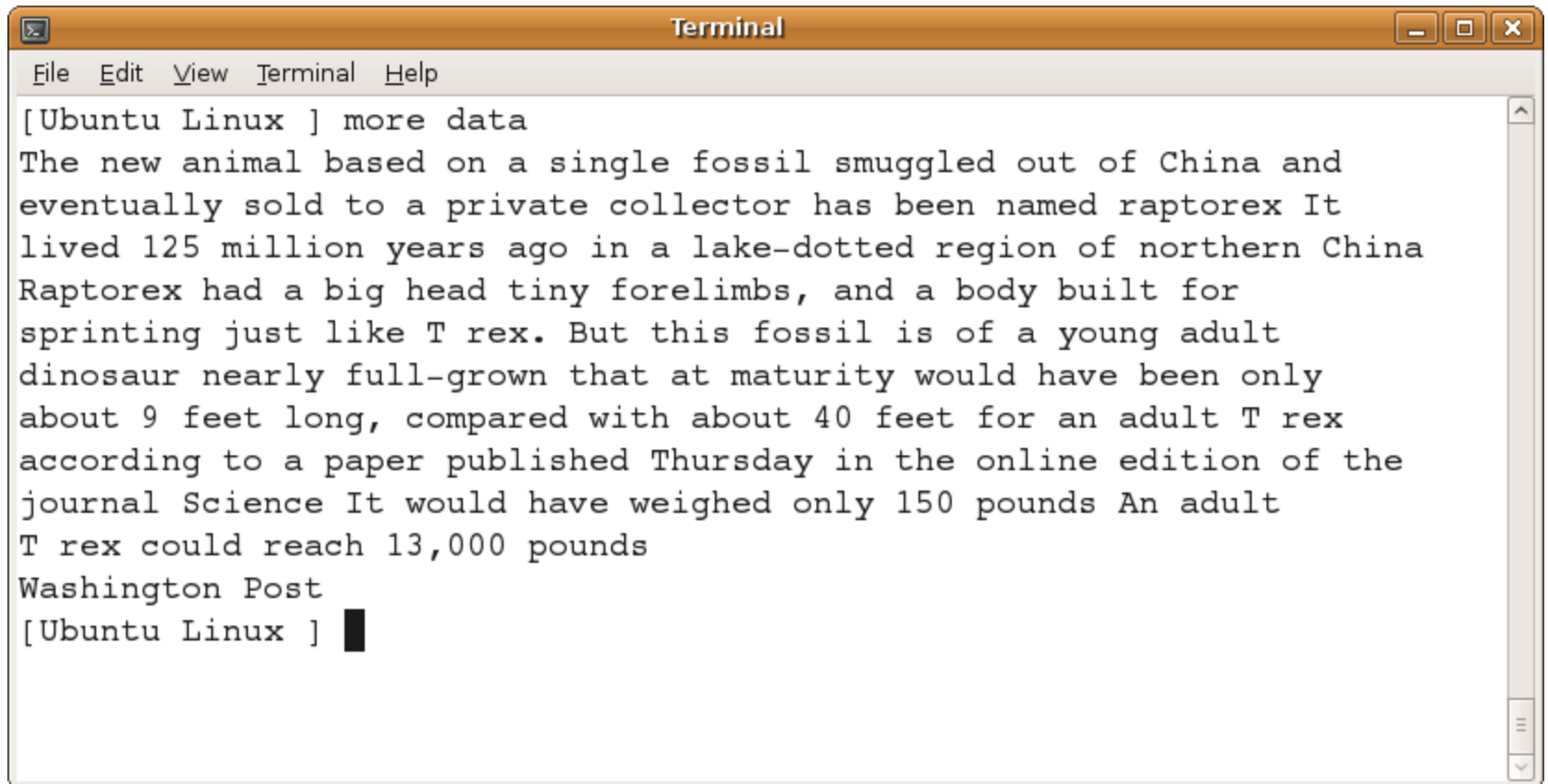


Sorting a File Line by Line Using the n-th Words

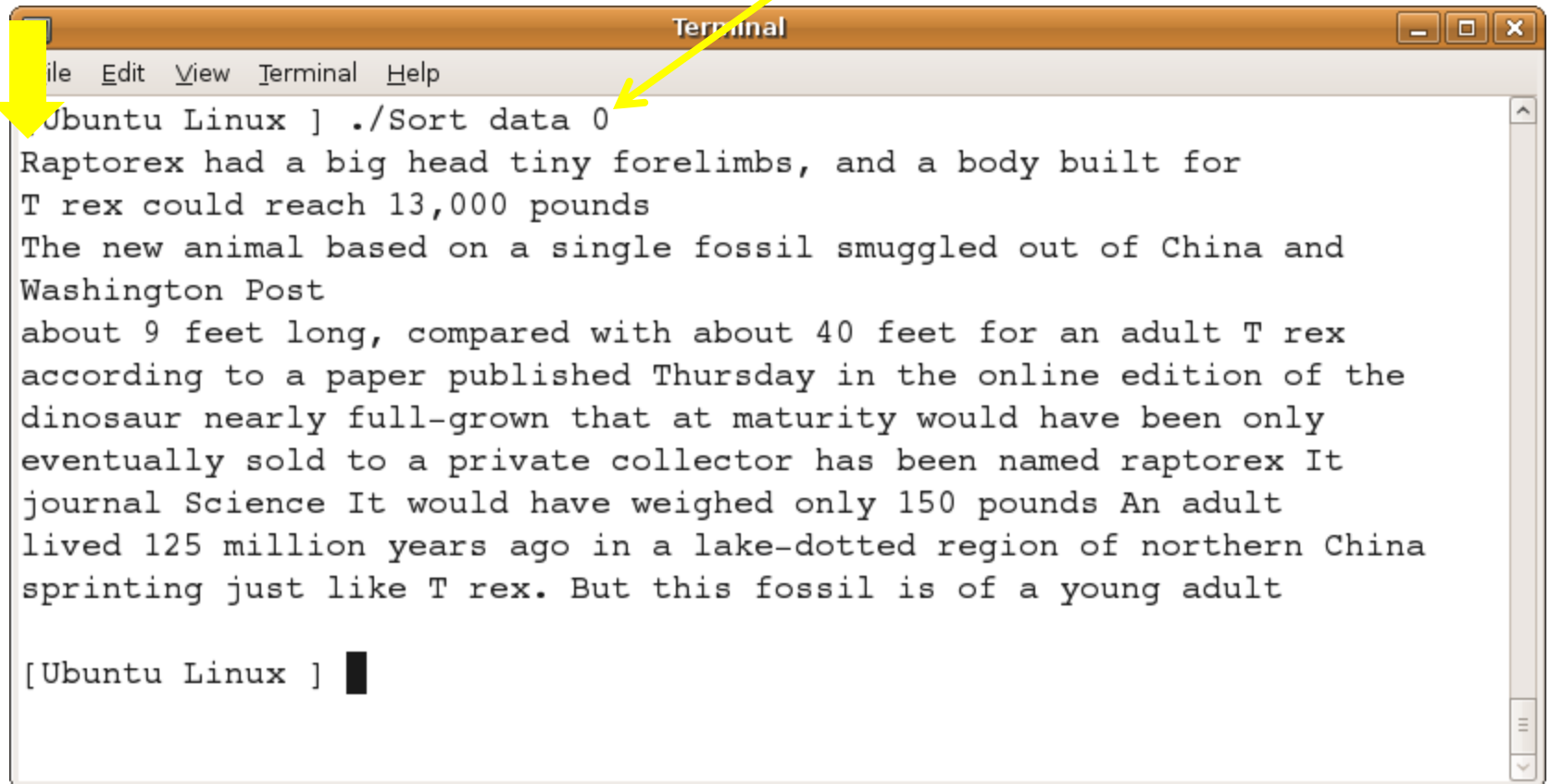
Yung-Hsiang Lu

Input

A screenshot of a terminal window titled "Terminal". The window has a menu bar with "File", "Edit", "View", "Terminal", and "Help". The terminal content shows a prompt "[Ubuntu Linux]" followed by the command "more data". The output is a multi-line text block about a dinosaur fossil named Raptorex, mentioning its age (125 million years), location (northern China), and weight (150 pounds). The text ends with "Washington Post" and another prompt "[Ubuntu Linux]" followed by a black cursor block.

```
[Ubuntu Linux ] more data
The new animal based on a single fossil smuggled out of China and
eventually sold to a private collector has been named raptorex It
lived 125 million years ago in a lake-dotted region of northern China
Raptorex had a big head tiny forelimbs, and a body built for
sprinting just like T rex. But this fossil is of a young adult
dinosaur nearly full-grown that at maturity would have been only
about 9 feet long, compared with about 40 feet for an adult T rex
according to a paper published Thursday in the online edition of the
journal Science It would have weighed only 150 pounds An adult
T rex could reach 13,000 pounds
Washington Post
[Ubuntu Linux ] █
```

Sorting by the First Word

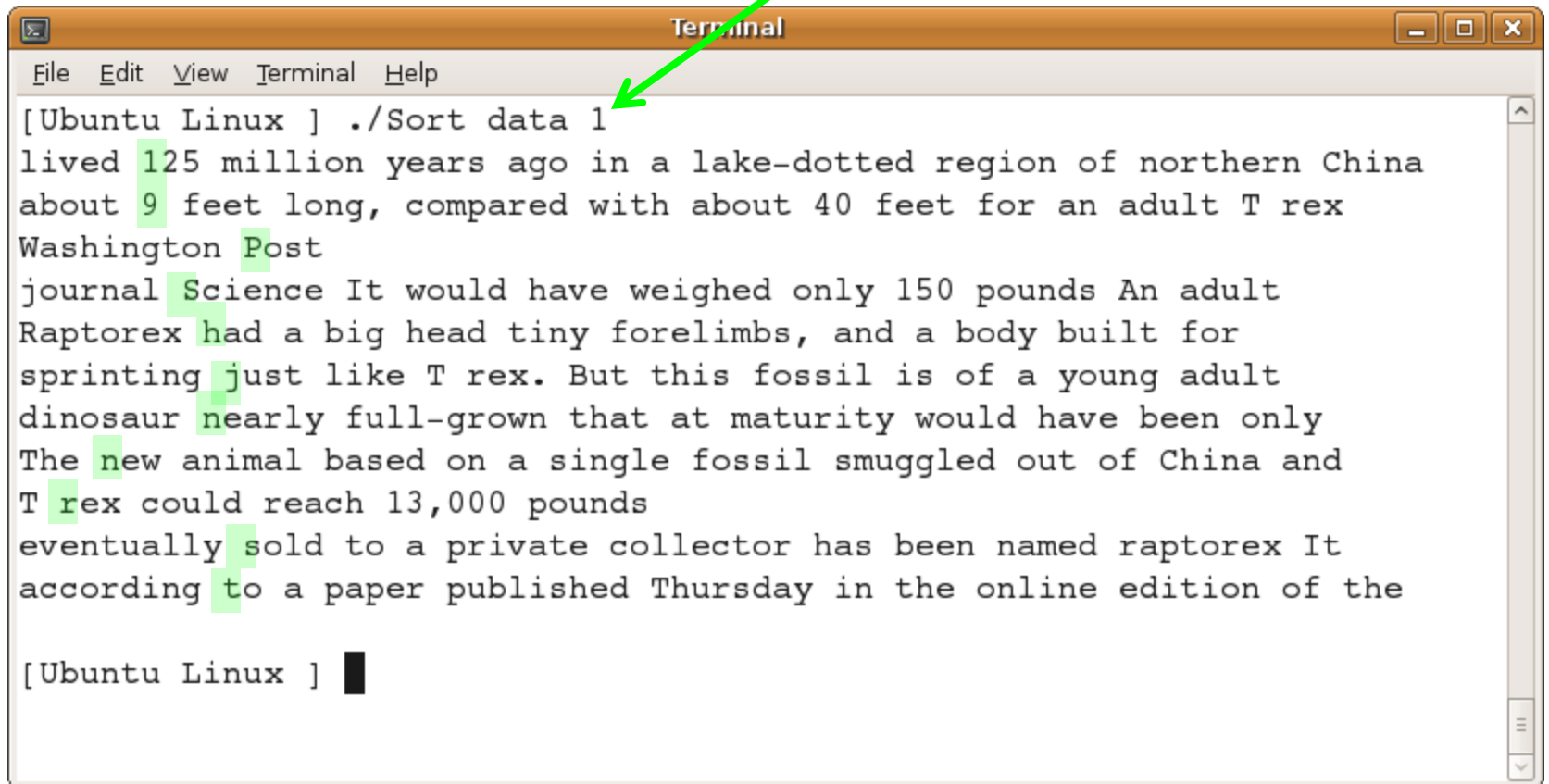


A terminal window titled "Terminal" with a menu bar containing "File", "Edit", "View", "Terminal", and "Help". The prompt is "[Ubuntu Linux]". A yellow arrow points from the word "First" in the title to the command `./Sort data 0`. The output is a paragraph of text about a dinosaur fossil. A second yellow arrow points from the word "First" in the title to the first word of the first line of output, "Raptorex".

```
[Ubuntu Linux ] ./Sort data 0
Raptorex had a big head tiny forelimbs, and a body built for
T rex could reach 13,000 pounds
The new animal based on a single fossil smuggled out of China and
Washington Post
about 9 feet long, compared with about 40 feet for an adult T rex
according to a paper published Thursday in the online edition of the
dinosaur nearly full-grown that at maturity would have been only
eventually sold to a private collector has been named raptorex It
journal Science It would have weighed only 150 pounds An adult
lived 125 million years ago in a lake-dotted region of northern China
sprinting just like T rex. But this fossil is of a young adult

[Ubuntu Linux ]
```

Sorting by the Second Word

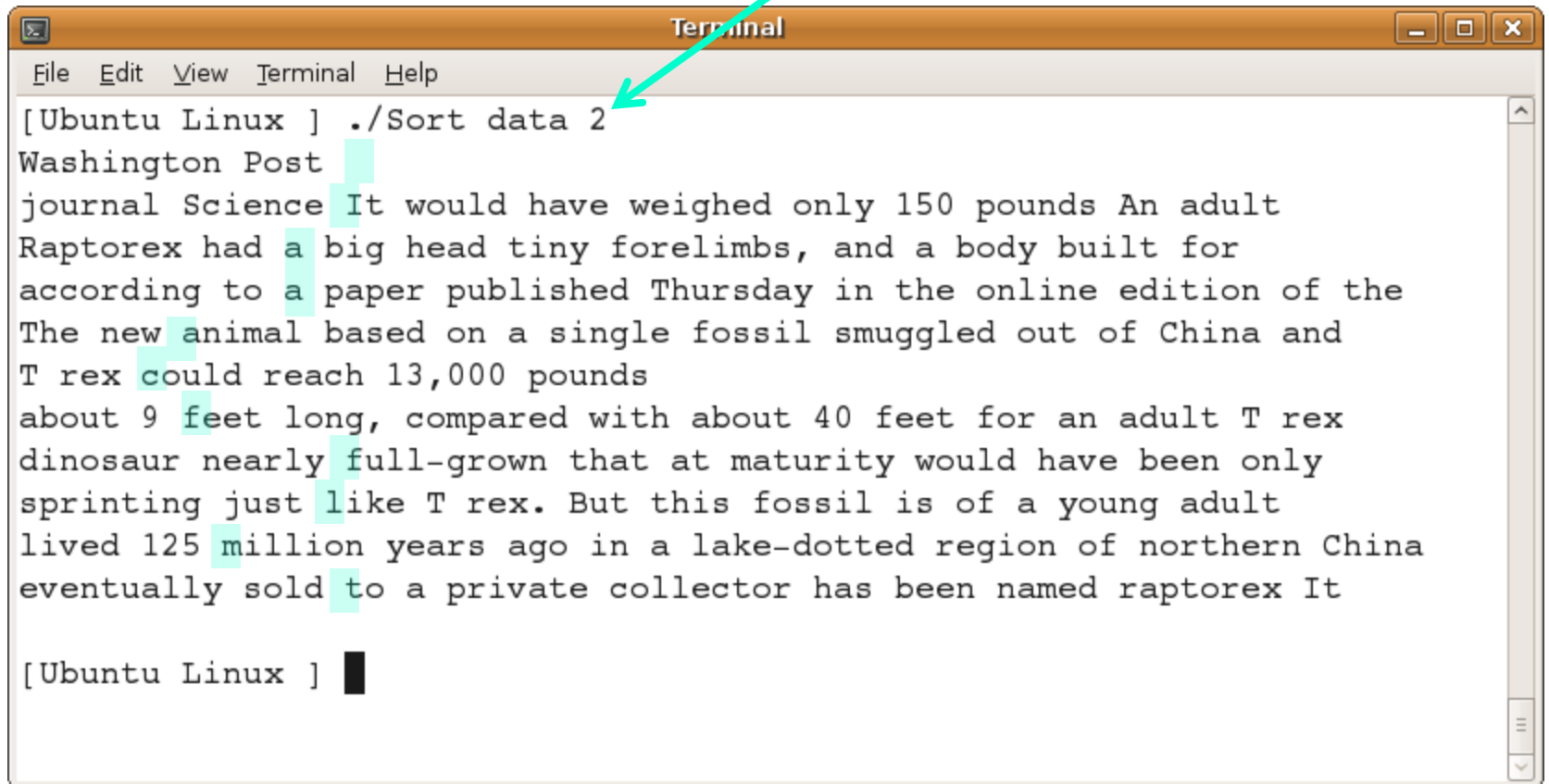


A terminal window titled "Terminal" with a menu bar (File, Edit, View, Terminal, Help). The prompt is "[Ubuntu Linux]". The command entered is `./Sort data 1`. The output is a list of text lines, some of which are highlighted in green. A green arrow points from the word "Second" in the title to the number "1" in the command, indicating the sort key.

```
[Ubuntu Linux ] ./Sort data 1
lived 125 million years ago in a lake-dotted region of northern China
about 9 feet long, compared with about 40 feet for an adult T rex
Washington Post
journal Science It would have weighed only 150 pounds An adult
Raptorex had a big head tiny forelimbs, and a body built for
sprinting just like T rex. But this fossil is of a young adult
dinosaur nearly full-grown that at maturity would have been only
The new animal based on a single fossil smuggled out of China and
T rex could reach 13,000 pounds
eventually sold to a private collector has been named raptorex It
according to a paper published Thursday in the online edition of the

[Ubuntu Linux ]
```

Sorting by the Third Word

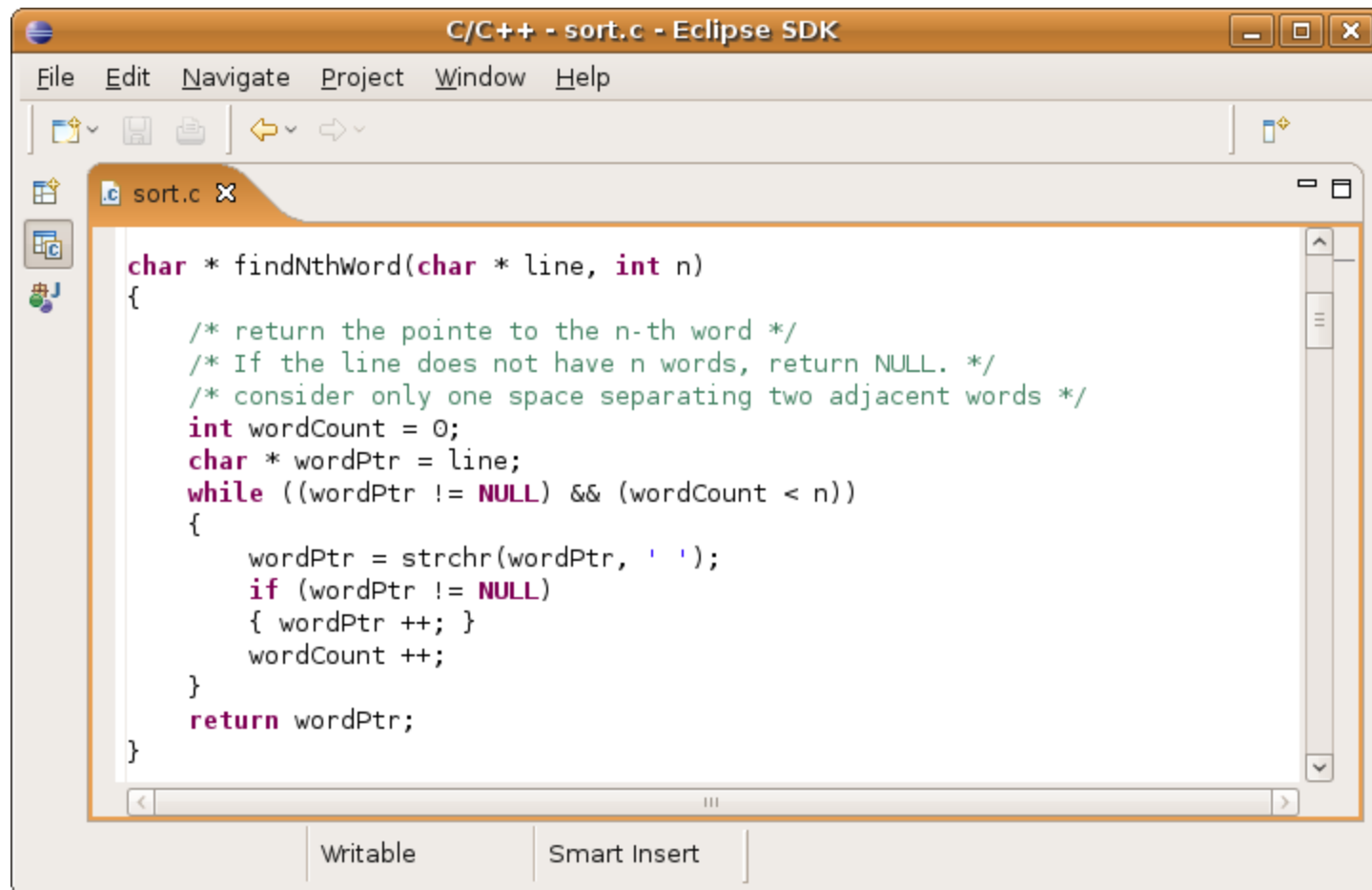
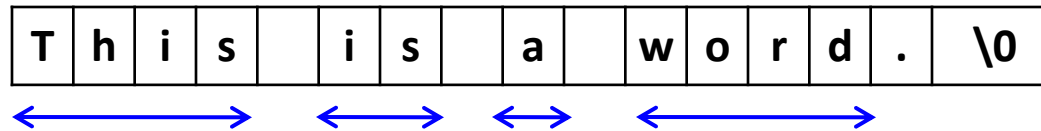


A terminal window titled "Terminal" with a menu bar (File, Edit, View, Terminal, Help). The command `./Sort data 2` has been executed. The output is a list of text lines, where the third word of each line is highlighted in cyan. A red arrow points from the word "Third" in the title to the number "2" in the command, indicating that the sorting is performed based on the third column of the data file.

```
[Ubuntu Linux ] ./Sort data 2
Washington Post
journal Science It would have weighed only 150 pounds An adult
Raptorex had a big head tiny forelimbs, and a body built for
according to a paper published Thursday in the online edition of the
The new animal based on a single fossil smuggled out of China and
T rex could reach 13,000 pounds
about 9 feet long, compared with about 40 feet for an adult T rex
dinosaur nearly full-grown that at maturity would have been only
sprinting just like T rex. But this fossil is of a young adult
lived 125 million years ago in a lake-dotted region of northern China
eventually sold to a private collector has been named raptorex It

[Ubuntu Linux ]
```

For simplicity, we assume two adjacent words are separated by one and only one space.



The screenshot shows the Eclipse IDE with a C file named `sort.c`. The code defines a function `findNthWord` that takes a character pointer `line` and an integer `n`. The function iterates through the string, counting words separated by single spaces. It returns a pointer to the start of the `n`-th word, or `NULL` if there are not enough words.

```
char * findNthWord(char * line, int n)
{
    /* return the pointer to the n-th word */
    /* If the line does not have n words, return NULL. */
    /* consider only one space separating two adjacent words */
    int wordCount = 0;
    char * wordPtr = line;
    while ((wordPtr != NULL) && (wordCount < n))
    {
        wordPtr = strchr(wordPtr, ' ');
        if (wordPtr != NULL)
        { wordPtr++; }
        wordCount++;
    }
    return wordPtr;
}
```

Find the Third Word (n is 2)

wordPtr = line ↓

T	h	i	s		i	s		a		w	o	r	d	.	\0
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	----

`wordPtr = strchr(wordPtr, ' ');`

line ↓ **wordPtr** ↓

T	h	i	s		i	s		a		w	o	r	d	.	\0
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	----

`wordPtr ++;`

line ↓ **wordPtr** ↓

T	h	i	s		i	s		a		w	o	r	d	.	\0
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	----

`wordPtr = strchr(wordPtr, ' ');`

line ↓ **wordPtr** ↓

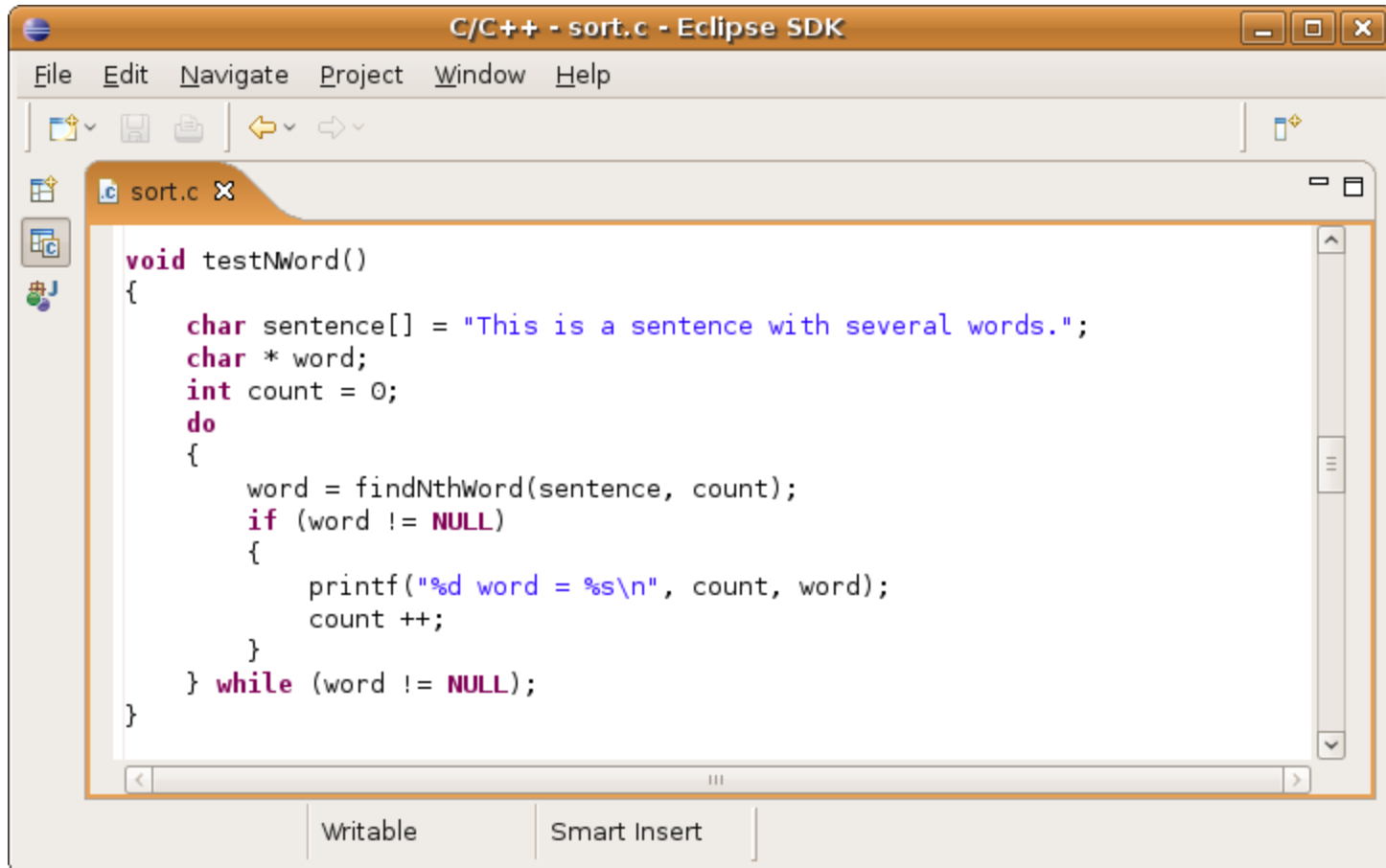
T	h	i	s		i	s		a		w	o	r	d	.	\0
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	----

`wordPtr ++;`

line ↓ **wordPtr** ↓

T	h	i	s		i	s		a		w	o	r	d	.	\0
---	---	---	---	--	---	---	--	---	--	---	---	---	---	---	----

Test findNthWord

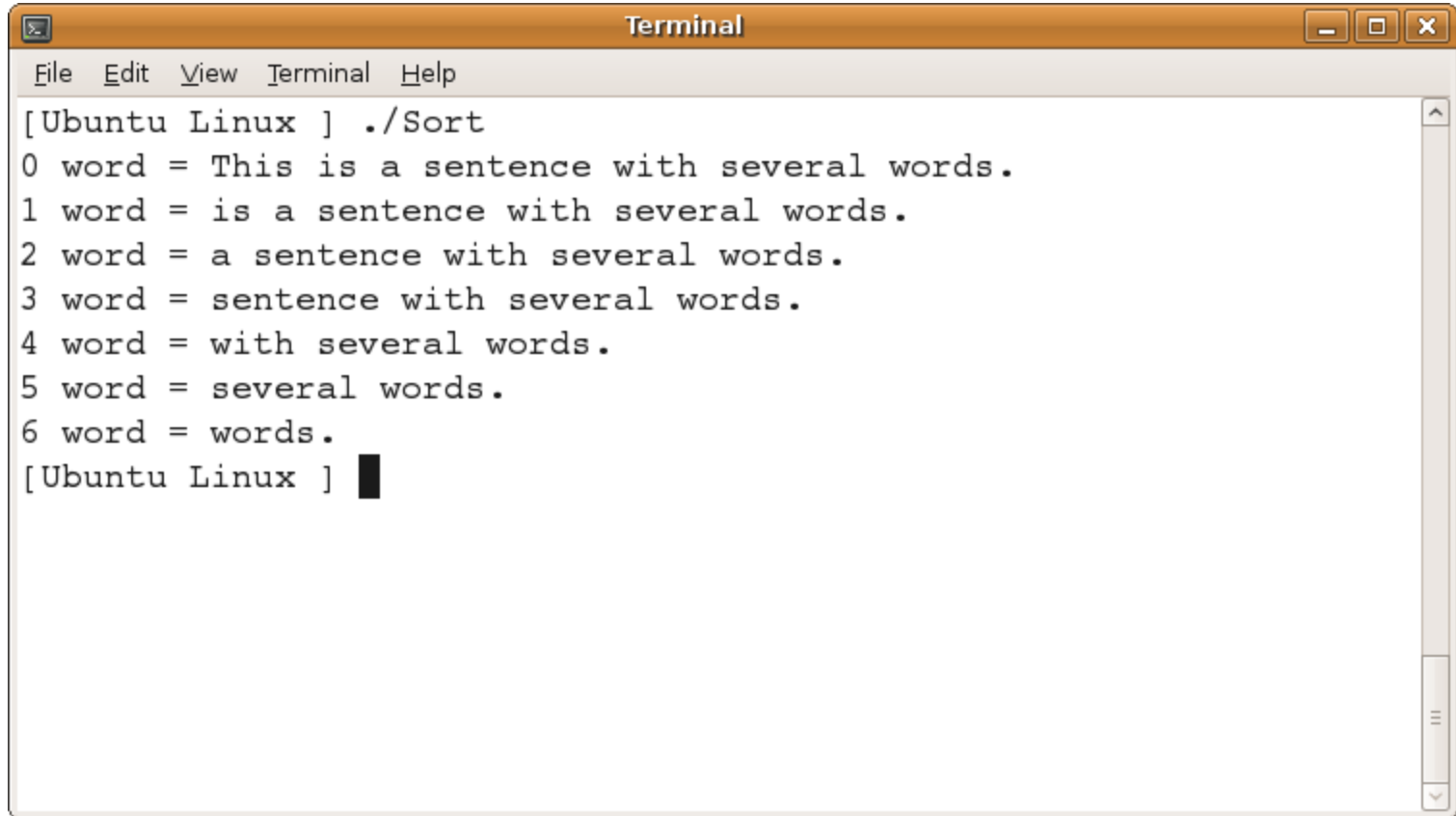


The screenshot shows the Eclipse IDE window titled "C/C++ - sort.c - Eclipse SDK". The menu bar includes File, Edit, Navigate, Project, Window, and Help. The toolbar contains icons for opening, saving, and navigating files. The editor window shows the following C code:

```
void testWord()
{
    char sentence[] = "This is a sentence with several words.";
    char * word;
    int count = 0;
    do
    {
        word = findNthWord(sentence, count);
        if (word != NULL)
        {
            printf("%d word = %s\n", count, word);
            count ++;
        }
    } while (word != NULL);
}
```

At the bottom of the editor, there are tabs for "Writable" and "Smart Insert".

Test Output

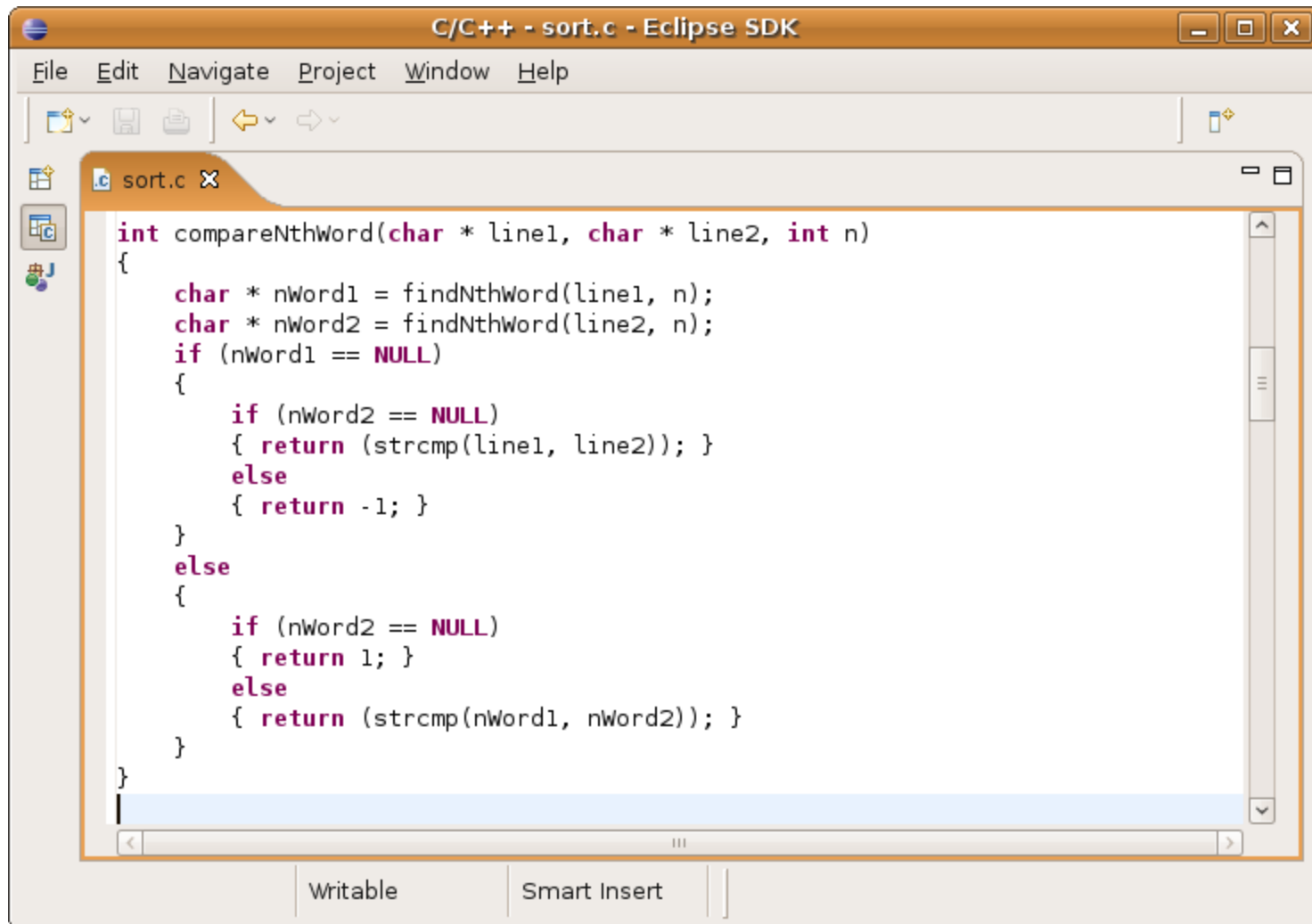
A screenshot of a terminal window titled "Terminal". The window has a menu bar with "File", "Edit", "View", "Terminal", and "Help". The terminal content shows the execution of the command `./Sort` on the string "This is a sentence with several words.". The output lists the words in order of their first occurrence: "0 word = This is a sentence with several words.", "1 word = is a sentence with several words.", "2 word = a sentence with several words.", "3 word = sentence with several words.", "4 word = with several words.", "5 word = several words.", and "6 word = words.". The prompt `[Ubuntu Linux]` is followed by a black cursor block.

```
[Ubuntu Linux ] ./Sort
0 word = This is a sentence with several words.
1 word = is a sentence with several words.
2 word = a sentence with several words.
3 word = sentence with several words.
4 word = with several words.
5 word = several words.
6 word = words.
[Ubuntu Linux ]
```

Define Line Order

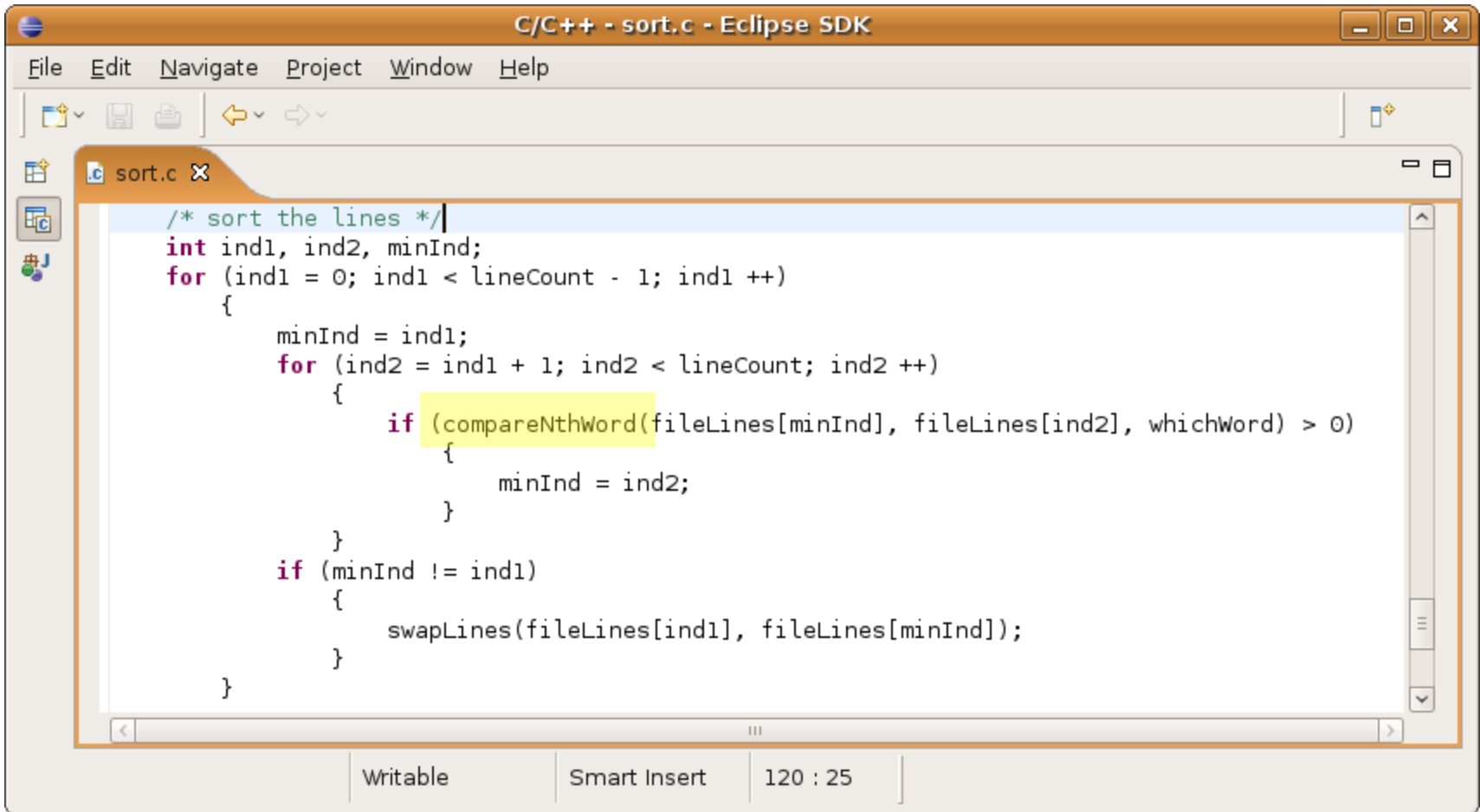
line1's nth word (word1)	line2's nth word (word2)	order
NULL	NULL	strcmp(line1, line2)
NULL	not NULL	line1 < line2 (-1)
not NULL	NULL	line1 > line2 (1)
not NULL	not NULL	strcmp(word1, word2)

Compare Lines by their n-th Words



```
int compareNthWord(char * line1, char * line2, int n)
{
    char * nWord1 = findNthWord(line1, n);
    char * nWord2 = findNthWord(line2, n);
    if (nWord1 == NULL)
    {
        if (nWord2 == NULL)
        { return (strcmp(line1, line2)); }
        else
        { return -1; }
    }
    else
    {
        if (nWord2 == NULL)
        { return 1; }
        else
        { return (strcmp(nWord1, nWord2)); }
    }
}
```

Sort the Lines by the n-th Words



The screenshot shows the Eclipse IDE interface with a C/C++ project named 'sort.c'. The code is a selection sort algorithm that sorts lines of text based on the word at a specific index 'whichWord'. The code is as follows:

```
/* sort the lines */
int ind1, ind2, minInd;
for (ind1 = 0; ind1 < lineCount - 1; ind1++)
{
    minInd = ind1;
    for (ind2 = ind1 + 1; ind2 < lineCount; ind2++)
    {
        if (compareNthWord(fileLines[minInd], fileLines[ind2], whichWord) > 0)
        {
            minInd = ind2;
        }
    }
    if (minInd != ind1)
    {
        swapLines(fileLines[ind1], fileLines[minInd]);
    }
}
```

The status bar at the bottom indicates 'Writable', 'Smart Insert', and the cursor position '120 : 25'.

Summary: What have we learned?

- structure of C programs: main, functions
- data types: int, char, float, array
- flow control: if, while, for, switch, function
- file: write and read
- memory: allocate and release
- algorithm: selection sort