

## THE ANALYSIS OF ROMANIAN MICRO-ENTITIES: FROM ONE-MAN BUSINESSES TO DRUM-BEATS ENTITIES

LUCIAN HANCU<sup>1</sup>

**ABSTRACT.** The rapid evolution of the Internet in the recent years has become a huge opportunity for businesses around the world, especially for the small ones, who do not have the strength and experience of the multinational corporations. The availability of a large amount of online information on the existing and acting entities is a fundamental step to the creation of more complex online services which are based on the raw material. In this article, we present a five-year analysis we have performed on the Romanian Micro-entities, based on their available financial statements. We have conducted our analysis on the last five financial years to investigate the source of their growth or fall and also their potential in the coming post-integration years.

**Keywords:** financial statements mining, micro-entities, taxation impact

**JEL Classification:** C81

### Introduction

Analyzing past economical behavior [Cam03, Efi04] has always been a major concern for social scientists, in order to predict future movements on the global, national and local economy [Cap04, Eco04]. The undoubted advantage of such an analysis is a broaden view of each economy or sector of the economy, contributing to a better understanding of the incoming opportunities and threatenings that each new day brings to the global business man.

A first disadvantage of the mentioned analysis publication is that they tend to ignore small opportunities that come out when analyzing smaller markets. Usually such opportunities are crucial for entrepreneurs in order to fundament their decisions to invest in a new company, market or sector, or to disinvest from these ones. The approach that we discuss in this article considers only the micro-enterprises from the Romanian companies, whose last five years financial statements were gathered and classified using automatic tools [Han07].

The article is structured as follows: the subsequent section introduces our method of gathering the data publicly available from Invisible Web sites [Han02] and analyzing it [Han07]. We present our results of applying the techniques to a large collection of financial statements, which were previously extracted from the Web, in order to classify the Romanian entities by considering two dimensions: the number of employees and the trend (positive, wavering or negative) of the profit margins. In our concluding section we present future directions for the types of

---

<sup>1</sup> SoftProEuro S.R.L., Cluj-Napoca, E-mail: lhancu@softproeuro.ro

analysis that we propose here and some insights on how to make available this kind of analysis to the large public.

### Material and methods

In the previous works [Han02, Han04, Han07], we investigated several methods of how to gather and index data from the Invisible Web (that part of the Web that is not publicly available to classical search engines such as Google, Yahoo, or MSN Live Search). Our recent article [Han07] reveals the importance of applying simple data extraction methods [Cha03, Han03, Kon06] in order to create new knowledge from the already available Web material.

As mentioned in the concluding part of the *Enhancing the Invisible Web* article, our intention was to tackle multi-year analysis on the financial statements of the Romanian companies.

Our method basically consists in classifying the material (228.938 Romanian entities) into two major groups: *micro-entities* (up to nine employees in each one of the analyzed financial year and a turnover up to 100.000 EUR during each one of the years) and *other entities*. The micro-entities are subsequently classified in other nine groups (*three groups* for the classes: [0-1] employees, [2-5] employees, [6-9] employees and *three groups* for the classes: continuous growth in profit margins, undecided direction or wavering in profit margins; continuous fall in profit margins).

### Results and discussions

We have investigated the financial statements of 228.938 Romanian entities from the last 5 financial years (2001-2005, since the 2006 financial statements are to be publicly released during the fall of 2007). From the analyzed data, we have extracted 189.410 micro-entities (whose employee's number is less than 10 and whose each year's turnover is less than 100.000 EUR).

From this data source, we have computed two dimensions for a more thorough micro-entities data analysis: the first one is the number of employees of each entity, which we have divided in three groups: **Micro 1**: the entities with zero or one employee (we call it: *one man business*), **Micro 3**: the entities with two to five employees (we call it: *middle-sized micro-entity*) and **Micro 9**: the entities with six to nine employees (we call it: *drum-beats entity*, since it uses the highest range of permitted number of employees to be categorized as micro-entity).

The second dimension takes into consideration the profit margin (gross profit divided by turnover) of each one of the entity, on the 5-years interval. We obtain three distinct groups: **the first**: the group of entities which had a continuous increase in the profit margins (which is that the profit margin of the first available year was less than or equal to the one of the second available year, whereas the profit margin of the fourth available year was less than or equal to the one of the last available year); we call this group **the runners**. The second group consists of entities which had a wavering of the profit margins (increase, followed by decrease, or decrease followed by increase, or any discontinuous combination of

the two in the analyzed interval): we call it: the **undecided**. The last group is the group of the companies whose profit margin was in continuous decrease over the last five years: we call this group the **turtles**, as their evolution is left behind the overall evolution of the last years Romanian economy.

By analyzing the data from the two mentioned dimensions' perspective, we compute the micro-entity matrix analysis shown in [Table 1], which depicts the number of entities from each analyzed group:

**Table 1.**  
The Micro-entities matrix analysis

<b>Micro 11</b> 9373	<b>Micro 12</b> 97007	<b>Micro 13</b> 1819
<b>Micro 31</b> 3875	<b>Micro 32</b> 59896	<b>Micro 33</b> 1218
<b>Micro 91</b> 763	<b>Micro 92</b> 15142	<b>Micro 93</b> 317

### The One-man business (Micro 11)

The One-man business is the entity which has zero or one employee. It is usually a small family business, in which the owner (the founder) has also executive powers in that business. We furthermore classify the one-man business into three groups: *the runner one-man business*, *the undecided one-man business*, and *the turtle one man business*.

#### 1. The Runner

We define the *Runner One-man business* as the entity whose profit margin was in a positive trend during the last five financial years and whose numbers of employees is less than two. Our system classified 9,373 Micro 11 entities, that is 4.95% of all classified entities.

By analyzing the distribution of Micro 11 entities among the host county, we discovered that the *champion* is the Cluj county (see **Figure 1**), with a maximum value (*maxVal*) of 1472, followed by Bucharest (column *BUC* in the cited figure), with only 1205 classified entities. The year distribution (**Figure 10**) reveals that the years 1992, 1993 and 1994 are the years on which the Micro 11 entities mostly started, whereas the 1998 year was the worst in this segment.

Interesting insights come out when analyzing the two distributions: from the county distribution we get the top counties (which are Cluj, followed by Bucharest, Bacau, Galati, Dolj), and the worst economic years (1998, followed in ascending order by 1999 and 2000).

We also need to point out that our data model was built on a large but not exhaustive database. We have analyzed active companies whose financial statements were publicly available, but different results could come out if analyzing 400.000 or more active companies. We believe that the Cluj county could, in this latter case, be surpassed by Bucharest, as the total number of entities in Bucharest is 5 times more than in Cluj [Recom]

## 2. The Undecided

The Undecided One-Man Business is by far the most classified entity from our data repository. We classified 97007 Micro 12 entities, that is 51.22% of all classified ones. The county distribution of Micro 12 (see **Figure 2**) reveals that Bucharest has the largest number of entities (20704, see also *maxVal* in the image), followed by Cluj (8961), Timis (5490), Arges (3436). We notice that there is a group of counties which have an appropriate number of classified entities (Brasov 3220, Dolj 3093, Galati 3300, Mures 3168).

The year distribution shows an exponential evolution from 1998 to 2004 (556, followed by 2194, 2665, 4106, 7610, 19568 and 33096 – in 2004 – of classified entities). We mention that – compared with Figure 10 – the current figure is shapely different. We shall see that it is unique in its shape among the other classified micro-entities groups.

## 3. The Turtle

The Turtle One-Man Business (Micro 13) is that business which moves slower than the rest of the economy. It consists on those entities having zero or one employee, and which has a negative trend in profit margins or no profit at all. We have classified 1819 of all the analyzed entities (0.96%). The county distribution shows a peak in Bucharest county (*maxVal* – 270), followed by Cluj (197), Bacau (107) and Timis (102). We notice that Cluj is a serious competitor of Bucharest in terms of Micro 13 entities.

The year distribution shows a two-peaks shape (1994 with 330 classified entities, followed by 1992 with 320 classified entities). The shape is similar to the Micro 11 year distribution, the same positive trend is revealed when analyzing the 1998-2001 years. Surprisingly, Micro 13 entities have no representation in the range 2002-2004. We explain this phenomenon by the greater number of **Micro 12** entities, which have the peaks in the 2002-2004 years interval. Fiscally speaking, this phenomenon was possible by the introduction of the 1.5% tax, which could explain the naissance of a greater number of new and profitable entities.

## The Middle-sized entity (Micro 31, Micro 32 and Micro 33)

We named the Middle-sized entity that entity which has at least 2 employees and at most 5 employees. We also classify the *runner middle size entities* (sub-section 3.2.1), the *undecided middle size entities* (sub-section 3.2.2) and the *turtle middle-size entities* (sub-section 3.2.3) by considering the evolution of the profit margins during the last financial years.

## 1. The Runners

The Runner Middle-sized entity (**Micro 31**, see **Figure 4** for county distribution, **Figure 13** for year distribution) has 2 to 5 employees and a continuous growth in profit margins. Notice that if one entity has 2 employees in 2001, followed by only one employee in 2002-2004, it will fall into the Micro 31 category,

as our classifier considers the minimum and the maximum number of employees during the analyzed interval.

The county distribution shows a peak on Bucharest county (388 entities), followed by a closer Cluj county competitor (305 entities), then Galati (234), Timis (202) and Bacau (198).

The year distribution shows two peaks (1994, followed by 1992), then minimum values on 1998, 2002-2004. The shape is similar but not identical to the year distribution of **Micro 11** and **Micro 12**, with few differences.

## 2. The Undecided

The number of undecided middle-sized entities is the second among the all 9 micro entities classifications: 59.896 (31.62%), when compared to only 3875 (2.05%) for **Micro 31** and 1218 (0.64%) for **Micro 33**.

The county distribution shows the same peak of Bucharest county (10874 entities), followed by Cluj (4932), Timis (3756), Galati (2227), Mures (2188) and Arges (2126). The year distribution (see **Figure 14**) reveals an unexpected 2003 peak (10485), followed by 8296 companies which began in 2004. The year shape has a different form when compared to any of the above cited Micro entities groups (**Micro 11**, **Micro 12**, **Micro 13**, **Micro 32**), the unexpected 2003 peak can be also motivated by the fiscal facilities for the Romanian micro-entities.

## 3. The Turtles

The *Turtles Middle-sized entities* have a negative trend in the profit margins during the last five financial years. Fortunately, the number of the classified entities is almost insignificant (less than 1% of all the classified entities). Unfortunately, Bucharest remains on the top of the list with 192 entities, followed by Cluj (94), Timis (71), Galati (59) and Arges (58). The county distributions shape of **Micro 31**, **Micro 32** and **Micro 33** companies are similar, the same counties appear in the first two places, whereas the third, fourth and fifth places do not show many variations.

The year distribution (depicted in **Figure 15**) shows two peaks (1994 and 1992), no classification for 2003 and 2004. Its shape is similar to Micro 31, which had less evidence in the years 2002-2004.

## The Drum-Beats Entities

### 1. The Runners

The Runners Drum-Beats entities are classified as the entities with more than six and less than 10 employees, whose profit margin was in a continuous rise since 2001. We found only 763 companies through the analyzed data, being only a small fraction (0.40%) from the total number of our dataset of available companies.

The county distribution (see **Figure 7**) shows the same Bucharest peak (89 companies), followed by a closer position of Cluj county (80 companies – notice the small difference between the two groups), then Timis (47), Galati (41) and Bacau (37).

The year distribution (see **Figure 16**) shows a two peaks shape (1992 and 1994), then small evidence on 2003 and 2004. The shape is similar to the one of **Micro 31** (see **Figure 13**), thus making us to believe that entities from the same columns of the matrix act in similar ways.

## 2. The Undecided Drum-Beats Entities (Micro 92)

The Undecided Drum-Beats entities (see **Figure 8** for county distribution and **Figure 17** for year distribution) have an uncertain profit margin trend (rise, followed by downfall or downfall followed by a rise). The county distribution shows a peak in Bucharest data (2754 entities), followed by Cluj (1228 entities), Timis (1008 entities), Mures (574) and Galati (547). The total amount of Micro 92 classified entities (15142) occupies the third place in the overall classification (surpassed by Micro 12 and Micro 32). It seems that the uncertainty was the major trend in the last 5 financial years, probably because the years 2001-2002 suffered from the 1997-1999 recession interval.

The Year distribution reveals a peak in 2003, followed by a strong interval 1991-1994, then 2002 and 2003. The similarity with the **Micro 32** year distribution reveals that medium-sized micro entities and drum-beats entities have close behavior patterns.

## 3. The Turtles

The Turtles Drum Beats entities (depicted in **Figure 9** – county distribution, then **Figure 18** – Year distribution) occupy only 0.17% (317 entities) of the classified data. This is a positive result, as the *turtle entities* describe companies with negative trend of profit margins.

The county distribution reveals a peak for Bucharest County (50), then Cluj (21), Galati (19), Timis (18) and Bacau (15). We notice that the hierarchy of counties is almost unchanged, the closer group of Cluj, Galati, Timis and Bacau counties can be motivated by the small number of overall **Micro 93** companies, which prohibits a distribution with greater differences between groups.

The year distribution (**Figure 18**) depicts a similar shape as the **Micro 13** and **Micro 33** year distribution: it has three peaks (notice the difference of only two peaks for **Micro 13** and **Micro 33**) in the years 1991, 1992 and 1994.

During the previous paragraphs, we have pointed out and discussed several patterns that we found when classifying the Romanian micro-entities. The similarities that we found between **Micro 11**, **Micro 31** and **Micro 91** Year distributions show that there is a relationship between the companies whose profit margins follow the same trend, independently of their number of employees. The same similarity appears to a second group (**Micro 32** and **Micro 92**) and a third (**Micro 13**, **Micro 33** and **Micro 93**). An interesting result that this analysis reveals is that **Micro 12** companies have a different year distribution shape and that they are more than half of the all classified micro entities. This shows that the fiscal facilities can conduce to the naissance of profitable companies (the year distribution of **Micro 12** entities reveals an increase in the number of companies during 2003 and 2004). We also intend to analyze the impact of the supplementation of the micro-entities tax due to come in the post-integration years.

## Conclusions

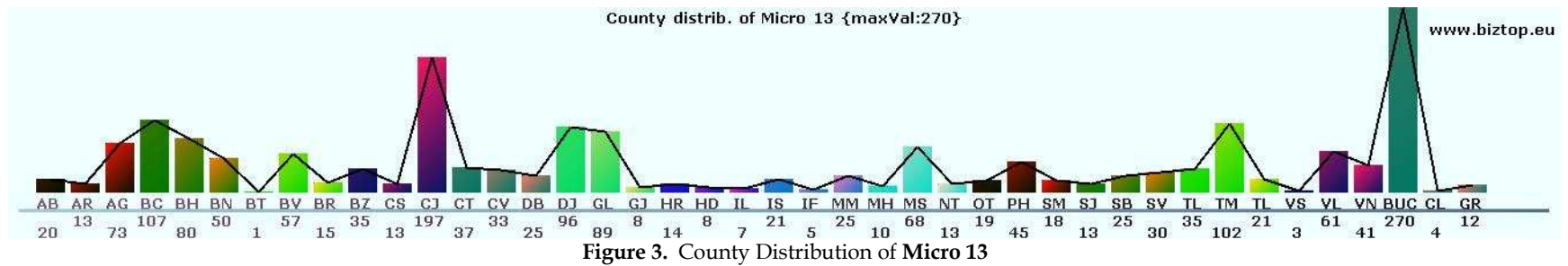
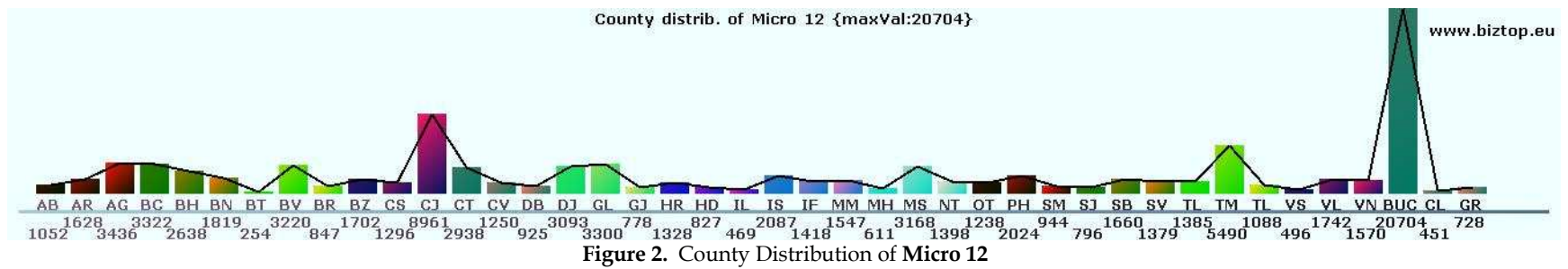
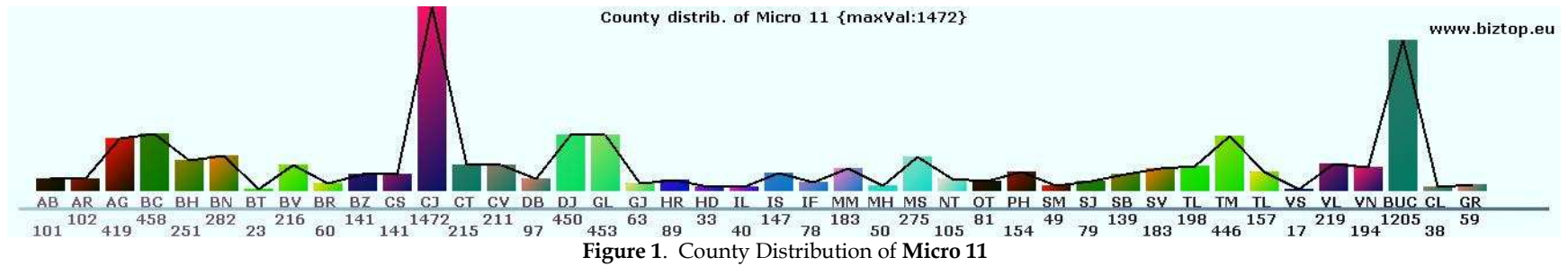
We have presented a method for analyzing the financial statements of more than 200.000 active Romanian companies and build a model of analysis. We intend to include our model in classifying the top of *Romanian Entities* [TopBiz, Recom] in order to extract the entities from the same classification group. Our previous method was to include small entities in the same group as large entities, thus ignoring a large part of potentially efficient companies whose turnover was below the top competitors.

Furthermore, we explore techniques for applying the analysis to small entities (10-49 employees), medium-sized and large entities. Even if micro-entities are by far the most representative group in any economy (by the number of active entities), we cannot ignore the entities whose number of employees exceeds ten employees.

The analysis can also be extended to derive strategic dependencies of one company on its supplier: by classifying the supplier as Micro 11 to Micro 93, we subsequently decide the inherent risks that come out from the business relationship. By contrast, classifying an entity as a *Large entity* with profit margins continuously growing would imply fewer risks for the client.

## REFERENCES

1. [Cam03] Chamber of Commerce and Industry Brasov, 2002 - *Top Companies Brasov County*, 2003, <http://www.ccibv.ro>.
2. [Cap05] *Romania Forecast, Lumea in 2006, Romania in 2006*, Revista Capital, 2005, pp. 49.
3. [Cha03] S. Chakrabarti, *Mining the Web - Discovering Knowledge from Hypertext data*, Morgan Kaufmann Publishers, 2003, pp. 290-295.
4. [Eco04] The Economist, *Romania Forecast, The World in 2005*, The Economist Newspaper Limited, 2004, pp. 36, 93.
5. [Efi04] E-finance, *Romania Top 100 Companii - Topuri regionale, Topuri sectoriale, Top bursier, Companii cu rating A+*, Bucharest, 2004, <http://www.efinance.ro/>.
6. [Han02] L. Hancu, *Discovering Hidden Web Content*, Babes-Bolyai University, 2002, Graduation Paper.
7. [Han04] L. Hancu et al, *Participatory Search, Proceedings of the IADIS International Conference WWW/Internet 2004*, Madrid, Spain.
8. [Han07] L. Hancu, *Enhancing the Invisible Web*, Knowledge Engineering - Principles and Techniques Conference, 2007, Cluj Napoca, Romania.
9. [Han03] D. Hand, M. Manilla, P. Smyth, *Principles of Data Mining*, MIT Press, 2003, pp. 456-470.
10. [Kon06] M. Konchady, *Text Mining Application Programming*, Charles River Media, 2006, pp. 155-182.
11. [Recom] Registry of Commerce, <http://recom.onrc.ro>
12. [Top04] TopBiz 2004, Romanian Business Top, <http://www.softproeuro.eu/topbiz/>





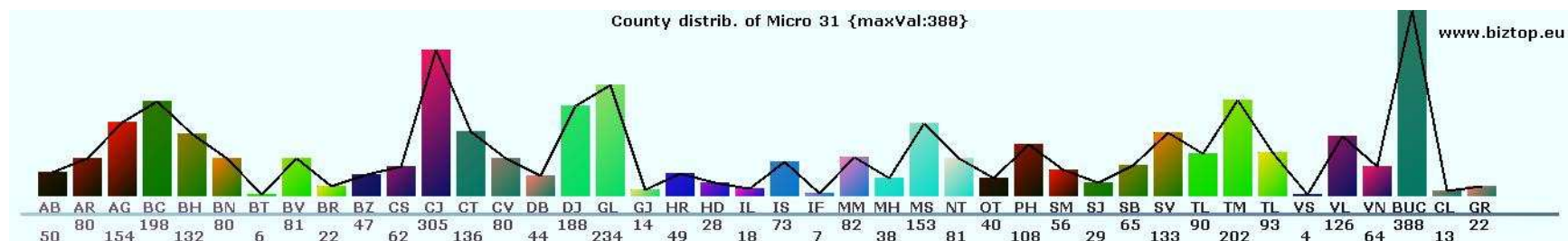


Figure 4. County Distribution of Micro 31

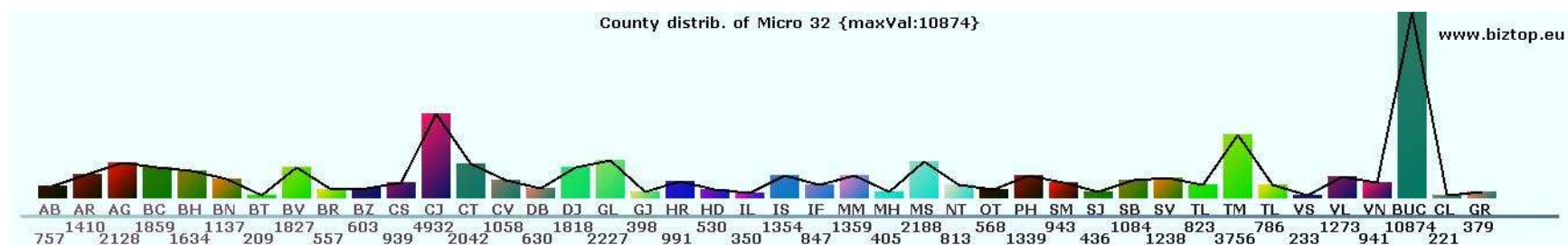


Figure 5. County Distribution of Micro 32

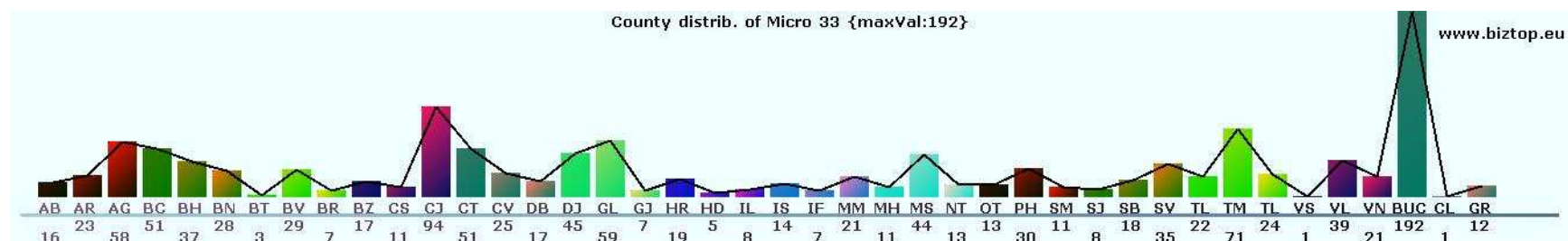


Figure 6. County Distribution of Micro 33

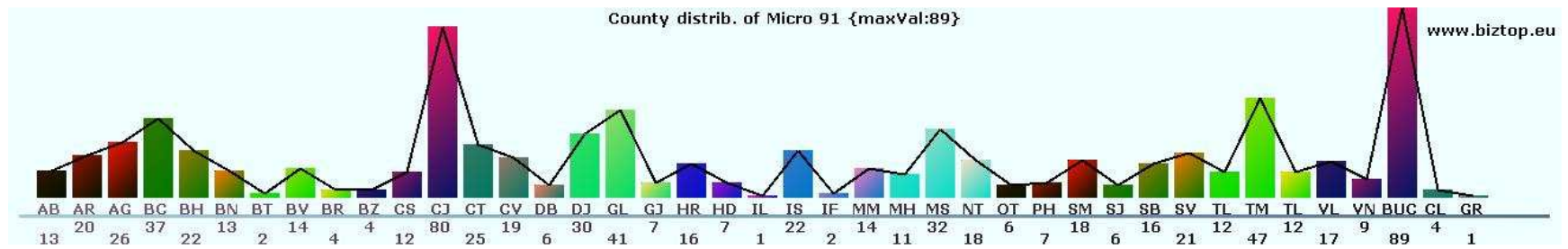


Figure 7. County Distribution of Micro 91

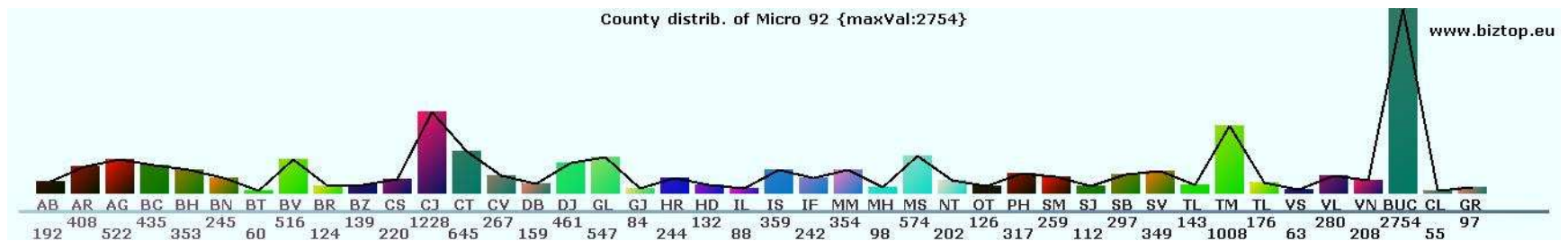


Figure 8. County Distribution of Micro 92

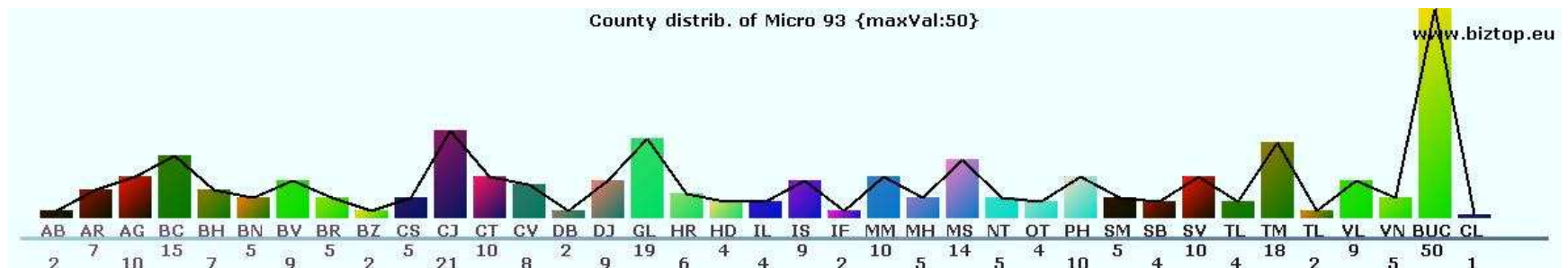


Figure 9. County Distribution of Micro 93

